

Economia Industrial



Preencha a **ficha de cadastro** no final deste livro

E receba gratuitamente informações sobre os
lançamentos e as promoções da Elsevier.

Consulte também nosso catálogo completo, últimos lançamentos e
serviços exclusivos no site
www.elsevier.com.br

Economia Industrial

Fundamentos Teóricos e Práticas no Brasil

Organizadores
David Kupfer
Lia Hasenclever



Amsterdam Barcelona Beijing Boston Filadélfia Londres Madrid
Mexico Milan Munique Orlando Paris Roma Sidney Tókio Toronto

© 2013, Elsevier Editora Ltda.

Todos os direitos reservados e protegidos pela Lei n. 9.610, de 19/02/1998.

Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida sejam quais forem os meios empregados: eletrônicos, mecânicos, fotográficos, gravação ou quaisquer outros.

Copidesque: Maya Indra Souarthes de Oliveira

Revisão: Carla Camargo

Coordenação editorial: Oitava Rima Produção Editorial

Editoração eletrônica: Thomson Digital

Elsevier Editora Ltda.

Conhecimento sem Fronteiras

Rua Sete de Setembro, 111 – 16º andar

20050-006 – Centro – Rio de Janeiro – RJ – Brasil

Rua Quintana, 753 – 8º andar

04569-011 – Brooklin – São Paulo – SP

Serviço de Atendimento ao Cliente

0800-0265340

sac@elsevier.com.br

ISBN: 978-85-352-6368-8

Nota: Muito zelo e técnica foram empregados na edição desta obra. No entanto, podem ocorrer erros de digitação, impressão ou dúvida conceitual. Em qualquer das hipóteses, solicitamos a comunicação ao nosso Serviço de Atendimento ao Cliente, para que possamos esclarecer ou encaminhar a questão.

Nem a editora nem o autor assumem qualquer responsabilidade por eventuais danos ou perdas a pessoas ou bens originados do uso desta publicação.

CIP-BRASIL. CATALOGAÇÃO-NA-FONTE
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ

E22

Economia industrial: fundamentos teóricos e práticas

no Brasil / organizadores, David Kupfer e

Lia Hasenclever. - 2.ed. - Rio de Janeiro: Elsevier, 2013.

ISBN 978-85-352-6368-8

1. Organização industrial (Teoria econômica). 2. Política industrial - Brasil 3. Economia - Brasil. I. Kupfer, David. II. Haguenauer, Lia.

12-7730.

CDD: 338.761

CDU: 338.45

*Ao meu filho Eduardo e a todos
aqueles que querem ver mais longe*
DAVID KUPFER

Aos meus alunos
LIA HASENCLEVER

Agradecimentos

Muitas pessoas contribuíram para a realização deste livro ao longo dos quase dois anos decorridos desde a concepção do projeto até a sua finalização em 2002, data de sua primeira edição. Posteriormente, por ocasião da produção de sua segunda edição, em 2012, outro conjunto de contribuições ortográficas ou de correções dos gráficos e textos, bem como sugestões de acréscimos foram recebidos dos leitores e está sendo incorporado a esta edição. Deixamos aqui registrado nossos agradecimentos especiais a Alexandre Florindo Alves, professor da Universidade Estadual de Maringá, Paraná, e a Henrique Cavalieri, aluno de doutorado do Instituto de Economia, da Universidade Federal do Rio de Janeiro, por suas contribuições na presente edição.

Agradecemos muito especialmente a Ricardo Redisch, responsável pelo projeto na Editora Campus/Elsevier, pelo incentivo permanente e principalmente pela paciência com os atrasos e as dificuldades impostos pelo árduo – e não raras vezes desalentador – cotidiano de trabalho em uma universidade pública brasileira. Esses agradecimentos estendem-se a André Wolff e Vanessa Vilas Bôas Huguenin durante os trabalhos para a segunda edição. Com certeza, a confiança depositada em nosso projeto e em nossa capacidade de realização foi decisiva para alcançarmos o objetivo inicial e agora podermos apresentar a segunda edição ampliada e revisada a nossos leitores.

Agradecemos também ao conjunto de autores colaboradores que mostraram excelente boa vontade ao aceitarem nossa recorrente intromissão, seja na delimitação dos escopos dos capítulos, seja nos inúmeros pedidos de revisão e, principalmente, pelo espírito acadêmico que demonstraram ao dedicarem seu precioso tempo a um projeto coletivo em uma época em que a regra parece ser o “cada um por si”. Nesta nova edição, três novos colaboradores foram adicionados: Jorge Chamie, Marta Lemme e Ricardo Torres, sendo os dois primeiros professores e o terceiro aluno de doutorado, todos do Instituto de Economia. Menção especial merece ser feita aos professores Carlos Frederico Leão Rocha, Helder Queiroz Pinto Junior, João Luiz de Simas Pondé e Jorge Nogueira de Paiva Britto que, além de contribuírem com seus respectivos capítulos, colaboraram ativamente nas discussões iniciais sobre a estrutura do livro.

Agradecemos ainda a Isleide Rosario Maeda e Patricia Moura Ferreira, mestrandas do IE/UFRJ, que muito nos ajudaram nas tarefas de edição do livro e a Letícia Teixeira e Marcia Espírito Santo pelo apoio de secretaria.

Por fim, agradecemos o apoio institucional da Direção do Instituto de Economia, na pessoa de seu diretor João Carlos Ferraz que, com seu entusiasmo, foi fundamental para viabilizar o projeto.

Os Autores

Alexis Toríbio Dantas fez doutorado em Economia (IEI/UFRJ). Atualmente ocupa o cargo de professor adjunto da Faculdade de Ciências Econômicas da Universidade do Estado do Rio de Janeiro e de professor titular do Instituto Universitário Cândido Mendes/Ipanema. Autor do livro *Empresas, Mercado e Concorrência*.

Carlos Eduardo Frickmann Young fez bacharelado em Economia (FEA/UFRJ), pós-graduação em Políticas Públicas (Ilpes-Cepal), mestrado em Economia (IEI/ UFRJ) e é doutor em Economia pela University College London. Atualmente é professor adjunto do Instituto de Economia da Universidade Federal do Rio de Janeiro. Entre 1996 e 1999 foi presidente da Sociedade Brasileira de Economia Ecológica. Autor/colaborador de diversos artigos e livros sobre desenvolvimento econômico, economia do meio ambiente, contas nacionais e economia industrial.

Carlos Frederico Leão Rocha doutorou-se em Economia Industrial (IEI/UFRJ,1995). Desde 1998 é professor do Instituto de Economia da Universidade Federal do Rio de Janeiro. Publicou artigos nacionais e internacionais na área de Economia Industrial, principalmente nos temas referentes a progresso técnico e processo de expansão da firma.

David Kupfer é engenheiro químico formado pela EQ/UFRJ, com mestrado e doutorado em Economia pelo IE/UFRJ onde atua como professor de graduação e pós-graduação e pesquisador. Na graduação, leciona as disciplinas Introdução à Microeconomia, Microeconomia I e II, Economia Industrial e diversas disciplinas eletivas sobre indústria brasileira além de ser o atual coordenador da área de microeconomia e economia industrial. Na pós-graduação é responsável pelas disciplinas Microeconomia e Organização Industrial. É também coordenador do Grupo de Pesquisa em Indústria e Competitividade (GIC/IE-UFRJ), tendo sido membro da equipe responsável pelo Estudo da Competitividade da Indústria Brasileira, uma das maiores pesquisas sobre a indústria nacional já conduzidas, realizada entre 1992 e 1994. É um dos autores do livro *Made in Brazil: Desafios Competitivos para a Indústria Brasileira* (Editora Campus), ganhador do Prêmio Jabuti na categoria Economia e Negócios em 1996.

Edmar Luiz Fagundes de Almeida fez doutorado em Economia Aplicada pelo Institut d'Economie et Politique de l'Energie (IEPF) da Université Pierre Mendès France, França (1999). Atualmente é pesquisador recém-doutor (CNPq) do Instituto de Economia da Universidade Federal do Rio de Janeiro e coordenador acadêmico do Programa de Formação de Recursos Humanos da Agência Nacional do Petróleo (ANP). Este é um programa interinstitucional que envolve o IE/UFRJ, o Programa de Planejamento Energético (PPE) e a Engenharia de Produção, ambos da Coppe/UFRJ. Também ocupa o cargo de editor responsável da revista eletrônica *Petróleo & Gás Brasil*, com dez volumes mensais, publicados desde novembro de 2000. (www.ie.ufrj.br/infopetro).

Germano Mendes de Paula é bacharel em Ciências Econômicas pela Universidade Federal de Uberlândia. Mestre e doutor em Economia pela Universidade Federal do Rio de Janeiro. Fez ainda o pós-doutorado em Economia na University of Oxford (Inglaterra). Desde 1990 é professor adjunto do Instituto de Economia da Universidade Federal de Uberlândia, onde leciona na graduação em Ciências Econômicas e no mestrado de Desenvolvimento Econômico as disciplinas de Economia Industrial, Economia de Empresas e Microeconomia.

Helder Queiroz Pinto Junior é doutor em Economia Aplicada pela Université Pierre Mendès France des Sciences Sociales/Institut d'Economie et Politique de l'Energie – Grenoble (França,1993). No momento, ocupa o cargo de professor adjunto no Instituto de Economia da Universidade Federal do Rio de Janeiro (IE/UFRJ), sendo membro do Grupo de Energia. Na pós-graduação, ministra o curso de Indústria de Infraestrutura e Desenvolvimento. Já na graduação, atua na área de Microeconomia e já ministrou cursos de Introdução à Economia, Economia Industrial, Regulação das Indústrias de Energia, Economia da Energia e Economia dos Recursos Naturais. Atuou como consultor da Diretoria Geral da Agência Nacional de Petróleo (1998-2000). Em 2001, esteve como professor visitante na University of Oxford no Centre for Brazilian Studies.

Hugo Pedro Boff é mestre e doutor em Ciências Econômicas pela Escola de Pós-Graduação em Economia da Fundação Getúlio Vargas (FGV-RJ); mestre em Estatística e licenciado em Ciências Econômicas pela Universidade Católica de Louvain (Bélgica). Atualmente é professor adjunto do Instituto de Economia (IE/UFRJ), onde ocupa também as funções de coordenador da área de Métodos Quantitativos e do Seminário de Pesquisa Econômica e coordenador do curso de atualização em Teoria Econômica (Cate/Ensync). Recebeu da Sociedade Brasileira de Econometria o prêmio Adriano Romariz Duarte (1998), com o trabalho “Cournot Competition under Knightian Uncertainty” (coautoria de Sergio Ribeiro da Costa Werlang).

Jacques Kerstsnetzky fez doutorado em Economia Política pelo Instituto de Economia da Universidade Federal do Rio de Janeiro (IE/UFRJ). Atualmente ocupa o cargo de professor adjunto na mesma instituição. Autor do livro *Empresas, Mercado e Concorrência*.

João Carlos Ferraz é economista e fez doutorado em Política Científica e Tecnológica na Sussex University. Atualmente ocupa o cargo de diretor geral do Instituto de Economia da Universidade Federal do Rio de Janeiro (IE/UFRJ) onde também é professor adjunto, ministrando as disciplinas de Política Industrial Comparada e Economia Industrial. É autor de muitos artigos e livros sobre Economia Industrial, sendo uma de suas principais publicações o livro *Made in Brazil: desafios competitivos para a indústria brasileira* (Editora Campus, 1996) em coautoria com David Kupfer e Lia Haguenauer.

João Luiz de Simas Pondé fez graduação em Economia na Faculdade de Ciências Econômicas (UFBA). Completou o mestrado e o doutorado no Instituto de Economia da Universidade Federal do Rio de Janeiro (IE/UFRJ). Desde 1996 exerce o cargo de professor do IE/UFRJ, onde leciona as disciplinas de Microeconomia (graduação) e Metodologia Econômica (graduação e pós-graduação). Autor de vários artigos publicados em revistas nacionais e anais de congressos sobre Economia Institucionalista, Teoria da Firma e Políticas de Defesa da Concorrência.

Jorge Chami Batista é engenheiro de produção, bacharel e mestre pela UFRJ, com pós-graduação em Economia –Ph.D. pela Universidade de Cambridge e M.Sc. pela LSE). Professor da UFRJ desde 1985, orientou inúmeros trabalhos acadêmicos de graduação e pós-graduação (mestrado e doutorado) nas áreas de Economia Internacional e Crescimento Econômico. Foi professor/pesquisador visitante nas Universidades de Kobe (Japão), de Roma Tor Vergata (Itália) e Técnica de Lisboa (Portugal), secretário de Comércio Exterior durante a elaboração e implementação do Plano Real e do Mercosul, colaborador do The Economist Intelligence Unit (1993;1995-2000), participou dos consórcios que avaliaram para o BNDES três empresas siderúrgicas (Usiminas, CSN e Acesita), a Rede Ferroviária Federal e a Vale do Rio Doce, elaborou estudos para diversas instituições e empresas (BID, BNDES, MRE, Cepal, CNI/Funex, Ipea, Brahma, Gerdau, Açominas, Rede Globo, Lloyds Bank, entre outras), e publicou vários artigos e livros no Brasil e no exterior.

Jorge Nogueira de Paiva Britto é bacharel em Ciências Econômicas pela UFRJ, com mestrado e doutorado pelo IE/UFRJ. Professor adjunto do Departamento de Economia da Universidade Federal Fluminense, ministra cursos na graduação e pós-graduação, nas áreas de Microeconomia e Organização Industrial. Autor de diversos artigos na área de Economia Industrial, com ênfase em temas relacionados a cooperação interindustrial e progresso técnico.

Larry Carris Cardoso é doutorando em Economia pelo Instituto de Economia da Universidade Federal do Rio de Janeiro. Atualmente é coordenador de negociação de câmbio da Petrobras. Foi consultor do Instituto de Pesquisa Econômica Aplicada (Ipea).

Lia Haguenauer fez bacharelado em Economia pela Universidade Federal do Rio de Janeiro (UFRJ) e pós-graduação na Escola de Pós-Graduação em Economia, Fundação Getúlio Vargas. Foi professora e pesquisadora do Instituto de Economia da UFRJ. Autora, juntamente com João Carlos Ferraz e David Kupfer, do livro *Made in Brazil, Desafios Competitivos para a Indústria* (Editora Campus, 1996) e, com Luiz Bahia, Paulo Furtado de Castro e Marcio Ribeiro, de *Evolução das Cadeias Produtivas Brasileiras na Década de 90* (Ipea, 2001). Faleceu em 24 de maio de 2002.

Lia Hasenclever é economista formada pela Faculdade de Economia e Administração, mestre em Economia Industrial pelo Instituto de Economia Industrial e doutora em Engenharia de Produção pela Coppe, todos da UFRJ. Desde 1978 é professora do Instituto de Economia (UFRJ) e, atualmente, é membro do grupo de pesquisa Economia da Inovação. Leciona as disciplinas de Metodologia de Pesquisa e Elaboração de Monografia, Análise Microeconômica, Economia Industrial, Economia do

Empreendedorismo e Gestão da Inovação no âmbito dos cursos de graduação e da pós-graduação. Entre 1978 e 1989 ocupou as funções de coordenadora do curso de Economia, assessora da Decanato do Centro de Ciências Jurídicas e Econômicas para assuntos de graduação e também a de conselheira do Conselho de Ensino de Graduação da UFRJ, quando se destacou a implantação do novo currículo de economia. Autora de vários artigos e capítulos de livros sobre o tema economia da inovação, sendo a mais recente publicação o livro *O Acordo Trips na OMC e a Proteção Patentária no Brasil: Mudanças Recentes e Implicações para a Produção Local e o Acesso da População aos Medicamentos* (editora ENSP/Fiocruz), em coautoria com Jorge Bermudez, Ruth Epstejn, e Maria Auxiliadora Oliveira.

Lucia Silva Kubrusly bacharelou-se em Física na PUC/RJ em 1974, e em 1979 ingressou no Programa de Engenharia de Sistemas Coppe/UFRJ, onde fez seu mestrado e doutorado, terminando em 1985. Atualmente leciona Estatística, Econometria, Modelos Lineares, no Instituto de Economia/UFRJ.

Luciano Dias Losekann é doutorando do Instituto de Economia/UFRJ. Lecionou a disciplina Economia dos Recursos Naturais (em conjunto com a professora Carmen Alveal) no curso de graduação em economia do Instituto de Economia da UFRJ e Introdução à Economia no curso de graduação em Relações Internacionais da Escola de Relações Internacionais da UniverCidade. Publicou diversos artigos sobre Economia da Energia, Telecomunicações e Energia Elétrica.

Luís Otávio de Figueiredo Façanha fez mestrado em Economia na Universidade de Vanderbilt (Tennessee/EUA) e doutorado em Economia Industrial pelo IEI/UFRJ. Atualmente é professor adjunto da Universidade Federal do Rio de Janeiro, onde leciona disciplinas nos cursos de graduação e pós-graduação, principalmente as matérias ligadas à Economia Industrial nacional e estrangeira e à Teoria dos Jogos; além disso, trabalha também na orientação de alunos em bolsas de pesquisa, monografias e teses de mestrado. Prestou assessoria e consultoria para algumas instituições governamentais como BID, MCT, Ipea e Finep. É autor e coautor de diversos artigos e capítulos de livros em publicações especializadas em economia e administração.

Luiz Martins de Melo é doutor em Economia Industrial pelo Instituto de Economia da UFRJ. Atualmente é professor adjunto da UFRJ onde leciona as disciplinas Microeconomia I e II, Economia Industrial, Gestão da Inovação, Economia de Empresas e Estratégia Industrial, Finanças Corporativas. É também consultor de várias empresas e projetos, entre os quais: Programa Nacional de Metrologia – Inmetro, Rio-2004 – Coordenador do Projeto Olímpico, Consejo Federal de Inversiones (CFI), Banco Interamericano de Desenvolvimento (BID) e da Organização de Pré-Inversion Latino-Americana y del Caribe (Opalc). Entre 1986 e 1988 foi diretor de Planejamento da Financiadora de Estudos e Projetos (Finep). Publicou diversos artigos sobre financiamento da inovação industrial, sistemas produtivos locais, sistemas de inovações de empresas e estratégia empresarial.

Magdalena Cronemberger Góes é economista e atualmente exerce a função de Gerente do Programa de Modernização das Pesquisas Econômicas do IBGE.

Marcelo Resende de Mendonça e Silva é doutor em Economia (D.Phil. in Economics), University of Oxford, 1997; mestre em Economia (M.A. in Economics), University of Pennsylvania, 1993; mestre em Economia, PUC/RJ, 1989; bacharel em Psicologia, Uerj, 1990; bacharel em Ciências Econômicas, Uerj, 1985. Atualmente exerce o cargo de professor adjunto do Instituto de Economia da UFRJ, onde leciona como principais disciplinas Microeconomia, Teoria dos Jogos, Estatística, Organização Industrial. Publicou numerosos artigos em periódicos nacionais e internacionais, principalmente na área de Economia de Infraestrutura.

Margarida Maria Gomes Pereira Sarmiento Gutierrez fez doutorado na área de Economia Industrial (IEI/UFRJ). Atualmente é professora adjunta do Instituto de Economia onde leciona as disciplinas de Introdução à Economia (Microeconomia e Macroeconomia) e Contabilidade Social no curso de graduação. É também redatora do Boletim de Conjuntura do Instituto de Economia da UFRJ – Seção Comércio Exterior. Ministrou vários cursos sobre Economia Empresarial em diversos cursos de Especialização do Instituto de Economia. Foi pesquisadora na Fundação Getúlio Vargas no Departamento de Estudos Governamentais. Trabalhou como economista da Fundação Instituto Brasileiro de Geografia e Estatística (IBGE). Foi professora da Sociedade Brasileira de Instrução Cândido Mendes e do Departamento de Economia da Pontifícia Universidade Católica (PUC/RJ). Atuou também como professora Assistente do Curso de Contabilidade Nacional da Escola Nacional de Ciências Estatísticas (Ence). Trabalhou como economista da Financiadora de Estudos e Projetos (Finep) e também como economista da Petrobras, na área de Planejamento Estratégico. Possui vários trabalhos publicados.

Maria Cecília J. Lustosa é doutora em Economia (IE/UFRJ), mestre em Economia (IEI/UFRJ) e bacharel em Ciências Econômicas (PUC/RJ). É professora da UFAL e professora colaboradora do mestrado em Sistemas de Gestão (Latec/UFF) na disciplina Economia do Meio Ambiente. É pesquisadora do Grupo de Economia do Meio Ambiente e Desenvolvimento Sustentável (IE/UFRJ) e diretora da Sociedade Brasileira de Economia Ecológica. É coautora do livro *Greening of Brazilian Business* (Unesco, 1999) e possui diversas publicações na área de meio ambiente.

Maria Tereza Leopardi Mello é advogada e doutora em Economia pelo Instituto de Economia da Unicamp. Atualmente exerce o cargo de professora do Instituto de Economia da UFRJ, onde leciona disciplinas de Instituições de Direito, Direito Econômico e Defesa da Concorrência. Desenvolve estudos e pesquisas sobre temas de Direito e Economia, Defesa da Concorrência, Regulação.

Mariana Iooty de Paiva Dias é economista formada pelo Instituto de Economia da UFRJ e doutoranda em Economia pelo mesmo Instituto. Atualmente exerce a função de pesquisadora visitante do Departamento de Economia da University of Reading (Inglaterra). Desde 1997 está associada ao grupo de Estrutura Industrial e Competitividade do IE/UFRJ, como pesquisadora. Lecionou as disciplinas de Microeconomia 1 e 2 e Economia Industrial na Universidade Cândido Mendes. Autora de diversos artigos em revistas nacionais e estrangeiras especializadas em Economia, normalmente versando sobre temas como Economia Industrial, Fusões e Aquisições na Indústria Brasileira e na Indústria Internacional.

Marina Honorio de Souza Szapiro é doutoranda do Instituto de Economia da UFRJ e pesquisadora associada à Rede de Sistemas Produtivos Locais do Grupo de Inovação do Instituto de Economia da UFRJ. Mestre em Economia Industrial e Tecnológica no Instituto de Economia da UFRJ, com a dissertação “Capacitação Tecnológica em Telecomunicações: Desenvolvimento e Impactos da Reestruturação do Setor” (1999). É graduada pelo Instituto de Economia da UFRJ (1992). É autora de vários artigos publicados por editoras nacionais e estrangeiras, geralmente ligados ao tema Economia de Infraestrutura, mais especificamente Telecomunicações.

Mario Luiz Possas graduou-se em Engenharia de Comunicações pelo Instituto Militar de Engenharia (IME, 1972). Fez mestrado (1977) e doutorado (1983) em Economia pela Universidade Estadual de Campinas (1977). Atualmente é professor titular do Instituto de Economia da Universidade Federal do Rio de Janeiro, onde ministra disciplinas de graduação (Macroeconomia) e pós-graduação (nas áreas de Microeconomia e Economia Industrial, além de Dinâmica Macroeconômica). Também participa de cursos de especialização, inclusive coordenando curso de extensão sobre Economia e Direito da Concorrência e da Regulação. Já orientou cerca de 40 teses de Mestrado e Doutorado, na UFRJ e na Unicamp. Coordenou pesquisas em diferentes áreas, especialmente em Economia Industrial, abrangendo entre outros temas estrutura industrial e competitividade, teoria e políticas de concorrência e regulação de setores de infraestrutura. É consultor da Capes (onde presidiu a Comissão de Consultores da área de Economia), do CNPq e da Faperj. É membro do conselho editorial e referência de várias revistas acadêmicas de Economia, tais como *Revista Brasileira de Economia*, *Revista de Economia Política*, *Pesquisa e Planejamento Econômico* e *Revista de Economia Contemporânea*. Consultor econômico na área antitruste, na qual tem elaborado pareceres e opiniões técnicas sobre diversos casos de concentração econômica e práticas anticompetitivas. Autor de numerosos artigos em revistas acadêmicas nacionais e internacionais, livros e capítulos de livros versando principalmente sobre Estrutura de Mercado, Concorrência e Economia Industrial.

Marta Calmon Lemme é bacharel em Ciências Econômicas pela UFRJ, com mestrado pelo IE/UFRJ. Atualmente é professora do Instituto de Economia da UFRJ, onde leciona como principais disciplinas Comércio e Investimento Internacionais e Introdução à Microeconomia, no curso de graduação. Em sua experiência profissional na área de comércio internacional, desde 1986, destaca-se: painelista em várias controvérsias na OMC referentes à aplicação de medidas *antidumping* e compensatórias, a partir de 1998; Coordenadora Geral de Apoio ao Exportador e Negociações Internacionais, do Decom/Secex/MDIC; participação em negociações na OMC, Alca e Mercosul-União Europeia; elaboração dos regulamentos brasileiros sobre defesa comercial, resultantes da Rodada Uruguai; integrante da equipe que formulou as Reformas Tarifárias de 1988, 1989 e 1990; coordenação de equipe responsável por análises de pedidos de alteração de imposto de importação e de ex-tarifários; e integrante da equipe negociadora da Tarifa Externa Comum.

Patricia Moura Ferreira é economista especialista pela Universidade Federal do Rio de Janeiro e mestrandona Instituto de Economia da UFRJ. Atualmente exerce o cargo de analista na Financiadora de Estudos e Projetos do Ministério da Ciência e Tecnologia (Finep).

Paulo Bastos Tigre é economista (UFRJ), mestre em Engenharia da Produção (Coppe/UFRJ) e Ph.D. em Science and Technology Policy (SPRU/Universidade de Sussex). Atualmente exerce a função de professor titular de Organização Industrial no Instituto de Economia da UFRJ, onde coordena o Grupo de Economia da Inovação e é consultor de organismos internacionais, incluindo Unido, Cepal, OECD, Aladi, BID, Flasco para assuntos de economia industrial e da tecnologia. Além disso, foi diretor da Cobra Computadores (1986/1988) e pesquisador visitante nas universidades da Califórnia, Berkeley (1997) e Paris XIII (1996). É autor de mais de 60 livros e artigos, incluindo *Technology and Competition in the Brazilian Computer Industry*, Pinter Publishers (London) e Editora Campus (Rio de Janeiro) e *Tecnologia e Meio Ambiente: Oportunidades para a Indústria*, Editora da UFRJ.

Reinaldo Gonçalves é bacharel em Economia (UFRJ), mestre em Engenharia da Produção na Coppe (1974) e em Economia pela EPGE-FGV (1976). Ph.D. em Economia pela University of Reading (Inglaterra, 1986) e Livre-docente em Economia Internacional (UFRJ, 1991). Atualmente é professor titular de Economia Internacional da UFRJ e acumula os cargos de diretor da Sociedade Brasileira de Economia Política, diretor da Associação Nacional dos Cursos de Graduação em Economia e conselheiro titular do Conselho Federal de Economia. Foi professor visitante da *Oirecteur d'Etudes, Ecole des Hautes Études en Sciences Sociales, Maison des Sciences de l'Homme* (Paris, 1996) e economista das Nações Unidas (Unctad, Genebra, 1983-87). Publicou mais de uma centena de trabalhos em 16 países na Europa (França, Inglaterra, Itália, Suécia, Suíça, Portugal e antiga Iugoslávia), Ásia (Japão, Coreia do Sul e Índia), África (Cabo Verde), América do Norte (Estados Unidos e México) e América do Sul (Brasil, Chile e Venezuela). Dentre os seus principais trabalhos no Brasil pode-se mencionar os livros mais recentes: *Globalização e Desnacionalização* (Paz e Terra, 1999) e *O Brasil e o Comércio Internacional* (Contexto, 2000). Este último livro recebeu o Prêmio Jabuti 2001, como primeiro lugar da categoria Economia, Direito e Administração.

Ricardo Torres é bacharel em Ciências Econômicas pela Universidade Federal do Paraná, mestre em Economia pela Universidade Federal de Santa Catarina, aluno de doutorado no Programa de Pós-Graduação em Economia do Instituto de Economia da UFRJ, pesquisador nas áreas de Economia da Indústria e da Inovação, professor do curso de Ciência Econômicas da Universidade Federal Fluminense nas disciplinas de Microeconomia, Macroeconomia e Economia do Setor Público.

Ronaldo Fiani é bacharel em Economia pela Faculdade de Economia e Administração, mestre em Economia Industrial e, atualmente cursa o doutorado em Economia, com projeto de doutorado sobre reforma das telecomunicações na Inglaterra nos anos 1980/1990, todos pela UFRJ. Desde 1990, exerce o cargo de Professor Assistente no IE/UFRJ e leciona as cadeiras de Microeconomia e Teoria da Regulação Econômica, além de trabalhar como assessor da Superintendência de Abastecimento desde janeiro de 2001. Exerceu outros cargos públicos, entre eles o de assessor do Departamento de Política Econômica (Ministério da Fazenda) e uma outra função na Agência Nacional de Petróleo (ANP). Publicou artigos em revistas especializadas sobre Economia e Políticas Públicas.

Victor Prochnik é professor doutor do Instituto de Economia e do mestrado em Contabilidade da UFRJ, onde leciona as disciplinas de Economia de Empresas, Economia da Informação e Estratégia Empresarial. Publicou vários artigos sobre Economia Industrial, sendo os dois últimos: *Transferência de Rotinas em Empresas Transnacionais* (coautoria com Tiago Fernandes) e *Cadeias Produtivas e Oportunidades de Investimento no Nordeste Brasileiro* (coautoria com Lia Haguenauer).



Introdução

Lia Hasencllever e David Kupfer

1.1 Antecedentes e a Segunda Edição

Desde os anos 1970, o Instituto de Economia (IE) da Universidade Federal do Rio de Janeiro (UFRJ) vem desenvolvendo uma importante linha de pesquisa voltada para a análise da dinâmica dos diversos setores da indústria brasileira. Em 1978 foi criado o Instituto de Economia Industrial (IEI), órgão complementar da UFRJ, responsável então pela pós-graduação em Economia Industrial e da Tecnologia. Em 1985, com a implementação do novo currículo mínimo de Economia, a então FEA, responsável pelo curso de graduação, incluiu entre as matérias de escolha a matéria Economia Industrial que, desde então, é ensinada por meio de uma disciplina obrigatória e diversas disciplinas eletivas para os cursos de graduação e pós-graduação. Em 1994, a FEA e o IEI se fundiram para dar lugar ao atual Instituto de Economia.

Entretanto, o ensino da matéria ressentia-se da ausência de um livro-texto que não somente fosse escrito em português, mas também contivesse um escopo adequado ao tipo de trabalho que se fazia – e se faz – nas diferentes disciplinas em que a matéria é ministrada: fornecer aos alunos uma visão abrangente da evolução dos principais instrumentos analíticos para o estudo das empresas e dos mercados e discutir as particularidades da indústria brasileira.

O livro-texto, que ora temos o prazer de apresentar aos nossos leitores, foi preparado justamente com esse escopo para ser utilizado nas disciplinas de Economia Industrial dos cursos de graduação de Economia, Administração, Engenharia de Produção e outros cursos afins, bem como em cursos de especialização e MBAs em Comércio Exterior, Economia das Infraestruturas, Regulação e Defesa da Concorrência dentre outros. Ele é o resultado de um esforço coletivo dos professores da área de Microeconomia e Economia Industrial do IE/UFRJ, que buscou reunir a experiência acumulada na casa através do ensino de graduação e pós-graduação e da realização de diversos estudos empíricos na área.¹

A presente segunda edição traz uma versão revista da primeira edição, ampliada de dois capítulos adicionais. O primeiro é um capítulo que apresenta o paradigma fundador da Economia Industrial e que, na edição anterior, havia sido mencionado apenas na introdução e o segundo é um capítulo que aborda a questão da política comercial, complementando a Parte VI do livro.

1.2 Filiação Teórica

Os termos Economia Industrial, oriundo da língua francesa, e Organização Industrial, oriundo da língua inglesa, são indistintamente utilizados no Brasil para denominar a matéria Economia Industrial. É uma área de conhecimento relativamente recente que veio a florescer somente a partir dos anos 1950, motivada principalmente pela busca de novos meios e métodos para estudar a dinâmica real dos diversos setores industriais empreendida por diferentes autores insatisfeitos com a tradição microeconômica neoclássica.

A Economia Industrial abriga uma grande diversidade de linhas de pensamento, que podemos agregar em duas correntes principais, que serão denominadas abordagem tradicional (*mainstream*) e abordagem alternativa (schumpeteriana/institucionalista). Sem sermos exaustivos, podemos dizer que essas correntes partem de um conjunto de questões empíricas comuns: qual é a natureza e qual o funcionamento real das empresas, dos mecanismos de coordenação de suas atividades e, portanto, de seus mercados? A partir da resposta a essas questões empíricas comuns, as correntes teóricas divergem radicalmente em relação aos seus métodos de análise e ao papel representado pelas empresas em sua estrutura teórica, bem como ao que entendem por concorrência.

A primeira corrente estruturou-se progressivamente a partir do trabalho de Joe S. Bain, culminando com a representação teórico-analítica proposta por F. M. Scherer, conhecida como modelo Estrutura-Conduta-Desempenho (Modelo ECD). Tem como principal objetivo a análise da alocação dos recursos escassos sob as hipóteses de equilíbrio e maximização dos lucros. Recentemente, alguns desenvolvimentos na matematização dos modelos de empresa e de interação entre essas (teoria dos jogos) levaram os estudiosos a rebatizar essa corrente de Nova Economia Industrial (NEI). Nesse desdobramento há um aumento da importância das condutas empresariais na determinação das estruturas de mercado; a empresa deixa de ser um agente passivo para adotar estratégias discricionárias. Os principais fundamentos da ação governamental na preservação da concorrência (regulação) e seus efeitos sobre a estrutura da indústria e sobre a estratégia das empresas (defesa da concorrência) são oriundos desta corrente.

A segunda corrente filia-se diretamente a Joseph Schumpeter e tem como objetivo central o estudo da dinâmica da criação de riqueza das empresas. Essa corrente tem uma preocupação menos normativa que a anterior, levando em consideração as instituições e a história como elementos fundadores da teoria. Nesse sentido, a organização interna da empresa não resulta de um procedimento de minimização de custos, mas da constituição de capacidade de inovação. A empresa é um objeto de estudo relevante, razão pela qual o estudo de suas estratégias se torna obrigatório para a compreensão da dinâmica dos setores industriais.

Ainda dentro da corrente alternativa, podemos incluir a contribuição de Oliver Williamson, que ampliou e consolidou a tradição inaugurada por Ronald Coase, ao enfatizar a natureza institucional da empresa visando explicar as diferentes formas de organização interna das corporações, as configurações industriais daí decorrentes e as implicações sobre o funcionamento dos mercados. Ainda que a contribuição de Oliver Williamson tenha unidade de análise distinta da corrente schumpeteriana – transação *versus* produção – ambas as contribuições têm interseções importantes no que diz respeito ao conceito de firma e suas fronteiras.

O fato de a NEI – ramo mais recente da corrente tradicional, oriunda do Modelo ECD – também dar destaque maior às estratégias empresariais, aproxima-a da corrente alternativa sem, entretanto, convergirem. Na próxima seção iremos discutir um pouco mais essa questão para que o leitor possa compreender a evolução dos assuntos e modelos abrangidos pelo livro.

Enfim, é propósito deste livro apresentar uma visão abrangente das duas correntes teóricas acima apresentadas visando a refletir o conhecimento acumulado sobre a matéria em ambas as perspectivas sem, no entanto, dedicar muito espaço ao debate entre as correntes, como é mais adequado para um livro-texto.

1.3 Empresas, Mercados e a Economia Industrial

Relações entre empresas, mercados, instituições, processos. Essa complexidade é o cerne da Economia Industrial, cujo objetivo é o estudo do funcionamento real dos mercados. Até que ponto, no entanto, é possível generalizar os princípios que explicam a dinâmica concreta de um mercado sem que seja necessário resgatar os elementos específicos da evolução histórica desse mercado é o desafio com que a disciplina se depara e a motivação das diferentes contribuições teóricas que têm surgido desde os anos 1950.

A rapidez e a intensidade com que as tecnologias e as formas de organização da produção industrial vêm se transformando desde meados do século XX têm atribuído à Economia Industrial e à temática a ela associada – preços, custos, inovação, crescimento das empresas, competitividade – um lugar central na análise econômica contemporânea. No Brasil, observa-se também um interesse crescente no estudo desses temas, a partir dos anos 1980, quando a matriz industrial se completa, e posteriormente, nos anos 1990, com a abertura comercial e o fim do regime de regulação apoiado no modelo de substituição de importações, e o consequente aumento de concorrência entre as empresas.

A concorrência é o fenômeno mais característico das economias capitalistas. É indiscutível, no entanto, que a construção do conceito de concorrência encerra uma grande complexidade. Desde as noções que lhe são preliminares como as de empresa, indústria e mercado, até a identificação das variáveis básicas descritivas das estruturas dos mercados e das condutas das empresas,

a noção de concorrência apresenta-se como um objeto analítico que insiste em se situar além da capacidade explicativa das formulações teóricas disponíveis.

Para os economistas neoclássicos – corrente tradicional –, a concorrência surge como um estado no qual prevalecem certas premissas sistêmicas que garantem o equilíbrio através da transformação de todos os agentes em tomadores de preço – na verdade, ausência de rivalidade entre as empresas. Ainda dentro dessa corrente, os desenvolvimentos mais recentes da NEI consideram concorrência um jogo em que as empresas disputam parcelas de um mercado e os lucros nele gerados mediante a adoção ativa ou reativa de políticas de preços, esforço de venda, diferenciação de produtos e outras. Para os neo-schumpeterianos – corrente alternativa –, a concorrência é analisada como um processo em que cada agente busca se diferenciar dos demais para reter ganhos monopólicos, sendo, no entanto, a inovação de processo, de produto ou organizacional o principal fato gerador dessas quase-rendas.

O **mercado** é pensado como um espaço abstrato no qual se definem preços e quantidades das mercadorias transacionadas por consumidores (demanda) e empresas (oferta). Em cada mercado vigora um dado padrão de concorrência definido a partir da interação entre as características estruturais dominantes e as condutas praticadas pelas empresas que nele atuam. A definição do padrão de concorrência vigente em cada mercado apresenta uma intensa controvérsia ainda não estabilizada em Economia Industrial. Nesse debate, estão englobadas desde as visões nas quais a estrutura de mercado é considerada um dado e condiciona univocamente o comportamento das empresas na tradição do chamado Modelo ECD até a visão virtualmente oposta, segundo a qual é a estrutura que é endogenamente determinada como resultado das estratégias concorrenceis adotadas pelas empresas, em um dado mercado, na abordagem alternativa e também na NEI, com um menor grau de liberdade do que na abordagem alternativa.

1.3.1 A ABORDAGEM TRADICIONAL E SUAS REVISÕES

A teoria econômica neoclássica, que sustenta os pilares da visão tradicional, encara uma decisão econômica como uma escolha. O comportamento decisivo é tomado como uma atitude racional que informa o processo de escolha, um simples problema de maximização. É possível sofisticar um pouco essa abordagem, levando em conta que uma empresa em vez de um ator é o resultado de comportamentos de múltiplos atores.

Nesta tradição, baseada no atomismo, as decisões das empresas estão subordinadas à determinação da existência de um vetor de preços que compatibilize as decisões individuais. Em particular, com algumas suposições acerca das preferências dos agentes e das características das técnicas produtivas e supondo-se que os agentes são tomadores de preços, garante-se a existência desse vetor. Nesse nível de abstração, é justificável a adoção da hipótese de concorrência perfeita, onde os agentes não rivalizam entre si – ausência da importância do estudo das políticas de preços e outras estratégias competitivas –, mas simplesmente se deixam disciplinar pelo mercado.

Desde a sua formulação pioneira, o modelo de concorrência perfeita foi alvo de severo questionamento, em particular, no que toca à baixa aderência de suas premissas à realidade econômica observada. Do debate quanto à existência de preferência dos consumidores, de funções de produção com rendimentos constantes de escala, de estruturas oligopolistas estáveis e outras questões, foram surgindo diversas revisões das proposições neoclássicas originais.

A partir da década de 1950, as proposições que utilizam a heurística do Modelo ECD passaram a ocupar o posto de paradigma teórico por excelência das teorias microeconômicas preocupadas com as questões práticas ligadas às empresas, às indústrias e aos mercados. É sob este paradigma que a Economia Industrial (ou Organização Industrial) consolida-se como uma matéria especializada da ciência econômica.

É consensual para os estudiosos desta matéria o caráter seminal da obra de Joe S. Bain na constituição da metodologia estrutura-conduta-desempenho como ferramenta básica de análise. Na tradição de Bain, que depois passou a ser reconhecida como hipótese estruturalista básica, as condutas não importavam, a ponto de se considerar que a estrutura – representada por variáveis como grau de concentração ou de barreiras à entrada – determinava direta e inequivocamente o desempenho do mercado. O desempenho, por sua vez, é avaliado em termos do desvio da taxa de lucro efetiva em relação à taxa ideal em eficiência alocativa – o ótimo de Pareto – o que significa de fato o desvio do preço efetivo em relação ao custo marginal de produção.

A partir das formulações pioneiras de Bain, basicamente ligadas à determinação de preços-limite na presença de barreiras à entrada, as proposições teóricas foram se aprofundando, mas também se diversificando. O aprofundamento consistiu, basicamente, na ampliação das variáveis incluídas no esquema analítico original, principalmente quanto aos elementos de conduta, como propaganda e pesquisa e desenvolvimento, e não somente à política de preços das empresas. A busca desse aprofundamento expressou-se na realização intensiva de pesquisas empíricas, em particular durante a década de 1960. Esse movimento, de certa

forma, contribuiu inicialmente para ampliar a crença no poder explicativo do Modelo ECD e tornar mais abrangente o seu escopo normativo. Mas certos resultados empíricos e, principalmente, certos questionamentos teóricos levaram a um processo de revisão do paradigma e à busca de diversificação das teorias de organização industrial. Nesse processo, o questionamento crítico de alguns supostos fundamentais da teoria terminou por expor graves lacunas na concepção original. De certa forma, as tentativas de completar a teoria, ao contrário de bem-sucedidas, acabaram por desfigurá-la seriamente.

Uma das lacunas do Modelo ECD pioneiro era a falta de importância atribuída às condutas das empresas no processo de concorrência. A resposta foi a aceitação da existência de causalidades menos rígidas, que se expressavam em uma relação interativa entre as variáveis de estrutura, conduta e desempenho. Com isso, passou-se a avaliar empiricamente todos os possíveis *feedbacks* entre as três categorias, enfraquecendo o modelo diante da múltipla causalidade das relações e da necessidade de encontrar soluções simultâneas para essas relações. Duas alternativas foram buscadas: estudos de caso e soluções matemáticas (teorias dos jogos). Ambas as alternativas se mostraram infrutíferas. Os estudos de caso eram muito particulares e pouco generalizáveis e o uso intenso da matemática se, por um lado, permitiu que soluções de oligopólio tivessem uma maior formalização e, na visão de alguns, maior rigor científico, por outro, voltou a privilegiar a conduta das empresas – e sua rivalidade – como a principal variável explicativa do funcionamento dos mercados, desconsiderando o papel das suas características técnico-econômicas (condições básicas da oferta e da demanda e grau de concentração).

Outra lacuna do paradigma ECD era a sua incapacidade de lidar com a existência de diferenciais de lucratividade entre empresas em uma mesma indústria. O problema é que, empiricamente, um dado grau de concentração de uma indústria pode abrigar variadas distribuições de tamanhos das empresas. Mesmo que se aceite correlação positiva entre grau de concentração e lucros excessivos em uma indústria, não há por que imaginar que todas as empresas de uma indústria concentrada partilhem igualmente esses lucros excessivos entre si. Como ademais, muitas das grandes empresas são diversificadas, pareceria mais pertinente que a unidade analítica adequada para as análises de Economia Industrial passasse a ser as grandes empresas e não mais os mercados (indústrias), tornando questionável o próprio objeto de análise do Modelo ECD.

Mas o principal questionamento com que o paradigma se defrontou foi a chamada questão da endogeneidade: se cada empresa escolhe seu nível de produção (e preços) em função de suas curvas de custos, funções de demanda e de expectativas que mantenham sobre a conduta das empresas rivais, o preço de mercado e os produtos de todas as empresas, para uma indústria em equilíbrio, são conjuntamente determinados. Isso por sua vez implica que, tanto o grau de concentração quanto os lucros, sejam variáveis endogenamente determinadas e não possam guardar relações de causalidade predefinidas. Ambas dependem, na verdade, das variáveis exógenas, assumidas como as curvas de custo, as funções de demanda e as expectativas de ação e reação dos concorrentes que cada empresa apresenta.

Claro está que, sendo pertinente a questão da endogeneidade, a noção de concorrência ver-se-ia obrigada a dar conta de variáveis muito mais complexas. E, ainda mais, essas variáveis incluiriam a própria conduta das empresas, baseada em expectativas de ação e reação, um objeto de difícil apreensão.

A hipótese da endogeneidade constituiu o ponto de partida, já na década de 1970, de uma corrente alternativa de análise da organização industrial baseada em teoria dos jogos, na qual as premissas do tipo ECD foram deixadas de lado. Na teoria dos jogos, ou NEI, formula-se um comportamento de equilíbrio das empresas em que estas ajustam quantidades, preços ou outras variáveis, de forma cooperativa ou não, resgatando assim os modelos de Cournot, Bertrand, Nash ou outros, basicamente ligados aos primórdios das teorias do oligopólio (em geral, duopólios).

Comparado metodologicamente com o paradigma ECD, as condições básicas e as condutas são as variáveis exógenas na teoria dos jogos, ao passo que a estrutura e o desempenho são as variáveis endógenas. As condutas são firmemente baseadas em expectativas, podendo, como é feito em jogos mais sofisticados, ser introduzidas incertezas quanto ao futuro.

Apesar de todas as críticas anteriormente sumarizadas, o paradigma ECD é ainda tanto um programa de pesquisa válido como um importante guia para a ação política. Fornece um conjunto estruturado de ideias e conceitos, um volume impressionante de resultados empíricos que sugerem que a estrutura de mercado está sistematicamente relacionada com o desempenho no mercado, sugerindo que concentração industrial e barreiras à entrada devem ser objetos de preocupação das autoridades regulatórias.

Mas também é verdade que a chegada dos anos 1980 trouxe novas questões que não puderam ser tratadas pelo Modelo ECD devido à sua intensa fragmentação. Com Scherer, o paradigma ECD havia perdido causalidade, a ponto desse autor, para muitos o responsável pelo mais completo e preciso livro-texto dessa linha teórica, se autoqualificar não como um estruturalista, mas como um “behaviorista”. A vertente empiricista econométrica mostrava-se esgotada, enredada em discussões de natureza muito mais estatísticas que econômicas. A NEI, apoiada no instrumental da teoria dos jogos, enfatizava a tal ponto a rivalidade concorrencial, expressa em suposições sobre as condutas das empresas, que vários de seus autores passaram a duvidar da importância da estrutura do mercado para a compreensão do seu funcionamento.

1.3.2 A ABORDAGEM ALTERNATIVA

A análise neoclássica parte de uma estrutura teórica universal a respeito da escolha racional e de comportamento para uma teoria dos preços e do bem-estar econômico. Nessa estrutura teórica pressupõe-se que os agentes se comportam racionalmente e maximizam suas funções de preferências; focam o estágio de equilíbrio alcançado, ou o seu movimento em direção ao equilíbrio; e excluem os problemas crônicos de informação, tais como as incertezas. O comportamento dos agentes é considerado um dado e, portanto, não é problematizado nas teorias e nas análises de cunho neoclássico.

As tentativas de desenvolver teorias não fundamentadas no equilíbrio por parte dos autores da corrente alternativa, denominados autores “neo-schumpeterianos” ou evolucionistas, com o acréscimo de contribuições dos autores institucionalistas, têm estimulado a construção de um novo paradigma microeconômico de natureza não determinística. Nessa busca, o caminho que tem se mostrado mais profícuo é o que toma por base uma visão evolucionista do processo de concorrência.

Esses autores têm como preocupação central a lógica do processo de inovação e seus impactos sobre a atividade econômica. Este é, claramente, um programa de pesquisas muito amplo e, neste contexto, as colocações sobre o processo de concorrência são ainda muito dispersas, fato que indica que ainda há um longo caminho a percorrer. Se concordam em substituir a noção de equilíbrio pela de trajetórias de evolução, em enfatizar o papel da mudança tecnológica na conformação das estruturas de mercado e no processo de mudança estrutural ou na atribuição de papel ativo por parte das empresas na definição da direção dessas mudanças, as formalizações dessas relações, em termos das variáveis-chave e das regularidades e causalidades relevantes são ainda pouco convergentes.

As ideias principais dos schumpeterianos-institucionalistas estão relacionadas com instituições, hábitos, regras e sua evolução. Não é comum na tradição schumpeteriana-institucionalista, entretanto, construir um único modelo geral fundamentado nessas ideias. Ao contrário, essas ideias facilitam uma abordagem específica e histórica para a análise. Com relação a isso, é possível inclusive fazer uma analogia com a abordagem utilizada na biologia. A biologia evolucionária tem poucas leis e princípios gerais pelos quais a origem e o desenvolvimento da vida podem ser explicados. A análise da evolução de um organismo específico exige dados detalhados sobre o organismo e seu ambiente, e também explicações específicas relevantes à espécie em consideração. A teoria evolucionária necessita tanto de teoria específica quanto de teoria geral. Em contraste, em física existem repetidas tentativas para formular uma teoria geral de todos os fenômenos materiais. Dessa forma, em sua relativamente maior ênfase em especificidades, a teoria econômica schumpeteriana-institucionalista está mais próxima da biologia do que da física.

A abordagem schumpeteriana-institucionalista parte de ideias gerais com relação ao homem, às instituições e à natureza evolucionária dos processos econômicos para ideias e teorias específicas, relacionadas com instituições econômicas singulares ou tipos de economia. De fato, existem vários níveis e tipos de análise, mas que são relacionados por meio dos conceitos de hábito e de instituição, ajudando a ligar o específico e o geral.

Todavia, a abordagem schumpeteriana-institucionalista não presume que a sua concepção do homem como um ser que toma decisões com base nos hábitos permita uma operacionalização imediata da teoria. Elementos adicionais são necessários. Em particular, um autor desta tradição precisa mostrar como grupos específicos de hábitos comuns estão embebidos em e são reforçados por instituições sociais específicas. Dessa forma, essa abordagem move-se do abstrato para o concreto. Os schumpeterianos-institucionalistas se valem das tradições de pesquisa psicológicas, sociológicas e antropológicas sobre o comportamento humano para entender esses hábitos e rotinas. Se há uma teoria geral, ela é indicativa de como desenvolver análises específicas de fenômenos igualmente específicos.

Os autores neo-schumpeterianos explicitam essas premissas de três formas: (1) a existência de assimetrias técnico-econômicas entre os agentes; (2) a existência de variedade tecnológica; e (3) a existência de diversidade comportamental entre os agentes. Os conceitos básicos da teoria que os autores buscam construir e que dão suporte a essas premissas são igualmente três: (1) a tecnologia é apropriável, cumulativa, tácita e irreversível; (2) existe incerteza quanto aos resultados dos esforços ou decisões tecnológicas (e não só em relação a elas); e (3) a despeito do anterior, existem paradigmas e trajetórias tecnológicas setoriais que ordenam o progresso técnico, fazendo da busca e seleção de inovações um processo não randômico, nem totalmente exógeno. O resultado dessa construção teórica é a obtenção de modelos evolucionistas que se contrapõem às formulações determinísticas habituais no pensamento neoclássico.

A questão central enfrentada pelos modelos evolucionistas é a tentativa de tratar a inovação, e, a partir dela, a concorrência, como um processo dependente do tempo, tanto lógico quanto cronológico. Concretamente, isso significa que a dinâmica a ser estudada não pode deixar de ser a dinâmica do processo de mudança. Nesse marco teórico, a preocupação de descrever a (falsa) dinâmica de ajuste de natureza estática comparativa não tem sentido e é irrelevante. O objetivo é tratar variáveis *path-dependent* e por isso a história tem que ser incorporada ao sistema teórico tanto no que diz respeito à história passada em decorrência da

natureza cumulativa das variáveis analisadas quanto em relação ao futuro que, em vista das condições de incerteza sob as quais se dá o processo decisório, não pode ser reduzido a sequências lógicas de tempos.

1.4 Estrutura e Características do Livro

Este livro dedica-se a registrar o que há de mais recente no campo da análise dos fenômenos que se observam na dinâmica dos mercados das economias capitalistas, com ênfase para o campo da concorrência industrial, no Brasil. Sua característica principal é a busca de homogeneidade no tratamento dos temas, diferentemente dos livros que se apresentam como uma coletânea de textos.

A homogeneidade foi buscada através de três princípios. O primeiro é a forma de organização das partes que preservam a evolução da matéria Economia Industrial e vai agregando a contribuição das diferentes filiações teóricas sobre as questões das empresas, dos mercados e da concorrência, apresentada neste capítulo. O segundo é a forma de organização dos capítulos. Cada capítulo aborda as definições dos termos, principais desenvolvimentos teóricos (sem enfatizar as controvérsias sobre o tema, e sim procurando as convergências para proporcionar ao leitor uma visão geral do tema), aplicações (relacionadas aos problemas, variáveis e indicadores decorrentes das proposições teóricas) e temas para discussão (envolve principalmente os resultados empíricos decorrentes das proposições teóricas e a problemática empírica resultante, com ênfase sobre os estudos empíricos brasileiros) e, quando for o caso, leituras sugeridas. O terceiro princípio, que busca manter a unidade e a coerência do livro, é o conjunto de termos econômicos utilizados pelos diferentes colaboradores. Fizemos um esforço de padronização de linguagem, inclusive no que tange às traduções, muitas vezes pela primeira vez, de termos já consagrados na literatura inglesa.

Outra importante característica do livro é que existem, para os temas tratados em cada capítulo, remissão para os encadeamentos com capítulos anteriores e posteriores, bem como um índice remissivo que ajuda o leitor a selecionar os assuntos de seu maior interesse.

O livro é composto de sete partes e 28 capítulos. A seguir descreveremos o conteúdo de cada parte e de cada capítulo.

A Parte I (Conceitos Básicos e Fundamentos) é formada de quatro capítulos que introduzem o leitor aos principais temas microeconômicos ao mesmo tempo em que procura adiantar em que medida as categorias analíticas ausentes ou insatisfatoriamente abordadas na visão tradicional darão margem a desdobramentos teóricos importantes para o estudo da Economia Industrial e para os fundamentos do Modelo ECD. O primeiro capítulo tem o objetivo de apresentar ao leitor os princípios mais gerais dos modelos básicos de concorrência da teoria econômica neoclássica, a saber, o modelo de concorrência perfeita e o monopólio, e discutir a eficiência alocativa associada a cada um desses modelos. O Capítulo 2 visa a analisar a evolução dos conceitos de empresa, indústria e mercado no âmbito da Economia Industrial. Inicia com uma apresentação da evolução do conceito de empresa – de um objeto estático e reativo para um organismo em crescimento e expansão –, define os conceitos de mercado e indústria e termina mostrando que a unidade de análise dos estudos microeconômicos tem evoluído do atomismo – empresa – para as redes – complexo industrial –, mostrando a crescente interdependência entre os agentes econômicos no processo de produção industrial. No Capítulo 3, examina-se a relação existente entre as estruturas de custos e o fenômeno das economias de escala e escopo, em que medida os comportamentos teóricos dos custos de longo e de curto prazos são importantes para compreender e analisar, na prática, a dinâmica industrial. Finalmente, no Capítulo 4, apresenta-se o Modelo ECD, seus fundamentos, sua evolução e seus limites.

A Parte II (Análise Estrutural dos Mercados) é composta de quatro capítulos e procura sumariar as questões mais consensuais hoje sobre o Modelo ECD quando interpretado em sua primeira versão estruturalista, ou seja, que as condições básicas de oferta e demanda, assim como a estrutura do mercado, afetam decisivamente os comportamentos estratégicos das empresas. O Capítulo 5 define a noção de grau de concentração e procura ensinar quais são os indicadores mais tradicionais para sua quantificação (Razões de Concentração; Índice de Hirschman-Herfindahl e Índice de Entropia de Theil). Seu grande mérito é discutir em que medida esses indicadores podem ou não captar o fenômeno da concentração com exemplos da literatura internacional e exemplos sobre a estrutura industrial brasileira. O Capítulo 6 irá aprofundar a discussão iniciada no Capítulo 1 sobre a inconveniência de se fazer a análise do processo de concorrência mantendo-se a hipótese de homogeneidade dos produtos. No mundo real, dificilmente as mercadorias que competem entre si são homogêneas aos olhos do consumidor, isto é, os consumidores não são indiferentes, por exemplo, quando dois produtos homogêneos são oferecidos associados a duas localizações distintas. Os dois modelos locacionais mais conhecidos da literatura são abordados em versões simplificadas: a cidade linear, desenvolvido por Hotelling, e a cidade circular, elaborado por Salop. No Capítulo 7 iremos apresentar a noção de barreiras à entrada, de uma perspectiva estrutural, também chamada de barreiras à entrada estáticas ou exógenas: são as

barreiras à entrada que decorrem das características técnico-econômicas dos setores e não do comportamento dos agentes. O capítulo apresenta os conceitos de concorrência real e potencial, identifica os principais fatores que geram barreiras à entrada e prossegue apresentando o modelo de preço-limite, baseados em diferenciais absolutos ou relativos de custos entre as empresas estabelecidas e as empresas entrantes. O Capítulo 8 encerra a Parte II apresentando para o leitor o papel da mudança e da inovação no processo concorrencial, abrindo a discussão sobre a hipótese schumpeteriana de endogeneidade da mudança. Discute a relação entre inovação e estrutura industrial na perspectiva dos modelos tradicionais e seus avanços bem como na perspectiva evolucionista. A resenha dos resultados empíricos, observados a partir da agenda estruturalista de pesquisa schumpeteriana, ilustra bem o problema de explicar a intensidade da inovação com base em variáveis de estrutura tais como tamanho das empresas ou grau de concentração industrial.

A **Parte III (Interação estratégica)**, composta de quatro capítulos, é dedicada à chamada Nova Economia Industrial (NEI), que tem como principal identidade o recurso à teoria dos jogos como ferramenta analítica. A teoria dos jogos pode ser definida, em princípio, como conjunto de técnicas de análise de situações de interdependência estratégica. O Capítulo 9 discute como usar a teoria dos jogos como instrumento de identificação, descrição e análise de regras de jogos e de conflitos nas e entre as organizações. Esse capítulo está centrado na elaboração de jogos não cooperativos, sem contemplar jogos com informação incompleta, que são situações nas quais características e regras do jogo não são de conhecimento comum dos jogadores. Dedicamos o Capítulo 10 à apresentação dos modelos de concorrência fundamentados em hipóteses de reação das empresas em oligopólios não cooperativos. Três modelos de competição imperfeita: Cournot (decisão simultânea de quantidades), Bertrand (decisão simultânea de preços) e Stackelberg (decisão sequencial de quantidades) mostram como o equilíbrio entre os agentes é buscado através das funções de reação, um conceito central para o estudo do equilíbrio com interdependência estratégica. Por fim, é analisada mais detalhadamente a atuação das empresas nos diferentes regimes competitivos, e realiza-se uma análise comparativa dos excedentes econômicos gerados em cada regime. O Capítulo 11 trata os principais aspectos que facilitam ou prejudicam a coordenação oligopolística, isto é, a adoção de acordos tácitos ou explícitos entre as empresas visando a evitar os efeitos negativos da concorrência sobre os lucros. A seguir, são examinadas algumas formas de superar esses problemas por meio da adoção de regras de bolso pelos agentes, com destaque para o princípio do custo total e a liderança de preços. Finalmente, o Capítulo 12 retoma a questão das barreiras à entrada associadas às estratégias das empresas na criação de custos irrecuperáveis (irreversibilidades) como importante ferramenta estratégica à disposição das empresas em processos de competição com rivalidade. Ao acentuar a importância das estratégias empresariais na prevenção da entrada e nos processos competitivos posteriores, essa teoria contrapõe-se àquela apresentada no Capítulo 7, baseada na noção de preço-limite, que enfoca a barreira à entrada em um ângulo eminentemente estrutural. A partir de uma crítica ao postulado de Sylos-Labini são apresentados tanto os modelos de A. Dixit quanto o de P. Milgrom e J. Roberts.

A **Parte IV (A Grande Empresa Contemporânea)**, formada por cinco capítulos, tem por foco a análise institucional da empresa, o estudo de sua natureza como enfatizou Ronald Coase. Toma como hipótese que “as instituições importam” na análise dos processos econômicos; que existe uma lógica de eficiência econômica por trás dos determinantes dos tipos de empresas e formas de organização predominantes. O Capítulo 13 introduz o importante conceito de custos de transação e suas principais aplicações na análise da integração vertical e na regulação econômica. Discute sob que condições os custos de transação deixam de ser desprezíveis e passam a ser um elemento importante nas decisões dos agentes econômicos, contribuindo para determinar a forma pela qual são alocados os recursos na economia. O Capítulo 14 irá apresentar o tipo específico de organização mais estilizado na literatura de Economia Industrial – a grande corporação – e as principais abordagens disponíveis na literatura para o seu estudo. São destacadas três abordagens explicativas sobre a organização interna da grande empresa. A primeira delas é de origem tradicional, mas procura estender o alcance do instrumental teórico da microeconomia tradicional para a análise das formas específicas de organização da empresa, partindo de uma redefinição conceitual desta última, que passa a ser considerada como uma rede de contratos. A segunda abordagem é aquela oferecida pela teoria dos custos de transação, apresentada no Capítulo anterior e que, neste Capítulo, foi aplicada ao estudo dos processos de mudança da organização interna das empresas. Por último, é apresentada uma abordagem alternativa que enfatiza o papel da estrutura organizacional da empresa na mobilização dos conhecimentos necessários ao aprendizado tecnológico. A principal origem dessa abordagem é o programa de pesquisa neo-schumpeteriano, embora influências de ideias geradas fora da Economia Industrial – basicamente de teorias das organizações de natureza mais sociológica – também se façam presente. O Capítulo 15 aborda as conexões entre diversificação e crescimento da empresa, enfatizando a discussão das possíveis direções que o processo de diversificação pode tomar – analisadas em termos do grau de similaridade com as atividades originais da empresa – bem como dos condicionantes internos e externos à empresa que interferem na dinâmica do processo. Apresenta o termo “competências essenciais” e discute as diversas formas de diversificação empresarial. O Capítulo 16 aborda a ocorrência de múltiplas formas de cooperação produtiva e tecnológica entre empresas. Aborda-se, principalmente, um tipo particular de estrutura em rede, as “redes de empresas”, procurando discutir

algumas características e propriedades desses arranjos que condicionam a sua capacidade de resposta diante dos estímulos ambientais. Sua principal contribuição é permitir ao leitor uma leitura sistematizada dos distintos exemplos apontados pela literatura de Economia Industrial acerca dos distritos industriais, *clusters*, arranjos cooperativos, entre outras redes. O Capítulo 17 encerra esta Parte, exemplificando mais uma importante forma institucional que a grande corporação pode adquirir: a Empresa Transnacional. De fato, esse tipo de organização se sobressai na atualidade como uma das mais importantes formas de organização, principal *locus* de acumulação e de poder econômico, exercido a partir do seu controle sobre ativos específicos de capital, tecnologia e de capacitações gerencial, organizacional e mercadológica.

A **Parte V (Estratégias Empresariais)**, formada por quatro capítulos, apresenta a teoria da concorrência esboçada em suas linhas básicas por J. Schumpeter na primeira metade do século XX, e aperfeiçoada nas últimas duas décadas por autores da corrente neo-schumpeteriana, em contraste com a análise estática da corrente tradicional. O Capítulo 18 inaugura essa parte mostrando a concorrência na economia capitalista como um processo evolutivo e, portanto, dinâmico, gerado por fatores endógenos ao sistema econômico, notadamente as inovações que emergem incessantemente da busca de novas oportunidades lucrativas por parte das empresas em sua interação competitiva. O capítulo encerra-se com a identificação de algumas implicações normativas e de política econômica – especialmente para a política de concorrência – dessas proposições teóricas. O Capítulo 19 apresenta um detalhamento das estratégias de inovação das empresas, sua organização interna e externa na aquisição, produção e difusão das inovações. A empresa é concebida como um organismo vivo em permanente mutação que recebe influências de seu ambiente (mercado) mas, ao mesmo tempo, é capaz de transformá-lo ou criar novos mercados ou indústrias a partir da introdução de inovações tecnológicas. Essa visão contrasta com a visão exposta no Capítulo 8, onde a propensão a inovar era tomada como função do padrão de concorrência existente no mercado em que a empresa estava situada. Outro elemento de grande relevância na estratégia competitiva das empresas é a propaganda. O Capítulo 20 analisa a influência da propaganda sobre o processo de concorrência, avaliando também seus efeitos no desempenho das empresas e sobre a estrutura da indústria. A propaganda é vista na análise tradicional como instrumento tanto de informação quanto de persuasão. Já para a corrente alternativa, ela pode introduzir maior qualidade e trazer também irreversibilidades ao mercado, criando ativos específicos, tal como a marca. O Capítulo 21 retoma a questão institucionalista de que as instituições evoluem *pari e passu* com a estrutura produtiva sob a ótica financeira. Procura responder de que forma as condições de financiamento e a estrutura de financiamento das empresas influenciam as decisões de investimento das mesmas. Especial ênfase é dedicada à questão das finanças corporativas e do *project finance*.

A **Parte VI (Políticas e Regulação dos Mercados)** trata da política econômica, tradicionalmente a principal aplicação da Economia Industrial. Aqui o leitor irá encontrar uma interessante discussão sobre os princípios que devem nortear a intervenção do Estado nos mercados e sobre a institucionalidade específica da economia brasileira. A principal lição que emerge de sua leitura é que o Estado não só tem um papel importante a desempenhar nas políticas de competitividade, com ênfase na inovação, mas também na regulação e defesa da concorrência. O Capítulo 22 analisa a política de defesa da concorrência na garantia da existência de condições de competição, preservando ou estimulando a formação de ambientes competitivos com intenção de induzir, se possível, maior eficiência econômica como resultado do funcionamento dos mercados. É uma excelente aplicação prática do Modelo ECD e do conceito de mercado relevante para a ação antitruste. Também compara a legislação brasileira com a legislação norte-americana, principal jurisprudência disponível sobre o assunto. O Capítulo 23, dedicado às políticas de regulação econômica, inicia-se apresentando o conceito do monopólio natural em duas circunstâncias distintas: monopólio natural com um produto e monopólio natural multiproduto, seguida de uma discussão sobre o fenômeno do monopólio natural nas indústrias de rede. Apresenta ainda as várias formas de regulação dos preços e, finalmente, uma visão histórica da regulação em outros países antes e depois do período de 1980, seguida de uma descrição sobre regulação no Brasil. O Capítulo 24 está organizado para permitir ao leitor identificar lógicas distintas da teoria econômica no que diz respeito ao papel do Estado: teorias que associam política industrial às falhas de mercado, ao desenvolvimento de nações e à evolução das competências de agentes econômicos. Analisa também a relação entre política industrial e outras políticas de Estado tais como a política macroeconômica, de comércio exterior, regulação de infraestruturas e de ciência e tecnologia, finalizando com o relato da experiência brasileira e da prática recente dos países da Organização para Cooperação e Desenvolvimento Econômico (OCDE). O Capítulo 25 apresenta os argumentos a favor e contra a adoção de políticas comerciais enfatizando seus efeitos para o desempenho das indústrias. Fechando essa parte, o Capítulo 26 discute as principais razões para a adoção de uma política ambiental, as soluções econômicas para esses problemas ambientais e uma pesquisa sobre a legislação ambiental brasileira e sua institucionalidade.

Finalmente, a **Parte VII (Guia para Análises Empíricas)**, composta de dois capítulos, tem a função de permitir que o leitor possa apreender os conceitos e ao mesmo tempo acessar as fontes (estudos e estatísticas) necessárias para desenvolver estudos empíricos aplicando os principais conceitos de Economia Industrial. O Capítulo 27 sugere como organizar os dados



categóricos utilizando-se as tabelas de contingência e algumas medidas de associação de variáveis categóricas. Sugere ainda algumas variáveis quantitativas, e a análise de correlações. As duas últimas seções são dedicadas aos modelos de análise estatística multivariada. São apresentados os modelos de Análise de Componentes Principais e Análise de Grupamento (*Cluster Analysis*). O Capítulo 28, ao apresentar as principais fontes de informação e suas metodologias sobre a indústria brasileira, é leitura obrigatória para os interessados em realmente praticar a Economia Industrial.

Esperamos que o leitor esteja suficientemente motivado para enfrentar a dificuldade do exercício da arte da Economia Industrial, para a qual não basta o conhecimento da lógica e dos instrumentos matemáticos; é preciso conhecer também as institucionalidades e particularidades dos mercados estudados.



PART

I

CONCEITOS BÁSICOS

Modelos Tradicionais de Concorrência

Luiz Martins de Melo

1.1 Introdução

A Competição Perfeita e o Monopólio são os dois modelos básicos de concorrência da teoria neoclássica. Eles representam, respectivamente, os extremos de atomização e concentração da produção. Concentração da produção significa que a empresa tem poder de mercado. Atomização significa ausência de poder de mercado; a empresa é tomadora de preços no mercado.

Esse modelo, até hoje dominante na lógica neoclássica de pensar as estruturas de mercado, começaram a sofrer uma contestação formal a partir do artigo “As leis dos rendimentos sob condições de concorrência” de Piero Sraffa, de 1926.¹

Esse artigo forneceu a inspiração teórica para Joan Robinson,² na Inglaterra, publicar *The Economics of Imperfect Competition*, em 1933. Nos Estados Unidos, no mesmo ano, Edward Chamberlin³ publicava *The Theory of Monopolistic Competition*. Nos dois livros os autores procuravam formular modelos de competição imperfeita, que apontavam para o comportamento monopolista como o padrão geral de competição nos mercados.

As hipóteses gerais que fundamentam o comportamento do agente neoclássico são a maximização de lucros e a informação completa ou perfeita. Isso pressupõe uma racionalidade plena na tomada de decisão pelo agente, conhecedor em última instância de todas as oportunidades presentes no mercado, qualquer que este seja.

Na primeira parte deste capítulo analisaremos os pontos principais do modelo da competição perfeita. A seguir discutiremos o monopólio e o comportamento monopolista de discriminação de preços. Na terceira parte será apresentada a competição imperfeita, sob o aspecto da diferenciação de produto. Na última parte, serão feitas algumas observações sobre os modelos, à guisa de conclusão.

1.2 O Modelo de Competição Perfeita

1.2.1 As HIPÓTESES

A estrutura de mercado definida como competição perfeita não prevê qualquer tipo de coordenação entre as empresas que atuam nesse mercado. As empresas tomam decisões de forma descentralizada, apenas sujeitas à disciplina do mercado, isto é, são tomadoras de preço.

A indústria é definida como um grupo de empresas que geram um produto homogêneo. As características do produto e dos serviços associados com a sua venda são as mesmas para todas as empresas. As hipóteses básicas do modelo de concorrência perfeita estão listadas no Quadro 1.1. Qualquer alteração dessas hipóteses produz um desequilíbrio entre oferta e demanda. A correção desse desequilíbrio é realizada pelas próprias forças do mercado perfeito, fazendo com que as quantidades voltem a se ajustar aos preços.

QUADRO 1.1 AS HIPÓTESES BÁSICAS DO MODELO DE COMPETIÇÃO PERFEITA

- H1: grande número de empresas
- H2: produto homogêneo
- H3: livre entrada e saída de empresas
- H4: maximização de lucros
- H5: livre circulação da informação
- H6: perfeita mobilidade dos fatores

Esse mercado inclui um grande número de empresas. Essas empresas podem ser grandes, mas não podem ter poder de mercado. Para isso têm de oferecer uma pequena parte da quantidade total vendida no mercado. Dessa forma, o tamanho absoluto da empresa deixa de ter relevância. O que interessa é o seu tamanho relativo; a parcela do mercado que está sob sua influência. A mesma observação se refere aos compradores: nenhum comprador pode concentrar uma parte significativa da demanda total do mercado, ou seja, não há poder monopsônico.

Nessas condições, produto homogêneo e ausência de poder de mercado por parte das empresas, cada uma das empresas é tomadora de preço. A sua curva de demanda é infinitamente elástica e cada empresa pode vender qualquer quantidade do produto ao preço do mercado. A curva de demanda do mercado da empresa individual é também a sua curva de receita média e de receita marginal.

Não existem barreiras à entrada ou à saída de empresas da indústria ou mercado. Se existirem barreiras ao livre movimento das empresas, o número de empresas pode diminuir, e cada uma delas pode adquirir poder para afetar o preço no mercado. Essa hipótese complementa a anterior referente ao grande número de empresas na indústria.

O objetivo das empresas é a maximização do lucro. O lucro, nessa concepção, é a remuneração do capital acima da taxa normal de mercado, dada pelo custo de oportunidade do investimento e a remuneração para a função de risco do empresário. O lucro é definido como a diferença entre a receita total (RT) e o custo total (CT). Se o lucro for nulo, isto é, $RT = CT$, isso quer dizer que a taxa normal de lucro está sendo obtida, uma vez que os custos de oportunidade do capital estão incluídos nos itens de custo das empresas. Se o lucro for positivo, isto é, $RT > CT$, a empresa terá lucros extraordinários ou renda econômica. Nesse caso, haverá entrada de empresas no mercado até o ponto em que o lucro volte a ser zero ($RT = CT$).

Os fatores de produção são livres para se moverem de uma empresa para outra por toda a economia. Do mesmo modo, os trabalhadores estão livres para trocarem de trabalho. Isto significa que as habilidades do trabalhador podem ser adquiridas facilmente, sem custo de aprendizado. As matérias-primas e outros fatores de produção não são monopolizados por nenhuma empresa, e a força de trabalho não é sindicalizada. Existe competição perfeita em todos os mercados de fatores de produção.

Todos os compradores e vendedores possuem perfeito conhecimento das condições do mercado. Esse conhecimento não se refere somente às condições correntes de produção, mas também às condições futuras. A informação é livre e sem custo. Nessas condições não existe incerteza sobre o comportamento futuro dos mercados.

Nesse momento não se considera a intervenção governamental no mercado. Tarifas, subsídios, fixação de salário e regulamentação institucional ou legal de qualquer atividade estão excluídos do modelo.

1.2.2 EQUILÍBRIO

O mercado está em equilíbrio quando nenhuma empresa modifica seus planos de produção, isto é, quando todas as suas empresas estão em equilíbrio. Por sua vez, a empresa está em equilíbrio quando produz a quantidade que maximiza o lucro, isto é, a diferença entre a receita total e o custo total. Para determinar o equilíbrio do mercado é necessário derivar a curva de demanda e a oferta do mercado. Cada uma delas é o somatório das curvas de demanda e oferta das empresas individuais.

Dadas as hipóteses apresentadas na seção anterior, em concorrência perfeita, a curva de demanda da empresa é horizontal ao preço de mercado. Isso significa que a preços maiores que o de mercado, a empresa não vende coisa alguma e, portanto, nenhuma tentativa de elevar os preços poderá ser bem-sucedida. A preços menores, embora ela se depare com toda a demanda de mercado, a limitação de quantidade que ela pode produzir também não proporcionará um aumento de receita que justifique essa escolha. Logo, a empresa é tomadora de preço.

Para que se determine o equilíbrio do mercado, é necessário determinar a curva de oferta da indústria. Isso requer a determinação da oferta individual de cada empresa, dado que a oferta do mercado é o somatório da oferta de todas as empresas na indústria.

1.2.2.1 O curto prazo

As funções de custo total (CT) e custo médio (CMe)⁴ refletem as diferentes condições de produção que vigoram no curto e no longo prazo. No curto prazo, definido pela presença de pelo menos um fator de produção fixo, a função de produção da empresa reflete as condições de operação da lei das proporções variáveis. Essa lei garante que existe um nível de produto além do qual a função de produção opera sob o impacto de retornos decrescentes do(s) fator(es) variável(is). Isto implica que o CMe passa por um ponto de mínimo.

Observe o Quadro 1.2.

O custo marginal (CMg) mede a taxa de variação dos custos quando aumentamos a produção em uma unidade. A relação que se estabelece entre CMg e CMe é que o primeiro deve estar abaixo do segundo, quando este for decrescente. Isso decorre do fato de que se a média for decrescente, os custos de cada unidade adicional produzida terão que ser menores que a média até aquele ponto. Se o CMe estiver crescendo, o CMg terá que ser maior do que o CMe, e “puxar” o CMe para cima. Portanto, a condição para o CMe mínimo é que ele iguale o CMg.

Para alcançar a posição de equilíbrio, a empresa tem que estar produzindo a quantidade na qual a receita marginal (RMg) é igual ao CMg. Mas esta condição não basta. Pode haver mais de um ponto na função lucro em que esta condição é satisfeita. Para satisfazer a condição de maximização, o CMg tem que ser crescente quando se iguala à RMg.

A função oferta de curto prazo é definida para alguma quantidade do fator de produção fixo. Ela inicialmente corresponde ao ramo ascendente da curva de CMg. No curto prazo, as empresas podem estar em uma posição de equilíbrio com lucros extraordinários ou prejuízos. Mas elas não irão operar se o preço for inferior ao custo variável médio mínimo. Assim, a curva de oferta de curto prazo é a parte ascendente da curva de CMg que está localizada acima da curva de custo variável médio.

1.2.2.2 O longo prazo

No longo prazo, quando nenhum fator de produção tem quantidade fixada, a empresa pode ajustar sua capacidade de produção para produzir no ponto de mínimo da função de CMe. Esse ponto será igual ao preço, determinado pela demanda do mercado. No longo prazo as empresas estarão ganhando apenas lucros normais ($\pi=0$). Se elas estiverem ganhando lucros extraordinários, novas empresas entrarão no mercado, fazendo com que o preço caia e o custo médio de produção aumente devido ao aumento do preço dos fatores de produção na medida em que a indústria se expande. Esse ajustamento seguirá até o ponto em que o preço se iguala ao CMe mínimo. Se as empresas estiverem operando com prejuízo no longo prazo, elas deixarão a indústria. A quantidade ofertada no mercado cairá e o preço de mercado aumentará até o ponto em que novamente os lucros se tornem normais ($p=CMe$ mínimo).

QUADRO 1.2 MAXIMIZAÇÃO E EQUILÍBRIO: CURTO PRAZO

Função de produção: $y=f(x_1, k)$

x_1 =quantidade do fator de produção 1

k =fator de produção fixo

w_1 =preço do fator de produção 1

w_k =preço do fator de produção k

y =quantidade produzida

p =preço de mercado

Receita média (RMe)= RT/y ; receita marginal (RMg)= $\partial RT/\partial y$

Custo fixo= $k \cdot w_2 = K$; custo fixo médio= K/y

Custo variável= $w_1 \cdot x_1$; custo variável médio= $w_1 \cdot x_1/y$

Custo médio (CMe)= $CT/y=(w_1 \cdot x_1 + K)/y$

Custo marginal (CMg)= $\partial CT/\partial y=\partial(w_1 \cdot x_1)/\partial y$

Equilíbrio: Maximização de Lucro

Lucro (π)= $RT - CT = p \cdot y - (w_1 \cdot x_1 + K)$

Condição de primeira ordem: $RMg = CMg$

$RMg = \partial RT/\partial y = \partial p \cdot y / \partial y = p$ (o preço é independente da quantidade da empresa)

$CMg = \partial CT/\partial y$

$\Rightarrow p = CMg$

Condição de segunda ordem: $\partial^2 RT/\partial y^2 < \partial^2 CT/\partial y^2$

$\Rightarrow \partial^2 CT/\partial y^2 > 0$

A função oferta de longo prazo da empresa mede a produção ótima, quando ela pode ajustar livremente a quantidade de fatores de produção que vai usar. A diferença entre o equilíbrio de curto prazo e o de longo prazo reside no processo de ajustamento. A função oferta de longo prazo envolve o CMg de produção quando o fator de produção é ajustado para sua utilização ótima. Os custos marginais de curto e longo prazos igualar-se-ão quando a escolha do fator de produção fixo associada ao CMg de curto prazo for a escolha ótima de longo prazo. Desse modo, as ofertas de curto e longo prazos da empresa se igualam.

O lucro que a empresa realiza no longo prazo tem que ser pelo menos zero, ou ela sai do negócio. Isso significa que o preço tem que ser no mínimo igual ao CMe. A parte relevante da função oferta de longo prazo é aquela que se situa acima do custo médio de longo prazo. Isto equivale à condição de curto prazo de ter o preço acima do custo variável médio.

Veja o Quadro 1.3.

No longo prazo a empresa tem mais escolhas para promover ajustes, quando preço e mercado variam. Em consequência, a função oferta de longo prazo será mais elástica e terá maior sensibilidade às variações nos preços do que a função oferta de curto prazo.

1.2.3 A ALOCAÇÃO ÓTIMA DE RECURSOS

O funcionamento do mercado de competição perfeita conduz para a alocação ótima de recursos. Esse estado é atingido no longo prazo quando prevalecem as seguintes condições:

1. A quantidade de produto é realizada no nível de custo médio mínimo.
2. Os consumidores pagam o preço mínimo, igual ao custo médio mínimo, que cobre o custo marginal do produto, isto é, preço igual ao custo de oportunidade.
3. As plantas estão operando com plena capacidade no longo prazo, portanto não existe desperdício de recursos.
4. As empresas ganham lucros normais.

No longo prazo, essas condições ocorrem em todos os mercados e os recursos estão otimamente alocados na economia como um todo. Se por simplicidade presume-se que apenas duas mercadorias são produzidas na economia, a curva de possibilidade de produção representa a dotação dada de recursos da economia. As preferências dos consumidores na economia podem ser representadas pela curva de indiferença da sociedade. Dada a curva de possibilidade de produção e as preferências dos consumidores, a competição perfeita levará para a alocação ótima de recursos sob as condições descritas a seguir.

1. Se a soberania dos consumidores, expressa pelo sistema de preços, funcionando sem qualquer intervenção governamental, refletir o correto posicionamento das preferências dos consumidores.
2. Se não existem economias de escala em nenhuma indústria.⁵
3. Se os recursos e a tecnologia são dados, não existem crescimento e progresso técnico na economia.

Nessas condições a economia está usando todos os recursos disponíveis e os consumidores alcançam o mais alto bem-estar possível.

QUADRO 1.3 MAXIMIZAÇÃO E EQUILÍBRIO: LONGO PRAZO

Função de Produção: $y = f(x_1, x_2)$

x_1 =quantidade do fator de produção 1

x_2 =quantidade do fator de produção 2

w_1 =preço do fator de produção 1

w_2 =preço do fator de produção 2

y =quantidade produzida

p =preço de mercado

Lucro (π)= $RT - CT = py - (w_1 \cdot x_1 + w_2 \cdot x_2)$.

Receita Média (RMe)= RT/y ; Receita Marginal (RMg)= $\partial RT/\partial y$

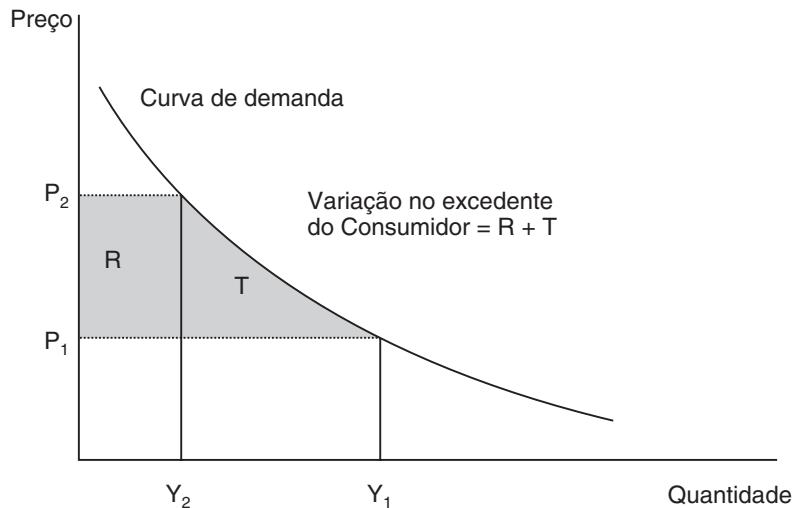
Custo Médio (CMe)= $CT/y = (w_1 \cdot x_1 + w_2 \cdot x_2)/y$

Custo Marginal (CMg)= $\partial CT/\partial y$

Equilíbrio: Maximização de Lucro

Condição de Primeira Ordem: $RMg=RMe=p=CMg=CMe$ mínimo.

Condição de Segunda Ordem: $\partial^2 RT/\partial y^2 < \partial^2 CT/\partial y^2$



Variação no excedente do consumidor: A variação no excedente do consumidor será a soma de $R + T$, que corresponde à área abaixo da curva de demanda.

Gráfico 1.1 O Excedente do Consumidor.

1.2.4 O EXCEDENTE DO CONSUMIDOR E DO PRODUTOR

O conceito de excedente do consumidor individual visa a medir o benefício do consumidor em adquirir certa quantidade de um bem, menos a redução do consumo nos outros bens. Em geral, não estamos muito interessados em medir o nível absoluto do excedente do consumidor. O importante é analisar a variação do excedente do consumidor em consequência de alguma variação do preço de um bem.

Suponhamos que o preço de um bem varie de p_1 para p_2 , sendo $p_2 > p_1$. A variação do excedente do consumidor é a área sombreada no Gráfico 1.1. Essa área pode ser subdividida em duas: a região T e a região R . A área retangular R mede a perda do consumidor resultante do aumento de preços do bem. Ele agora paga mais $(p_2 - p_1)$, por cada unidade do bem que ele consome. Para consumir y_2 unidades do bem, ele gasta mais $(p_2 - p_1)y_2$ do que antes.

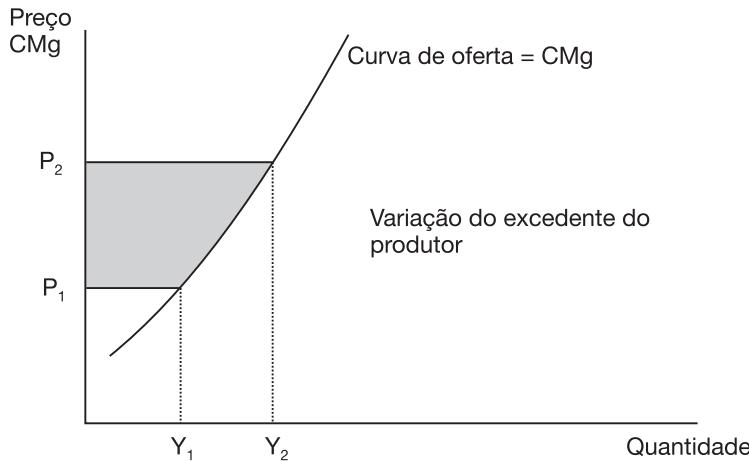
Porém, isto não representa a perda total de bem-estar do consumidor. O aumento do preço do bem faz com que ele consuma menos do que anteriormente. A área T mede o valor perdido do consumo do bem. A área $R + T$ representa a perda total de bem-estar do consumidor: R mede a perda pelo gasto maior para consumir as unidades do bem e T mede a perda pela diminuição do consumo. Portanto, a área abaixo da curva de demanda mede o excedente do consumidor.

O excedente do produtor é definido de forma análoga ao excedente do consumidor. A área acima da curva de oferta é denominada excedente ao produtor. Esse está intimamente relacionado aos lucros da empresa. Sabemos que a curva de CMg é a curva de oferta da empresa. No Gráfico 1.2, mostramos o excedente do produtor quando a empresa passa de uma produção y_1 para y_2 , sendo $y_2 > y_1$. Aqui, também, estamos interessados na variação do excedente do produtor, mais do que na sua quantidade total. A mudança do excedente do produtor ao passar de y_1 para y_2 é a mudança nos lucros ($RT - CT$) para a mesma variação de quantidade produzida. Isto é devido ao fato de que o custo marginal é o determinante para medir o impacto nos lucros de uma variação na produção. No curto prazo os custos fixos não variam, por definição, e no longo prazo todos os custos são variáveis. Logo, não precisamos nos preocupar com o CMe, para medir o excedente do produtor.

1.3 Monopólio

1.3.1 AS CAUSAS DO MONOPÓLIO

O monopólio é a estrutura em que há apenas um produtor no mercado. As causas da existência do monopólio são várias; algumas políticas, outras econômicas e outras técnicas. As principais causas apontadas pela teoria econômica neoclássica são as descritas a seguir:



Variação no excedente do produtor: Como a curva de oferta coincide com a parte de inclinação ascendente da curva de custo marginal, a variação no excedente do produtor será a área à esquerda da curva de oferta.

Gráfico 1.2 O Excedente do Produtor.

1. Propriedade exclusiva de matérias-primas ou de técnicas de produção.
2. Patentes sobre produtos ou processos de produção.
3. Licença governamental ou imposição de barreiras comerciais para excluir competidores, especialmente estrangeiros.
4. O caso do monopólio natural quando o mercado não suporta mais do que uma única empresa, pois a tecnologia de produção impõe que a operação eficiente tenha economias de escala substanciais. Observe o Quadro 1.4.

1.3.2 O EQUILÍBRIO NO MONOPÓLIO

O modelo teórico do monopólio supõe uma única empresa que domina o mercado. A demanda da empresa é a demanda da indústria. Os consumidores não têm outra alternativa senão comprar do monopolista. O monopolista tem o poder de impor o preço aos consumidores. Esse poder de mercado⁶ permite que o monopolista opere sempre com lucros extraordinários, isto é, vai impor uma margem (*mark-up*) sobre os custos marginais apresentado a seguir:

$$p = CMg / (1 - 1 / |\varepsilon_d|)$$

onde:

p =preço de mercado

ε_d =elasticidade-preço da demanda.

Esta fórmula de fixação de preços pelo monopolista resulta da condição de maximização de lucros do monopolista:

$$\begin{aligned} RT &= py \\ RMg &= \partial RT / \partial y = p(\partial y / \partial y) + y(\partial p / \partial y) \\ RMg &= p + y(\partial p / \partial y) \end{aligned}$$

Multiplicando o segundo termo por (p/p) e colocando p em evidência temos:

$$RMg = p(1 + 1 / \varepsilon_d) = p(1 - 1 / |\varepsilon_d|)$$

QUADRO 1.4 HIPÓTESES BÁSICAS DO MODELO DE MONOPÓLIO

H: um único produtor

H2: produto sem substitutos próximos

H3: barreiras à entrada

H4: maximização de lucros

No equilíbrio temos:

$$\begin{aligned} RMg &= CMg; \\ p(1 - 1/|\varepsilon_d|) &= CMg \\ p &= CMg/(1 - 1/|\varepsilon_d|) \end{aligned}$$

O monopolista, pela condição de maximização descrita anteriormente, só vai operar na região onde a demanda for elástica, para que tenha garantido $p > CMg$. A condição de segunda ordem do monopolista é a mesma de competição perfeita, $\partial^2 RT / \partial y^2 < \partial^2 CT / \partial y^2$. O Gráfico 1.3 mostra o equilíbrio do monopolista no caso da demanda linear.

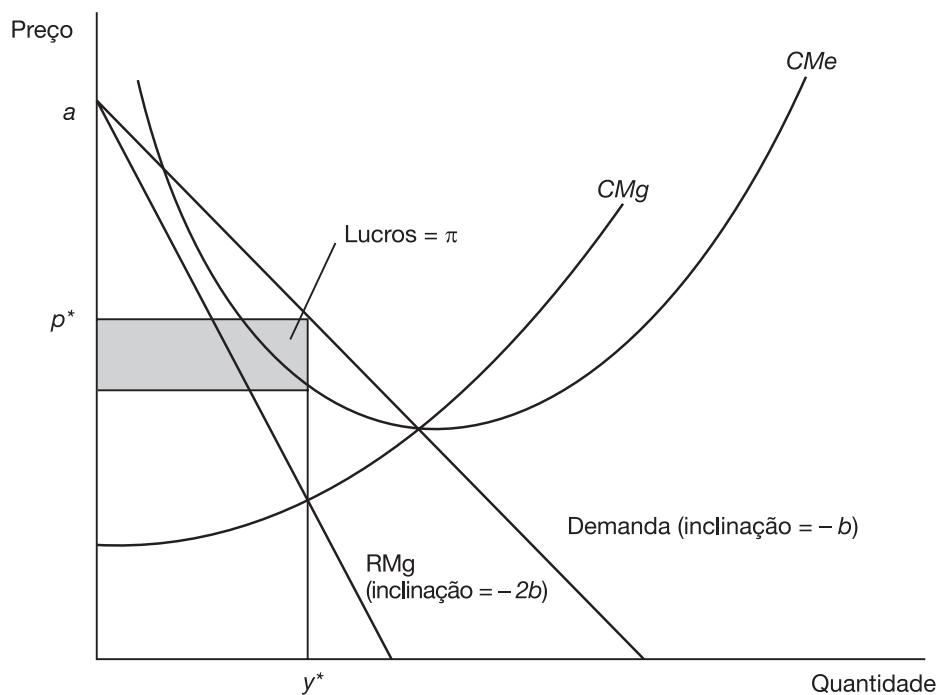
O monopolista pode expandir ou usar a capacidade existente da planta em qualquer nível para maximizar seu lucro. A entrada bloqueada torna desnecessário que o monopolista trabalhe no nível da escala ótima da planta. Não existe qualquer condição, técnica ou de mercado, que o obrigue a isso. O tamanho da planta e o grau de sua utilização dependem inteiramente da demanda de mercado.

O monopolista poderá operar com grau de utilização ótimo (CMe mínimo), subótimo e acima do ótimo. Operar em condições subótimas significa ter excesso de capacidade. O mercado, muito pequeno para o tamanho da planta, não permite que ele expanda a produção até o ponto de CMe mínimo.

Se o mercado for grande demais para sua planta, ele encara a decisão de construir uma nova planta ou trabalhar acima da capacidade existente. A sua decisão vai depender de quão grande é o mercado, isto é, qual o tamanho da diferença entre a demanda e o ponto ótimo da sua planta em operação. Caso opere com a planta atual, ele vai incorrer em custos mais altos, por duas razões: a quantidade produzida é maior que o ótimo e a empresa está operando acima da capacidade. Mesmo nessas condições ele auferirá lucros extraordinários.

O monopolista operará no tamanho ótimo (CMe mínimo) se o tamanho do mercado for suficientemente grande para permitir que este nível de produção seja aquele de capacidade plena.

Em resumo, não existem forças de concorrência que obriguem o monopolista a operar no ponto ótimo. Na hipótese de que ele o faça, nada garante que ele abra mão do lucro extraordinário. Assim, o preço cobrado pelo monopolista será sempre maior do que em competição perfeita e a quantidade vendida, menor.



Monopólio com uma curva de demanda linear: A produção que maximiza os lucros do monopolista ocorre onde a receita marginal se iguala ao custo marginal ($RMg = CMg$).

Gráfico 1.3 O Equilíbrio do Monopolista.

1.3.3 A INEFICIÊNCIA DO MONOPÓLIO

As empresas na indústria que operam em competição perfeita maximizam o lucro no ponto em que o preço é igual ao custo marginal ($p = RMg = CMg$). O monopólio maximiza o lucro quando a receita marginal é igual ao custo marginal ($RMg = CMg$). O preço será mais alto e a produção será menor no monopólio do que em competição perfeita. Portanto, o consumidor estará sempre em condições de bem-estar inferiores àquelas da competição perfeita. É por isso que o modelo de competição perfeita é considerado o padrão de comparação para aferir a eficiência dos outros modelos competitivos.

Porém, poderíamos argumentar que a empresa estaria melhor no monopólio do que em competição perfeita. Suponhamos que possamos fazer com que um monopólio se comporte competitivamente. Isto é, o preço de mercado seria definido exogenamente. Vamos definir um equilíbrio econômico de Pareto, como aquele em que não existe qualquer forma de melhorar a situação de um agente econômico sem piorar a de um outro.

Pela definição acima sabemos que no equilíbrio de Pareto não haverá qualquer agente econômico disposto a pagar mais por um bem do que custa para produzi-lo ($p = RMg = CMg$). Porém, no monopólio temos $p(1 - 1/|\varepsilon_d|) = RMg = CMg$. O que supõe que os consumidores estariam dispostos a pagar mais por uma unidade do bem do que custa para produzi-la. É óbvio que existe potencial para uma melhoria de Pareto aqui. Esse potencial compreende toda região entre o preço de monopólio $p_m = CMg/(1 - 1/|\varepsilon_d|)$ e o preço de concorrência $p_c = CMg$.

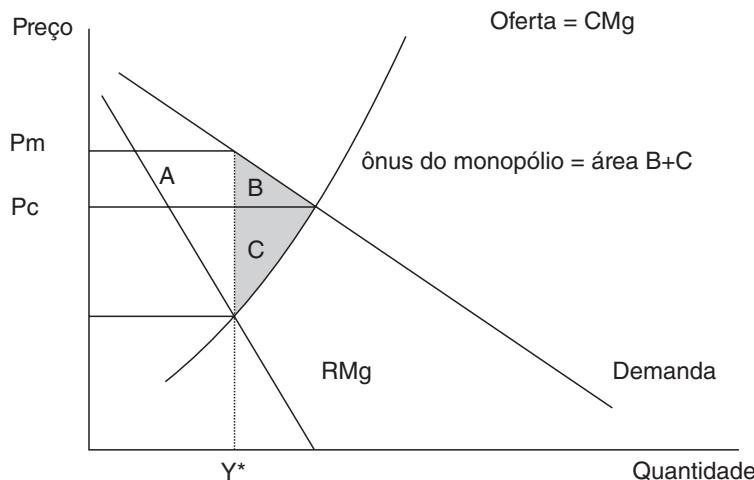
Essa diferença entre p_m e p_c é considerada a ineficiência do monopólio. A produção é considerada eficiente quando o consumidor paga por uma unidade extra do produto exatamente o custo de produzi-la. O monopolista estaria disposto a vender uma unidade adicional a um preço mais baixo, se não fosse preciso diminuir o preço de todas as unidades que estivessem à venda. Para isso ele vai adotar uma política de discriminação de preços que analisaremos na próxima seção deste capítulo.

O ônus do monopólio é uma forma de medir a sua ineficiência. Nós já analisamos como medir o excedente do consumidor e do produtor. Para medirmos o ônus do monopólio vamos considerar a empresa e os consumidores da sua produção como agentes simétricos. Para isso vamos somar os lucros da empresa ao excedente do consumidor. A variação no excedente do produtor – mudança nos lucros da empresa – mede quanto a empresa estaria disposta a pagar para obter o preço mais alto em condições de monopólio. A variação no excedente do consumidor mede quanto os consumidores deveriam receber para compensar as suas perdas com o preço mais alto. A diferença entre o excedente do produtor e do consumidor nos diz qual foi o benefício líquido ou o custo do monopólio.

O Gráfico 1.4 mostra as variações no excedente do produtor e do consumidor em consequência da mudança da produção sob monopólio para competição perfeita. Podemos notar três regiões no gráfico. A região A representa a diminuição do excedente do monopolista pelo preço mais baixo nas unidades que ele já vendia. A região C representa o ganho do monopolista devido às unidades extras que ele vende, devido ao preço menor.

Simetricamente, o excedente do consumidor aumenta em A, porque agora eles compram toda a quantidade do bem a um preço menor. A região B representa o excedente que os consumidores ganham com as unidades extras que são vendidas.

A área A representa a transferência de bem-estar do monopolista para o consumidor: um lado ganha, o outro perde, mas o excedente permanece o mesmo. A área B+C expressa o aumento no excedente: o valor que os produtores e consumidores atribuem para a produção extra que entra no mercado em função da diminuição do preço.



O ônus do monopólio: O ônus do monopólio é dado pela área B + C.

Gráfico 1.4 O Ônus do Monopólio.

Esta área, B+C, é o ônus devido ao monopólio. Mede a perda de bem-estar em função do estabelecimento de preço do monopólio em vez do preço competitivo. O ônus do monopólio pode ser definido como o valor da produção de cada unidade perdida ao preço que as pessoas estariam dispostas a pagar por elas. O valor de cada unidade extra (marginal) de produção é o preço de mercado. O custo de produzir cada unidade adicional é o CMg. O valor social de produzir uma unidade extra é a diferença entre o preço e o CMg. Ao nos movermos do nível de produção do monopólio para o de produção competitiva, somamos a distância entre as curvas de demanda e de CMg, para termos o valor da produção perdida devido ao monopólio. A área total formada pelo triângulo B+C é o ônus do monopólio.

1.3.4 DISCRIMINAÇÃO DE PREÇOS

O poder do monopolista permite que ele tenha uma política de discriminação de preços voltada para extrair o máximo possível de excedente do consumidor e para aumentar a sua receita total.

Para que haja discriminação de preços o mesmo produto tem que ser vendido a diferentes preços para diferentes compradores. O custo de produção é o do monopolista, isto é, o mesmo para todos os produtos vendidos.⁷ O produto é idêntico, os consumidores não percebem qualquer diferença nos produtos vendidos.

A discriminação de preços vai depender da renda dos consumidores, das suas preferências, da localização e da facilidade de encontrar substitutos para o produto. O monopolista vai procurar segmentar a sua curva de demanda, em diferentes elasticidades, para criar mercados distintos para o seu produto.

As condições necessárias para que a política de discriminação de preços do monopolista seja bem-sucedida são as descritas a seguir.

1. O mercado deve ser dividido em submercados com elasticidades distintas do preço da demanda.
2. O monopolista tem que ser capaz de estabelecer uma efetiva separação dos submercados, para que não exista a possibilidade de revenda de um produto adquirido em um mercado de baixo preço em outro mercado de alto preço.⁸

O monopolista, para discriminar preços, tem que decidir a quantidade total que vai produzir, quanto vender em cada mercado e a que preço para maximizar o lucro.

Basicamente existem três tipos de discriminação de preços, conforme apresentado a seguir.

1º grau: o monopolista vende cada unidade do produto a preços diferentes. É a discriminação perfeita de preços. Cada unidade é vendida ao consumidor, pelo preço máximo que ele está disposto a pagar. Não há excedente do consumidor neste mercado. O monopolista extrai todo o excedente do consumidor.

2º grau: o monopolista vende diferentes unidades do produto a preços diferentes, mas todos os compradores que adquirem a mesma quantidade pagam o mesmo preço. O preço por unidade não é constante, depende da quantidade que o consumidor compra. Os preços estabelecidos pelo monopolista vão ter um comportamento não linear de análise mais complexa que foge ao escopo deste capítulo. Exemplos são os prestadores de serviços, os serviços de eletricidade etc.

3º grau: o monopolista vende o produto para diferentes compradores por preços diferentes, mas cada unidade vendida para um grupo de compradores é vendida ao mesmo preço. Essa é a forma mais comum de discriminação de preços que se aplica bastante nos descontos para idosos, estudantes, ou as diferentes classes de tarifas aéreas.

Para entender a lógica da discriminação de preços do terceiro grau, será apresentado um exemplo do processo pelo qual se determina os preços que serão cobrados nos dois mercados. Suponhamos que existam dois grupos de pessoas. O monopolista deseja vender uma quantidade de produto, para cada grupo, a um preço diferente.

Curvas de demanda inversa: $p_1(y_1)$ para o grupo 1; $p_2(y_2)$ para o grupo 2.

Custo de produção: $c(y_1 + y_2)$

A maximização dos lucros será definida pela equação:

$$\max \pi(RT - CT) = \max [p_1(y_1)y_1 + p_2(y_2)y_2] - c(y_1 + y_2)$$

O ponto de maximização será alcançado quando:

$$RMg_1 = CMg(y_1 + y_2)$$

$$RMg_2 = CMg(y_1 + y_2)$$

Substituindo a RMg pela fórmula seguinte tem-se:

$$\begin{aligned} p_1(y_1)(1 - 1/|\varepsilon_{d1}|) &= CMg(y_1 + y_2) \\ p_2(y_2)(1 - 1/|\varepsilon_{d2}|) &= CMg(y_1 + y_2) \end{aligned}$$

ε_{d1} e ε_{d2} representam, respectivamente, as elasticidades do preço da demanda nos mercados 1 e 2. Das duas últimas equações concluímos que:

$$p_1(y_1)[1 - 1/|\varepsilon_{d1}|] = p_2(y_2)[1 - 1/|\varepsilon_{d2}|]$$

Se o monopolista quiser fixar $p_1 > p_2$ terá que obedecer a:

$$\begin{aligned} 1 - 1/|\varepsilon_{d1}| &< 1 - 1/|\varepsilon_{d2}| \\ 1/|\varepsilon_{d1}| &> 1/|\varepsilon_{d2}| \\ |\varepsilon_{d2}| &> |\varepsilon_{d1}| \end{aligned}$$

Isso significa que o monopolista tem que estabelecer o preço mais alto no mercado de menor elasticidade. O preço mais baixo será estabelecido no mercado mais sensível ao preço (elasticidade mais alta) e o preço mais alto no mercado menos sensível ao preço. Esta é a maneira dele maximizar o lucro de uma forma geral.

1.4 Competição Monopolística

1.4.1 ANTECEDENTES

A insatisfação com os modelos de concorrência existentes na teoria econômica, competição perfeita e monopólio, resultou na elaboração de um modelo alternativo que incorporasse algumas das críticas que vinham sendo feitas. Podemos citar, a seguir, as principais críticas da época, em especial referidas ao modelo de competição perfeita formuladas por Piero Sraffa, em seu famoso artigo de 1926, “As leis dos rendimentos sob condições de concorrência”.

- Não explicava vários fatos do mundo real.
- A hipótese de produto homogêneo não se adequava ao padrão de competição dominante nos mercados, onde a propaganda e outras técnicas de venda prevaleciam como instrumentos fundamentais para criar a fidelidade do consumidor.
- As empresas expandiam sua produção com custos decrescentes (retornos de escala crescentes), como era a previsão do modelo de competição perfeita, sem se tornar grandes empresas e ganhar poder de mercado.

Os modelos que surgiram para confrontar esse desafio teórico procuraram conciliar o poder de mercado das empresas com a hipótese de lucro econômico zero no longo prazo. Esta estrutura de mercado deveria combinar atributos do monopólio e da competição perfeita.

1.4.2 O MODELO

A definição dessa estrutura de mercado afirma que em uma indústria em competição monopolística existe livre entrada e as empresas se deparam com uma curva de demanda negativamente inclinada, e não horizontal como no caso da competição perfeita. Se as empresas entram na indústria toda vez que houver lucros positivos, cada empresa estará em equilíbrio a longo prazo, como em competição perfeita. Se as empresas se deparam com uma curva de demanda residual negativamente inclinada, elas têm poder de mercado.

O conceito-chave da competição monopolística é a diferenciação de produtos. A diferenciação de produtos pode surgir a partir de duas características. A primeira porque os consumidores pensam que um produto é diferente dos demais. A segunda porque os consumidores preferem os produtos que possuem características ou atributos diferenciados e estão dispostos a pagar um prêmio para comprar esses produtos. No primeiro caso, a propaganda e as técnicas de vendas desempenham um papel importante. No segundo, características como a localização geográfica ou aspectos técnicos e de qualidade dos produtos fazem a diferença.

Como sempre, as empresas em competição monopolística vão maximizar lucros no ponto em que a RMg se iguala ao CMg. A curva de demanda residual da empresa e, a partir dela, sua curva de RMg dependem da quantidade produzida por cada um dos seus competidores, e não apenas da quantidade total produzida no mercado. Quanto maior for o impacto da diferenciação, maior será a inclinação da curva de demanda, porque os outros produtos são substitutos mais distantes. Essa maior inclinação confere à empresa um poder de elevar seu preço acima do CMg.

O número de empresas no mercado é determinado pela facilidade de entrada na indústria. A entrada de novas empresas auxilia o consumidor porque baixa o preço dos produtos e aumenta a variedade de produtos que ele pode escolher. Porém, o poder da entrada de novas empresas regula o preço em favor do consumidor, em função do tipo de diferenciação que estamos tratando no modelo.

Uma análise mais detalhada do modelo de competição monopolística será apresentada no Capítulo 6 deste livro.

1.5 Conclusão

A estrutura de mercado, definida teoricamente como competição perfeita, não prevê qualquer tipo de rivalidade entre as empresas. Portanto, o significado teórico do conceito de competição perfeita tem uma dimensão radicalmente oposta ao uso que ele tomou pelo senso comum. Aqui, competição significa rivalidade. Na teoria significa ausência de rivalidade.

A teoria econômica neoclássica, que fornece a visão convencional corrente sobre o comportamento da empresa, não atenta, de forma adequada, às diferenças entre as empresas. As empresas são iguais, os comportamentos são idênticos. Pode-se representar todas as empresas com uma empresa-padrão. Isso resulta em um mercado ou indústria, como a agregação dos consumidores e das empresas individuais. Não há espaço para estratégias diferenciadas, nem quando se reconhece que pode haver diferenciação de produtos (competição monopolística).

A primeira razão para isso é a concepção da teoria neoclássica sobre o que é a atividade econômica. Desde a formulação da teoria do equilíbrio geral, há mais de um século, o foco tem sido fundamentalmente sobre a eficiência da alocação de recursos. A pergunta teórica central a ser respondida é: com que eficiência uma economia aloca os seus recursos, dadas as preferências dos consumidores e as tecnologias disponíveis para produção. Essa concepção resulta em uma visão muito limitada sobre o que as empresas fazem.

A segunda razão também decorre dessa orientação teórica mais geral. Embora não seja a única formulação possível do processo de tomada de decisão da empresa, compatível com a teoria neoclássica, presume-se que as empresas encaram o conjunto de escolhas, como dado e conhecido, restrinindo pelas tecnologias disponíveis. A partir daí, as empresas não têm qualquer dificuldade em escolher a ação que é melhor para elas, dado o seu objetivo de maximização de lucro.

Dessa forma, o problema econômico se resume basicamente a conseguir os preços certos, que funcionam como um mecanismo de incentivos, e não a identificar a melhor coisa a fazer, a qual foi presumida não ser um problema, pelo perfeito conhecimento de todas as alternativas possíveis.

A permanência da hipótese de perfeito conhecimento e maximização de lucros nos modelos de monopólio e competição monopolística, isto é, a racionalidade perfeita do tomador de decisões, leva a que todas as modificações teóricas feitas a seguir continuem a se constituir como um caso especial, do caso geral, a competição perfeita. Tudo mais que não se enquadre nas hipóteses básicas do modelo é considerado uma falha ou imperfeição de mercado.

Notas

1. Sraffa, P. (1926) "The Laws of Returns under Competitive Conditions". *Economic Journal*, 36(2), pp. 535-550, dezembro.
2. Robinson, J. (1933) *The Economics of Imperfect Competition*. Londres, Macmillan Press.
3. Chamberlin, E. H. (1933) *The Theory of Monopolistic Competition*. Cambridge, Mass., Harvard University Press.
4. A função custo médio mede o custo por unidade de produção. A função custo variável médio mede o custo variável por unidade de produção. A função custo fixo mede os custos fixos por unidade de produção. O Capítulo 3 traz uma análise detalhada dos custos de produção.
5. As economias de escala ocorrem quando as empresas aumentam a sua produção ou o tamanho da sua planta. As economias de escala moldam o formato da curva de custo médio de longo prazo, ao passo que a posição desta curva depende das economias externas, como mudanças na tecnologia e variações nos preços dos fatores de produção na indústria ou na economia como um todo. As mudanças nessas condições externas vão representar um deslocamento da curva de custo médio de longo prazo (custo unitário de produção). Existe também a possibilidade da ocorrência de deseconomias de escala, quando a empresa cresce além de determinados limites, tornando-a sujeita a altos custos de gerenciamento da produção. As economias de escala, portanto, descrevem o que acontece quando se aumenta a quantidade de todos os fatores de produção presentes na função de produção. As economias de escala serão detalhadamente discutidas no Capítulo 6.

6. O grau desse poder de mercado não é absoluto, pois depende da elasticidade-preço da demanda.
7. Na verdade, essa hipótese é um pouco forte. Os custos podem variar, mas não tanto quanto os preços. Da mesma forma, pode haver ligeiras diferenças entre os produtos como: encadernações diferentes dos livros, assentos distintos no teatro, no avião ou no trem.
8. Esta condição mostra por que é fácil aplicar a discriminação de preços em serviços de utilidade pública: eletricidade, gás, transporte, medicina, espetáculos etc.

Bibliografia

- Chamberlin, E. H. (1933) *The Theory of Monopolistic Competition*. Cambridge, Mass., Harvard University Press.
- Koutsoyiannis, A. (1983) *Modern Microeconomics*. Londres, Macmillan Press (Segunda edição).
- Robinson, J. (1933) *The Economics of Imperfect Competition*. Londres, Macmillan Press.
- Sraffa, P. (1926) "The Laws of Returns under Competitive Conditions". *Economic Journal*, 36(2): pp. 535-550, dezembro.
- Tolipan, R.; Guimarães, E. A. (1988) Uma nota introdutória ao artigo "As leis dos rendimentos sob condições de concorrência" de Piero Sraffa. *Clássicos de Literatura Econômica*. Rio de Janeiro, IPEA/INPES.
- Varian, H. R. (1999) *Microeconomia. Princípios básicos*. Rio de Janeiro, Editora Campus.

Empresa, Indústria e Mercados

Alexis Dantas, Jacques Kertsnetzky e Victor Prochnik

2.1 Introdução

Este capítulo visa analisar a evolução dos conceitos de empresa, indústria e mercado no âmbito da Economia Industrial. O objetivo básico consiste na avaliação dos desdobramentos teóricos da introdução de categorias analíticas ausentes, ou insatisfatoriamente abordadas, pela visão tradicional neoclássica, em particular no que se refere à natureza da empresa e seus objetivos e consequências para a delimitação dos conceitos de mercado e indústria. Na Economia Industrial, em várias correntes, destaca-se claramente a busca pela incorporação do crescimento e da acumulação de capital das empresas como determinantes fundamentais da dinâmica da economia capitalista. A crescente importância das grandes corporações, em geral acumulando várias atividades produtivas, contrasta claramente com a empresa idealizada pela escola tradicional neoclássica, gerando novas preocupações não só com a própria ideia de empresa como também com seus espaços de concorrência, sobretudo na definição de mercado e indústria.

Com relação a isso, a próxima seção se ocupa em examinar as diversas abordagens referentes ao papel observado pelas empresas na economia capitalista, buscando identificar modelos que permitam analisá-las em função de sua dinâmica de expansão e valorização de capital. Em seguida, a terceira seção aborda as formas tradicionalmente adotadas pelas empresas no que se refere à sua organização interna, associando sua configuração a estratégias típicas de expansão. A quarta seção discorre acerca dos conceitos de mercado e indústria, buscando refletir os novos traços característicos das empresas e suas formas tradicionais de expansão, com destaque para a diferenciação de produto e a diversificação industrial.

Finalmente, a quinta seção apresenta espaços alternativos de análise da concorrência empresarial – as cadeias e complexos industriais, cujo objetivo é explorar as interdependências dos agentes econômicos que extrapolam os limites estritos da definição de indústria como espaço concorrencial típico das empresas.

2.2 Natureza e Objetivos da Empresa

A questão da natureza e dos objetivos das empresas encontra na evolução da Economia Industrial uma diversidade de respostas.

Antes de expor essa diversidade é possível explicá-la, como fizeram Chandler e Penrose. Para Chandler, a empresa tem várias faces, que podem ou não ser enfatizadas na elaboração teórica.

Uma empresa é uma entidade legal que estabelece contratos com fornecedores, distribuidores, empregadores e, frequentemente, com clientes. É também uma entidade administrativa, já que havendo divisão do trabalho em seu interior, ou desenvolvendo mais de uma atividade, uma equipe de administradores se faz necessária para coordenar e monitorar as diferentes atividades. Uma vez estabelecida, a empresa se torna um conjunto articulado de qualificações, instalações e capital líquido. Finalmente, em nome de lucros, empresas têm sido e são instrumentos de economias capitalistas para a produção de bens e serviços e para o planejamento e a alocação para produção e distribuição futuras.

(Chandler, 1992, p. 483)

A proposição de Penrose é complementar, indicando que os teóricos fazem escolhas dentre os múltiplos aspectos da empresa.

Uma empresa [...] não é um objeto observável de maneira fisicamente separada de outros objetos, e é difícil de se definir a não ser com referência ao que faz ou ao que é feito em seu interior. Consequentemente, cada analista é livre para escolher quaisquer características da empresa nas quais esteja interessado, definir a empresa em termos destas características, e proceder de forma a chamar sua construção de “empresa”.

(Penrose, 1959, p. 10)

O retrato de empresa contido nas teorias econômicas depende da visão de funcionamento do sistema econômico mais geral desenvolvido pelas teorias.

Com essa observação em mente, podemos partir para a discussão da natureza e dos objetivos da empresa, sem pensar que estaremos buscando o que é realmente uma empresa e quais são realmente seus objetivos, mas sim como as teorias econômicas as retratam, sabendo que a diversidade desse retrato faz parte do mundo da ciência.

Podemos ainda levar em conta um aspecto adicional: a evolução histórica das unidades que organizam a produção também tem influência sobre a formulação teórica da empresa.

2.2.1 ANTES DA ESCOLA NEOCLÁSSICA: ACUMULAÇÃO DE CAPITAL E ELEMENTOS DE UMA TEORIA DA PRODUÇÃO

A escola clássica não chega a explicitar um agente de nome empresa. Nela estão presentes como agentes as classes sociais – trabalhadores, proprietários de terras e capitalistas. Para nosso uso, isso pode se explicar pelo fato de a empresa capitalista se ter constituído como agente de forma evolutiva, sendo identificável de forma separada de seus proprietários à medida que evoluiu juridicamente. Assim, antes da revolução industrial podem identificar-se empresas na esfera comercial, mas a produção se faz, em geral, de forma doméstica ou em oficinas de dimensões reduzidas. Mesmo no século XIX, as primeiras empresas industriais (da indústria têxtil, por exemplo) são empresas familiares ou sociedades de natureza jurídica simples, não separando a responsabilidade do patrimônio familiar dos compromissos assumidos pelas empresas. Neste sentido, a empresa da escola clássica se identifica com o capitalista, e seu objetivo é acumular capital em um ambiente competitivo representado por um sistema capitalista em expansão.

Do ponto de vista da compreensão da escola neoclássica, que será discutida a seguir, cabe ainda recolher na escola clássica elementos da teoria da produção desenvolvidos por seus autores, que comporão a agenda de questões discutidas nas teorias da produção desenvolvidas posteriormente.

Estamos falando especificamente das leis dos rendimentos.¹ Tais leis procuravam relacionar a ampliação da atividade econômica à produtividade e podiam ser verificadas seja na unidade individual da produção, seja no conjunto da atividade. É assim que Adam Smith propôs que quanto mais amplo o mercado (e maior a produção), mais profunda pode ser a divisão do trabalho (seja na unidade da produção, seja no conjunto da atividade econômica), apresentando um material que mais tarde será proposto como lei dos rendimentos crescentes. E é assim também que Ricardo apontou que a agricultura, por contar necessariamente com o recurso a terras adicionais para a expansão da atividade produtiva, visando atender ao aumento da população e da demanda de alimentos, apresenta produtividade decrescente porque as terras são incorporadas por ordem de fertilidade – o que compõe a lei dos rendimentos decrescentes. A unidade de produção agrícola também apresentaria individualmente rendimentos decrescentes às tentativas de aumentar a produção, porque a intensificação do uso de uma porção fixa das terras apresentaria resultados análogos à extensão do cultivo a terras adicionais de menor fertilidade.

2.2.2 A EMPRESA NA ESCOLA NEOCLÁSSICA: TRANSFORMAÇÃO NAS LEIS DOS RENDIMENTOS

A escola neoclássica trouxe para o centro da teoria econômica o problema da alocação de recursos escassos a necessidades ilimitadas. O problema da alocação de recursos entre setores da atividade já estava presente na escola clássica com o problema da expansão do sistema capitalista. Essa mudança de eixo teórico de uma escola para outra é acompanhada pelo deslocamento das leis dos rendimentos na discussão da expansão do sistema para a discussão do valor das mercadorias como solução do problema alocativo.

Vejamos primeiro a versão do equilíbrio parcial, referente ao desenvolvimento da teoria do valor a partir de proposições contidas nos *Principles of Economics* de Alfred Marshall. Como agente deste sistema alocativo, a empresa neoclássica é vista como um agente que toma decisões de produção (curto prazo) e de escolha do tamanho da planta (longo prazo), incluindo a entrada ou saída de mercados onde os lucros estejam acima ou abaixo dos lucros normais, de forma que as decisões do conjunto de empresas de uma economia conduzem as escolhas da aplicação dos recursos da sociedade – o que, como, quanto e para quem produzir. As escolhas individuais das empresas são governadas pelo objetivo de maximização de lucros, que corresponde à quantidade produzida que proporciona os maiores lucros dentre o conjunto de quantidades que uma planta permite produzir (curto prazo), ou à escolha da planta ótima, a que permite obter a maior lucratividade dentre o conjunto de dimensões alternativas de plantas no âmbito da tecnologia vigente.

A empresa é, assim, o local onde se combinam os fatores de produção de maneira a gerar os produtos, sendo a produção sujeita às leis dos rendimentos, que são discutidas primordialmente no interior de cada unidade de produção isolada. Os rendimentos são a base para a construção das curvas de custo médio e marginal de curto e longo prazos. A lei dos rendimentos decrescentes é generalizada para qualquer unidade que apresente fatores variáveis combinados a fatores fixos, não mais se restringindo à agricultura. A existência de ao menos um fator fixo compõe a noção de curto prazo.

Na versão do equilíbrio geral introduzida por Léon Walras, o retrato da empresa é ainda mais acentuadamente talhado para o modelo. A empresa aparece sob a forma de empresários que comparecem no mercado de fatores como demandantes de seus serviços e no mercado de bens como ofertantes dos produtos. Se as remunerações concebidas são os lucros do capital, salário do trabalho e a renda dos recursos naturais, os lucros extraordinários a que empresários poderiam eventualmente visar como resíduo – resultante do desconto dos custos de produção do valor dos bens – se anulam por pressão da competição. No equilíbrio não resta ao empresário qualquer remuneração salvo aquela(s) que aufera como proprietário de algum fator dentre os combinados na produção. Seu papel de auxiliar o “leiloeiro walrasiano” na coordenação de disposições de comprar e vender (bens e serviços de fatores), para igualar ofertas e demandas na economia, não é remunerado, assim como o do próprio leiloeiro. Essas limitações, devemos frisar, não são decorrentes de incapacidade intelectual ou de percepção da realidade por parte do formulador da teoria. São resultado da tentativa de oferecer uma solução ao problema do equilíbrio geral com um desenho institucional simples.

No entanto, a compreensão da atividade econômica e das empresas como agente pode ser aprofundada por meio da consideração de instituições, inclusive por meio da discussão institucional sobre a natureza da empresa. Neste âmbito, a empresa é uma organização hierárquica. Esta caracterização admite a discussão de aspectos adicionais na teoria da empresa.

2.2.3 EMPRESAS COMO INSTITUIÇÃO: A CONTRIBUIÇÃO DE COASE

De acordo com Ronald Coase, a empresa é vista como um arranjo institucional que substitui a contratação renovada de fatores no mercado por uma outra forma de contratação, representada por um vínculo duradouro entre fatores de produção. Na contratação entre capital e trabalho, por exemplo, seria a diferença entre contratar um autônomo para uma tarefa avulsa ou contratar um trabalhador mediante um contrato de trabalho, ou seja, para desempenhar essa tarefa ao longo de um futuro indeterminado, incluindo as variações que a tarefa pode assumir ao longo do tempo.

Dessa maneira, o autor identifica duas formas alternativas de alocação de recursos: uma pelo mercado, flexível, elástica, respondendo às mudanças nas condições e sinalizada por preços; outra, hierárquica, correspondendo às ordens emitidas pela hierarquia interna à empresa, que destina aos fatores contratados sua utilização produtiva. As duas podem conviver porque existem vantagens de parte a parte: as empresas economizam custos de transação, porque seus contratos dispensam-na de recorrer repetidamente ao mercado para cada utilização de serviços dos fatores de produção.

Mas essa economia possui limites na ineficiência gerencial que cresce com o número de fatores contratados pela hierarquia, de maneira que a partir de certa dimensão a economia de custos de transação seja compensada por deseconomias da ineficiência gerencial. Define-se dessa forma, empregando a análise marginal, o tamanho ótimo da empresa, que faz com que uma empresa hipotética não cresça ao ponto de a hierarquia resultante prescindir das relações de mercado. Como resultado, empresas convivem nos mercados transacionando entre si e com fatores que vendem seus serviços de forma avulsa.

Assim, a empresa de Coase é uma hierarquia que economiza custos de transação. Ao manter o problema alocativo como central e ao empregar o cálculo racional e a análise marginal na formulação do tamanho ótimo da empresa que maximiza lucros, a empresa de Coase deve ser entendida como um desenvolvimento teórico ainda no âmbito da abordagem neoclássica.

2.2.4 OUTRAS VISÕES DE EMPRESA COMO INSTITUIÇÃO: MARSHALL

Há, porém, um amplo conjunto de formulações teóricas que se apresentam como críticas ou como alternativas à discussão neoclássica das empresas. Tais formulações podem ser analisadas em conjunto devido a um elemento comum às suas abordagens.

Esse elemento pode ser representado pela expressão *capacitações organizacionais*, de uso recente, mas de conteúdo identificável já na obra de Alfred Marshall.

Marshall é, sem dúvida, o fundador da vertente neoclássica de análise do equilíbrio parcial. Efetivamente, Marshall utilizou nessa análise a figura de uma empresa idealizada, a empresa representativa, que seria madura o suficiente para estar de posse de capacitações representativas do desenvolvimento geral da indústria e do conjunto de empresas produtoras da mercadoria em análise de equilíbrio.

No entanto, antes de partir para a análise do mercado e seu equilíbrio no livro V dos *Principles of Economics*, Marshall discutiu a produção e seus agentes no livro IV. Para Marshall, as empresas se desenvolvem ao longo de um ciclo de vida, no qual nascem e sobrevivem se o seu fundador possui qualidades que o selecionem no ambiente. Ele deve trazer soluções adequadas aos problemas de organização e das técnicas de produção, comercialização das mercadorias e relacionamento com fornecedores.² Uma vez que a empresa sobrevive a seu nascimento e cresce, o empresário precisa resolver os problemas de seu crescimento, já que organizar a produção e a comercialização de quantidades crescentes requer novas soluções.

Para Marshall, o crescimento da empresa individual na indústria se faz sob rendimentos crescentes. O autor explica que empresas maiores se beneficiam de vantagens na adoção de técnicas, na compra de grandes volumes, no uso de instrumentos de comercialização acessíveis às empresas maiores. Se juntarmos estas vantagens estáticas às vantagens dinâmicas referentes à experiência, aos conhecimentos acumulados, aos relacionamentos comerciais estabelecidos, a uma estrutura organizacional interna que amadurece e se consolida, tudo isto em evolução ao longo do tempo, o resultado é que quanto maior a empresa mais competitiva ela é.

A primeira empresa a atingir dimensões suficientes não é, para Marshall, capaz de monopolizar o mercado. Esse autor discutiu a empresa familiar inglesa das últimas décadas do século XIX e primeiras do século XX. Historiadores revelam que, nesse ambiente, empresas eram criadas e iam à falência em profusão. A explicação marshalliana é a de que as empresas não podem reter e desenvolver eternamente as vantagens do tamanho porque, ao final da primeira geração fundadora da empresa, há perda de vigor no trabalho da gerência; e sua substituição pelas gerações posteriores faz prosseguir a debilidade porque seus gerentes são determinados hereditariamente, não sendo selecionados pelo mercado em sua capacidade de conduzir as empresas acompanhando o ambiente em permanente mudança. Como resultado, as boas soluções criadas no passado envelhecem e a empresa entra em decadência.

No entanto, no século XX se desenvolve o fenômeno da separação da propriedade e do controle, alterando o ambiente que serve de inspiração e de aplicação da teoria da empresa. De qualquer forma, permanece a contribuição marshalliana de um ambiente em permanente mudança, da necessidade dos responsáveis pela condução da empresa acompanharem seu tempo e introduzirem mudanças em várias áreas (produção, administração, comercialização) que representem vantagens competitivas.

2.2.5 OUTRAS VISÕES DE EMPRESA COMO INSTITUIÇÃO: GERENCIALISTAS E PENROSE

Assim como outras abordagens alternativas à visão tradicional neoclássica, a corrente gerencialista³ rejeita o processo de maximização do lucro como o determinante exclusivo do comportamento decisório da empresa. Um elemento-chave na configuração dos modelos dessa corrente refere-se à separação entre propriedade e controle, uma nova característica organizacional das empresas ao introduzir a figura do gerente profissional – representando seu corpo executivo. Esses executivos, por sua vez, possuem objetivos próprios, não necessariamente coincidentes com os interesses dos acionistas proprietários. Com relação a isso, uma função-utilidade dos gerentes, que orienta as suas decisões, conteria não apenas os lucros, mas também os outros elementos que afetam suas carreiras, seu emprego, suas oportunidades de remuneração futura: parcelas de mercado das empresas, grau de risco, crescimento das vendas apareceriam ao lado dos lucros como variáveis, inclusive competindo com os lucros. Um gerente profissional poderia “trocar” um pouco de lucros por um pouco mais de vendas como objetivo, para valorizar seu prestígio entre os gerentes existentes na economia – variáveis associadas ao crescimento da empresa ocupam, portanto, papel preponderante nesses modelos.

Edith Penrose formulou uma teoria da empresa que se destaca deste conjunto por retomar a problemática esboçada por Marshall, dando maior organicidade à discussão de empresas que crescem acumulando capacidades e recursos. Para a autora, a empresa reúne e combina recursos, mas esta função contrasta com a empresa neoclássica porque não há relação biunívoca entre um recurso e os serviços que dele se podem obter. Estes dependem do ambiente (da empresa) em que os recursos são utilizados, com especial importância para os conhecimentos utilizados quando do seu emprego. Empresas encerram experiência e conhecimentos acumulados ao longo de sua existência que dela fazem um exemplar único, resultante de sua trajetória específica, dos problemas que enfrentam, das estratégias e soluções que escolhem.

Os conhecimentos necessários para a atividade da empresa são, em parte, tácitos, de forma que os recursos humanos que dela participam os adquirem pela experiência comum. Esta proporciona à atividade da empresa o caráter de trabalho de equipe. A própria elaboração de estratégias depende da avaliação dos membros da empresa e, consequentemente, de sua experiência passada e conjunta. Tudo isto em um ambiente hierárquico, onde se distinguem os recursos humanos como participantes de uma organização hierárquica e com divisão do trabalho.

Penrose destaca-se ainda dos demais gerencialistas por considerar que os vários objetivos, que podem ser identificados como perseguidos pelos gerentes, podem ser englobados no objetivo mais amplo de crescimento. Lucros, segurança, parcelas de mercado são vistos como resultado e fator de crescimento, em uma visão dinâmica da empresa que se transforma e cresce.

2.2.6 A VISÃO NEOSCHUMPETERIANA DE EMPRESA

Uma corrente de teoria econômica contemporânea em que a empresa se apresenta como agente que acumula capacidades organizacionais, é a dos neoschumpeterianos. Em uma das obras seminais desta corrente, Richard Nelson e Sidney Winter apresentam essas capacitações sob a forma de rotinas. Para estes autores, em vez da escolha racional e permanentemente renovada proposta pela corrente principal da teoria econômica, as empresas se comportam de acordo com rotinas cristalizadas por meio de sua experiência, que possuem o papel de coordenar a atividade interna dos membros da empresa, ao mesmo tempo em que encerram o conhecimento da organização, à semelhança de um código genético. Tal conhecimento é, em parte, apreciável de caráter tácito, não sendo transferível por meios formais, e compõe o caráter idiossincrático da atividade empresarial. A discussão das rotinas enfatiza um aspecto central do comportamento das empresas: o de que não bastam os equipamentos e seus manuais para sua utilização; a empresa não é uma planta operada com custos variáveis na forma de trabalho que pode ser contratado ou demitido; as rotinas encerram o conhecimento da empresa, e incluem produção, transmissão e interpretação das informações provenientes do ambiente externo e as geradas no interior da empresa. Sendo, em boa parte, conhecimento tácito e não formal, é adquirido por meio da participação na atividade rotineira.

A proposição de rotinas como descrição do comportamento das empresas não implica comportamento imutável. A abordagem evolucionista de Nelson e Winter discute a relação entre as rotinas e a inovação.⁴ Problemas detectados nas rotinas podem pôr em ação rotinas de solução de problemas ou demandar alterações nas próprias rotinas; a introdução de inovações pode implicar o desenvolvimento de novas rotinas ou adaptação das rotinas anteriores. A própria geração de inovações é uma atividade passível de organização em rotinas que consistem em princípios de busca de soluções de problemas por parte de cientistas, engenheiros e gerentes.

2.3 Estrutura Organizacional Interna da Empresa

A insatisfação com o tratamento conferido à empresa (e ao ambiente concorrencial) pela teoria neoclássica tradicional é um dos pontos de contato de maior relevo nos vários modelos que compõem a análise da Economia Industrial, especialmente em função da introdução de categorias necessariamente dinâmicas ao corpo analítico proposto. Isto se reflete, em particular, nos conceitos específicos de empresa, indústria e mercado.

Como visto na exposição precedente, de forma geral a empresa como instituição é entendida como uma entidade administrativa e financeira cujos objetivos predominantes são o crescimento e a acumulação interna de capital. A diversificação industrial apresenta-se, historicamente, como uma das formas mais tradicionais de expansão das empresas na economia capitalista (ver o Capítulo 15).

Essa empresa diversificada, todavia, pode apresentar diferentes formatos de organização interna. Como ponto de partida para essa discussão, é possível considerar as proposições de Oliver Williamson e Alfred Chandler, relativas à existência de dois modelos estilizados de estrutura organizacional – o formato unitário (forma U) e a empresa multidivisional (forma M).

Em seu formato unitário, a empresa organiza-se segundo uma perspectiva estritamente funcional; as divisões que a compõem estão envolvidas cada uma delas com uma atividade de características particulares (produção, marketing, finanças etc.), que se sobrepõem à linha de produtos gerados. O caráter centralizado do empreendimento faz com que cada divisão envolva-se, de acordo com suas características, com uma ampla linha de produtos. Como problema potencial desse tipo de estrutura, é possível salientar o fato de que a alocação dos fundos disponíveis para investimento tende a ocorrer de acordo com a barganha de interesses entre as várias divisões funcionais, o que pode negligenciar as oportunidades oferecidas pelos diferentes produtos em seus respectivos mercados.

Em contraste com essa estrutura unitária, é possível postular a existência de empresas com um formato multidivisional. Essas empresas funcionam a partir de um sistema de divisões organizadas por produto ou por região geográfica, cada uma delas comportando-se como instância operacional individualizada, responsável por um amplo elenco de decisões locais concernentes a preços e produção. Nesse caso, a empresa diversificada pode ser compreendida como uma federação de quase-empresas, às quais caberia a responsabilidade pela produção colocada em um mercado particular, de tal modo que a empresa diversificada estaria dividida em tantas quase-empresas quanto fosse o número de mercados em que atua.

Embora cada quase-empresa possua um corpo de gerentes próprio, responsável pelas decisões rotineiras relativas à produção e à comercialização no seu mercado específico, as decisões cruciais ficam a cargo da gerência central da empresa. Dessa forma, além de definir os cargos decisórios das quase-empresas, cabe à gerência central a tipificação das políticas e estratégias de investimento da empresa, além da alocação e distribuição de recursos entre as quase-empresas.

Este formato resulta na combinação de aspectos positivos associados à descentralização produtiva – permitindo a definição de espaços próprios ocupados pelas quase-empresas (que representam linhas de produtos específicos ou zonas geográficas definidas) – e à concentração decisória, referente à alocação de recursos entre as quase-empresas a partir da gerência central. Esta característica viabiliza uma importante diversidade de opções estratégicas para a conformação organizacional da empresa.

De maneira simplificada, é possível diferenciar os seguintes modelos organizacionais de empresas diversificadas.

1. *Empresa multiproduto*: produz vários bens colocados junto a mercados distintos, porém relacionados em termos das funções de P&D, fabricação e *marketing*. Sua expansão é, portanto, ou concêntrica, sendo induzida por similaridades tecnológicas e/ou mercadológicas das atividades previamente desenvolvidas – visando à exploração de economias de escopo e dos canais de comercialização disponíveis para a empresa.
2. *Empresa verticalmente integrada*: envolve a atuação da empresa em diversos estágios da cadeia produtiva associada à transformação de insumos em bens finais de determinada indústria. Em geral, a justificativa para esses movimentos prende-se à exploração de economias de escala – as quais se expandem para o conjunto de atividades da empresa integrada – que permitem a obtenção de ganhos de eficiência e a redução de custos de transação.
3. *Conglomerado gerencial*: corresponde a um tipo de empresa diversificada que está presente em vários mercados, envolvendo produtos pouco relacionados entre si. É caracterizada por uma capacitação gerencial genérica que pode ser utilizada em diferentes mercados, o que lhe confere uma vantagem concorrencial em relação a outras empresas que não dispõem desse tipo de capacitação. Utiliza essa capacitação gerencial para explorar oportunidades atrativas – embora possa também recorrer a um intercâmbio de recursos financeiros e gerenciais entre suas diversas unidades, o que resultaria em ganhos de eficiência para estas.
4. *Conglomerado financeiro*: corresponde a um tipo de empresa diversificada que está presente em diversos mercados que não se encontram relacionados entre si – nem do ponto de vista técnico-produtivo, nem mesmo do ponto de vista das capacitações gerenciais necessárias para operar aquelas unidades de maneira eficaz. Nesse caso, a interligação de atividades se dá basicamente por meio de controles financeiros, associados à distribuição de recursos líquidos pela gerência central que, em geral, dispõe de um acesso privilegiado aos circuitos financeiros.
5. *Companhia de investimento*: de modo semelhante ao conglomerado financeiro, também se baseia na distribuição de recursos líquidos entre atividades não relacionadas. No entanto, esse tipo de empresa apresenta uma grande volatilidade, em termos das áreas de atuação, para as quais seu projeto de diversificação se orienta. Nesse sentido, observa-se uma ênfase na maximização da rentabilidade do portfólio do conjunto de atividades para as quais a empresa direciona seus recursos. Assim, no caso de performance insatisfatória de determinada unidade, a tendência é que a empresa procure rapidamente se desfazer daquela atividade, em vez de procurar substituir sua gerência. Além disso, a companhia de investimentos não necessariamente detém o controle majoritário da propriedade das unidades operacionais que fazem parte de seu portfólio de negócios, podendo operar, com maior agilidade, na exploração de novas oportunidades rentáveis.

2.4 Os Conceitos de Indústria e Mercado

Assim como no caso da empresa, a evolução da Economia Industrial é também marcada pela tentativa de oferecer conceitos de indústria e mercado mais adequados à análise econômica. Destaca-se, mais uma vez, a insatisfação com o tratamento conferido ao tema pela abordagem tradicional neoclássica. No âmbito da dicotomia concorrência perfeita/monopólio que marca esta escola, o mercado é tratado como um espaço abstrato de encontro de oferta e demanda, adotando-se uma noção de produto como algo absolutamente bem definido e, portanto, perfeitamente distinguido na análise dos consumidores. Com relação a isso, o mercado reflete, em última instância, o conjunto de empresas (mono) produtoras desta mercadoria, de forma que a cada

indústria corresponda um mercado. Em consequência, o conceito de indústria assumido por esta corrente expressa espaços delimitados e estanques de competição.

Ao contrário, a suposição de um processo de crescimento da empresa marcado pela crescente diferenciação de produto, além de um movimento de expansão diversificante da atividade produtiva como uma estratégia fundamental, insere um alto teor de heterogeneidade de produto, ao menos no que se refere à percepção dos consumidores. Isto implica maior importância analítica da substituibilidade de produtos e do foco e direcionamento dos esforços competitivos, aspectos que passam a preponderar para uma definição mais adequada de mercado e indústria.

O mercado, portanto, corresponde à demanda por um grupo de produtos substitutos próximos entre si. Para uma empresa diversificada, no entanto, a ideia de mercado envolve também outros espaços concorrentiais em que pode atuar, definidos como área de comercialização por Edith Penrose. A indústria, por seu turno, é definida pelo grupo de empresas voltadas para a produção de mercadorias que são substitutas próximas entre si e, dessa forma, fornecidas a um mesmo mercado. Da mesma forma que para a noção de mercado, para uma empresa diversificada a indústria pode representar um conjunto de atividades que guardam algum grau de correlação técnico-produtiva, constituindo um conjunto de empresas que operam métodos produtivos semelhantes, incluindo-se em uma mesma base tecnológica de acordo com Penrose.⁵

De forma geral, portanto, mercado e indústria representam espaços de concorrência cuja delimitação não é (e não pode ser) estanque – nem no que se refere à definição do produto, nem quanto aos objetivos concorrentiais e de expansão. Do ponto de vista metodológico, a questão é definir o corte analítico – qual é efetivamente o grupo de produtos que compõem o mercado e, por conseguinte, que conjunto de empresas faz parte da análise de concorrência. A resposta não é, obviamente, tão simples, envolvendo, de forma geral, algum grau de arbitrariedade. Uma tentativa nessa direção é realizada a partir do desenvolvimento dos conceitos de cadeia produtiva e complexos industriais, que representam extensões da ideia de setor econômico em que são privilegiados os movimentos concorrentiais.

2.5 Cadeias Produtivas e Complexos Industriais

As experiências de ampliação da ideia de setor econômico são vistas como relevantes por causa da crescente interdependência econômica e social entre os agentes. Evidências empíricas da maior interdependência entre setores são numerosas, entre as quais a introdução e difusão dos métodos organizacionais japoneses e outras técnicas de gestão moderna (como *supply chain management*), a generalização das formas de parcerias e cooperação, crescente eletrônica da sociedade e o aumento das economias de escala e escopo das empresas. Estas duas últimas reproduzem, de forma ampliada, a necessidade de maior eficiência na operação intersetorial, como mostrou Alfred Chandler, para o crescimento da escala e da amplitude do escopo ocorrido na revolução industrial de fins do século XIX.

Assim, na medida em que a competitividade das empresas depende do seu meio ambiente imediato, a arena concorrencial se amplia, deixando de ser apenas a dos mercados imediatos de venda de mercadorias/serviços e aquisição de insumos, para também incorporar mercados acima e abaixo do nível da cadeia produtiva em que a empresa está atuando.

A análise de cadeias e complexos pretende dar conta desta questão. As cadeias produtivas resultam da crescente divisão do trabalho e maior interdependência entre os agentes econômicos. Por um lado, as cadeias são criadas pelo processo de desintegração vertical e especialização técnica e social. Por outro lado, as pressões competitivas por maior integração e coordenação entre as atividades, ao longo das cadeias, ampliam a articulação entre os agentes.

O conceito de cadeia produtiva pode se tornar uma ferramenta mais comum nos estudos econômicos. As aplicações existentes, algumas das quais mencionadas neste texto, demonstram, de forma convincente, sua utilidade. Do ponto de vista teórico, observa-se uma progressão, em várias correntes de pensamento econômico, na direção de uma melhor formatação do conceito de cadeia produtiva.

Cadeia produtiva é um conjunto de etapas consecutivas pelas quais passam e vão sendo transformados e transferidos os diversos insumos. Essa definição abrangente permite incorporar diversas formas de cadeias.

Segmentando-se longitudinalmente, pode-se ter uma *cadeia produtiva empresarial* onde cada etapa representa uma empresa (ou um conjunto de poucas empresas, que participam de um acordo de produção). Esse desenho é encontrado, por exemplo, em *supply chain management* e corresponde, também, à proposta, explicada adiante, de “subsistema vertical estritamente coordenado” (SVEC) – Zylberstajn e Farina (1999).

O recurso a esse tipo de cadeia é útil para a realização de análises empresariais, estudos de tecnologia e planejamento de políticas locais de desenvolvimento. Ele requer esforço de coleta de dados, pois os órgãos oficiais de estatística tendem a difundir dados mais agregados.

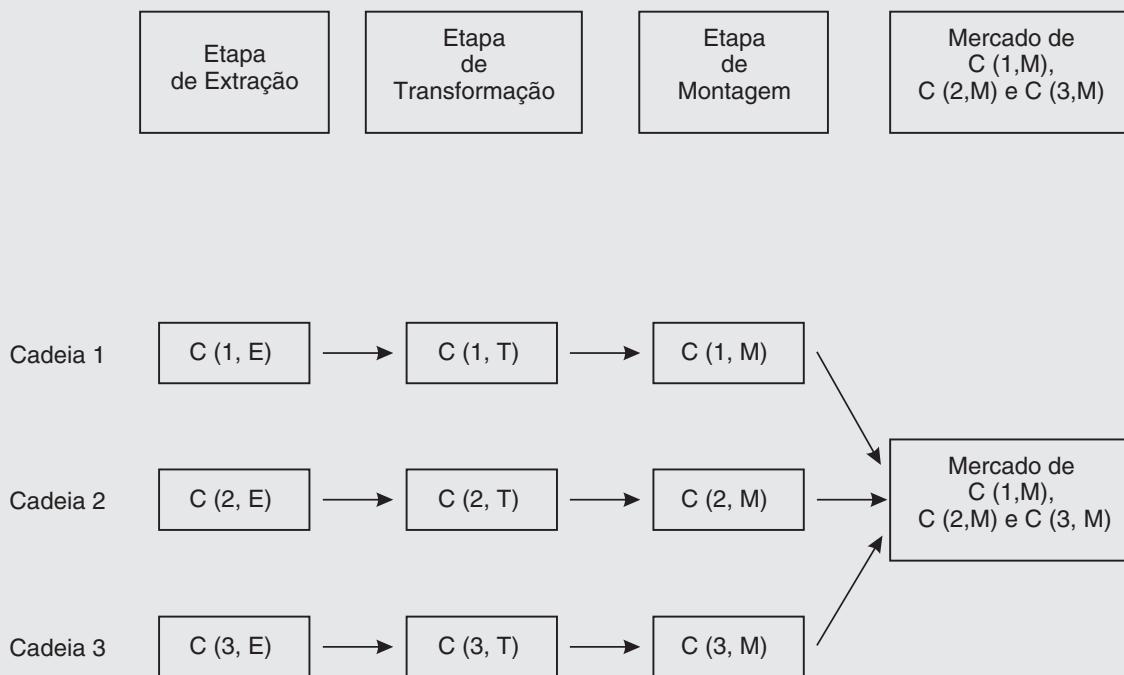
QUADRO 2.1 A NOÇÃO DE CADEIA INDUSTRIAL COMO ESPAÇO DE CONCORRÊNCIA


Gráfico 2.1 Esquema Simplificado de Cadeias e Etapas.

Para discutir como o conceito de cadeia produtiva se aplica à análise da concorrência, considere-se o Gráfico 2.1, em que os produtos substitutos de três cadeias competem pelo mesmo mercado. No gráfico, os retângulos representam indústrias; as setas, mercados entre indústrias consecutivas. São hipóteses do esquema simplificado: (1) cada indústria só adquire uma única matéria-prima, com exceção das indústrias extrativas, a montante das cadeias, que não compram insumos correntes; (2) a matéria-prima usada em uma indústria é totalmente produzida pela indústria que a antecede, com a óbvia exceção das indústrias extrativas; (3) o nível de integração vertical para trás e para frente é semelhante em todas as empresas de cada indústria e restrito à própria indústria e (4) são concorrentes os produtos das indústrias M1, M2 e M3. Os produtos das demais indústrias não são concorrentes entre si.

As cadeias são divididas em etapas: extração (E), transformação (T) e montagem (M). A divisão é arbitrária; a etapa de extração, por exemplo, poderia ser segmentada em extração e beneficiamento; e a de montagem em peças, partes e montagem final.

O Gráfico 2.1 é capaz de representar pelo menos três diferentes formas de concorrência. A primeira é a tradicional, entre as empresas de uma mesma indústria. A segunda forma de concorrência deriva-se do caráter estanque de cada cadeia. As vendas das indústrias finais, M1, M2 e M3, são iguais ao valor adicionado pelas suas respectivas cadeias. Assim, em cada cadeia, as empresas de uma indústria competem contra as empresas das demais, por uma parcela maior do valor adicionado. Acordos de preços setoriais e rompimento de acordos são dois exemplos de estratégias com este objetivo.

Pode-se sugerir também que, em uma cadeia, as indústrias com maior poder de mercado prevalecem sobre as demais. Acordos em uma indústria ou em uma cadeia envolvem, muitas vezes, empresas e indústrias com poder de barganha diferenciado. São vários os mecanismos que conferem o exercício de poder e de retaliação a algumas das empresas ou indústrias. Por isso, a análise das formas de concorrência entre indústrias sobre uma cadeia deve ser realizada a partir da determinação das indústrias motrizes que forem identificadas.

Uma terceira forma de concorrência, ao envolver duas ou mais cadeias, implica analisar a disputa entre indústrias motrizes diferentes. Os produtos das indústrias M1, M2 e M3 são substitutos e concorrem pelo mesmo mercado. Isto faz com que as empresas das demais indústrias dependam dos resultados da competição entre as indústrias finais. O processo de concorrência, nas indústrias terminais, envolve a todos e implica terceira forma de concorrência, o confronto entre as cadeias produtivas. Na construção civil, o exemplo mais conhecido é o que antepõe os produtores de aço aos de cimento.

Note-se, finalmente, a coexistência de competição e cooperação. As três formas de competição mencionadas – concorrência entre empresas de uma indústria, indústrias de uma cadeia e entre diferentes cadeias – se inter-relacionam de forma contraditória. As empresas de uma indústria competem entre si, mas têm interesses comuns diante das empresas das outras indústrias. As indústrias de uma cadeia, por sua vez, apesar de competirem entre si, são solidárias na disputa com outras cadeias, como mostra o debate sobre o uso do cimento ou do aço na construção civil.

Em um nível mais agregado, encontram-se as *cadeias produtivas setoriais*, nas quais as etapas são setores econômicos e os intervalos são mercados entre setores consecutivos. Variando a amplitude do leque de produtos considerados, nos setores econômicos obtêm-se cadeias mais ou menos desagregadas. Com relação a isso, pode-se ter, por exemplo, a cadeia dos calçados de couro ou a cadeia de calçados.

Duas cadeias são ditas concorrentes quando seus produtos finais servem a um mesmo mercado e as cadeias são relativamente independentes entre si. Cadeias concorrentes fabricam produtos substitutos. Em geral, o nível de desagregação de cadeias concorrentes está entre o das cadeias empresariais e o das setoriais. A cadeia produtiva de calçados de couro é diferente das cadeias de calçados de materiais sintéticos (embora ambas possam usar linhas de costura de náilon etc.). Manilhas de concreto estão em uma cadeia e manilhas de cerâmica em outra etc.

O entrelaçamento de cadeias é comum. Muitas cadeias se repartem e outras se juntam. Mas não há por que presumir que a teia de cadeias produtivas se espalhe, de maneira uniforme, sobre a estrutura econômica. Ao contrário, as cadeias de uma economia nacional podem ser agregadas em conjuntos, ou blocos, de forma que o valor médio das compras e vendas entre os setores constituintes de um bloco seja maior do que o valor médio das compras e vendas desses mesmos setores com os setores de outros blocos. Os blocos assim formados são denominados complexos industriais. Veja o Quadro 2.1.

Notas

1. Ver também o Capítulo 6.
2. Tais soluções significam um ambiente competitivo em permanente mudança a partir das iniciativas dos próprios empresários, com a ressalva de que as inovações não apresentam o aspecto radical da *destruição criadora* que Schumpeter irá propor mais tarde.
3. Destacam-se como principais referências Berle & Means, W. Baumol, R. Marris e A. Wood.
4. Do ponto de vista da empresa, a introdução de inovações requer uma ação sistematizada de busca por novas tecnologias, sujeita a rotinas específicas. Ademais, esta busca, por si só, não garante o sucesso do empreendimento inovativo, estando sujeita a um processo de seleção – ver Capítulos 8 e 19.
5. Ver o Capítulo 15.

Bibliografia

- Bellon, B. (1983) "La Filière De Production: Un Concept De Crise", Documento de Trabalho nº 106, Centre de Recherches en Économie Industrielle, Université de Paris-Nord, 1983.
- Berle, A. & Means, G. (1932) *The Modern Corporation and Private Property*. Nova York: Macmillan.
- Chandler, A. D. (1962) *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Cambridge, MIT Press.
- Chandler, A. D. (1977) *The Visible Hand: the Managerial Revolution in American Business*. Cambridge, Harvard U.P.
- Chandler, A. D. (1990) *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, Belknap P. of Harvard U.P.
- Chandler, A.D. (1992) "Organizational capabilities and the economic history of the industrial enterprise". *Journal of Economic Perspectives* 6(3):79-100, summer.
- Coase, R.H. (1937) "The nature of the firm". *Economica*, New Series, vol. 4, n. 16, nov.: 386-405.
- Haguenauer, L. & Prochnik, V. (2000) "A Delimitação de Cadeias Produtivas na Economia do Nordeste." Capítulo 2 de Haguenauer, L. & Prochnik, V. (orgs.) *Identificação de Cadeias Produtivas e Oportunidades de Investimento no Nordeste do Brasil*. Banco do Nordeste, Fortaleza.
- Haguenauer, L., Guimarães, E.A.A., Araujo, J.R. e Prochnik, V. (1984) "Complexos Industriais na Economia Brasileira". Texto para Discussão n. 62, IEI/UFRJ.
- Knight, F. (1921) [1971]. *Risk, Uncertainty and Profit*. Chicago, Chicago University Press.
- Marshall, A. (1920) *Principles of Economics*. London, Macmillan, 8^a ed.
- Nelson, R. e Winter, S. (1982) *An Evolutionary Theory of Economic Change*. Cambridge: Harvard U.P.
- Penrose, E.T. (1959) *The Theory of Growth of the Firm*. Oxford, Basil Blackwell, 1972.
- Penrose, E.T. (1959) *The Theory of the Growth of the Firm*. Oxford: Basil Blackwell. (second edition, 1980.)
- Perroux, F. (1977) "O Conceito de Pôlo de Crescimento". Schwartzman, J. Economia Regional – Textos Escolhidos, págs. 145/156; Convênio CEDEPLAR/CETREDE – MINTER, UFMG, ed. original Note sur la Notion de Pôle de Croissance, *Economie Appliquée*, 1955.
- Pitelis, C. (1993) "On transaction (costs) and markets and (as) hierarchies". IN: Pitelis, C. (ed.) *Transaction Costs, Markets and Hierarchies*. Oxford, Blackwell.
- Prochnik, V. (1987) "O Macrocomplexo da Construção Civil". Texto para Discussão n. 117, IE/UFRJ.
- Prochnik, V. (1989) "Cadenas y Etapas en el Complejo de la Construcción Civil". *El Trimestre Económico*, Vol. LVI (4), n. 224, out./dez. 1989, Ed. Fondo de Cultura Econômica, México.
- Richardson, G.B. (1972) "The organization of industry". *The Economic Journal* 82:883-896, September.
- Rocha, C.F.L. (1992) "A Teoria da Firma, o Agente Econômico e o Processo de Decisão", Cadernos ANGE, Textos Didáticos n. 3.
- Scherer (1970) *Industrial Market Structure and Economic Performance*, Chicago, Randy McNally.
- Schumpeter, J.A. (1942) *Capitalism, Socialism and Democracy*. London: George Allen and Unwin, 5a. edição, 1979.

- Schumpeter, J.A. (1961) *The Theory of Economic Development. An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*. Oxford: Oxford U.P.
Reimpressão da primeira edição em língua inglesa, de 1934.
- Simon, H. (1959) "Theories of Decision Making in Economics and Behavioural Science". *American Economic Review*, republicado em G.P.E. Clarkson (ed.) *Managerial Economics*. Harmondsworth: Penguin Books, 1969.
- Smith, A. (1776) [1983] *A Riqueza das Nações*. São Paulo, Abril Cultural.
- Steindl, J. (1952) "Maturidade e Estagnação no Capitalismo Americano", Edição *Os Economistas*, Ed. Nova Cultural, São Paulo, 1986.
- Trajtenberg R (1977) "Un Enfoque Setorial para el Estudio de la Penetración de las Transnacionales en América Latina". ILET, DEE/DI México.
- Williamson, O.E. (1975) *Markets and Hierarchies*, Nova York, Free Press.
- Zylberstajn, D. & Farina, E. M. M. Q. (1999) "Strictly Coordinated Food-Systems: Exploring the Limits of the Coasian Firm". *International Food and Agribusiness Management Review*, 2(2):249-265.

Economias de Escala e Escopo

Mariana Iootty e Marina Szapiro

3.1 Introdução

Os custos considerados pelos economistas são, em geral, diferentes daqueles utilizados pelos contadores. Estes últimos estão mais preocupados com os demonstrativos financeiros da empresa, isto é, com a contabilidade de todas as despesas, de fato, incorridas pelas empresas durante a produção. Os economistas, por sua vez, estão mais preocupados com o processo decisório, e, por isso, a análise econômica centra-se nos custos que poderão ocorrer no futuro e nos critérios que devem ser utilizados pela empresa para reduzir seus custos e melhorar sua lucratividade. Os custos considerados pelos economistas são os custos de oportunidade, definidos pelas oportunidades que serão deixadas de lado, caso a empresa (ou o indivíduo) não faça o melhor investimento. O custo de oportunidade de uma ação é dado pelo valor da melhor alternativa de alocação dos recursos empregados em tal ação. Vamos imaginar, assim, que o custo de oportunidade indica se uma atividade produtiva deve continuar (ou não).

Apresentar essas diferenças de entendimento sobre os custos é extremamente relevante para o desenvolvimento deste capítulo, no qual iremos analisar a relação existente entre as estruturas de custos e as economias de escala e escopo. Na Seção 3.2 apresentaremos os componentes básicos dos custos. Na Seção 3.3 discorreremos sobre os custos de curto prazo. Finalmente, na Seção 3.4 apresentaremos uma análise de custos de longo prazo.

3.2 Componentes Básicos dos Custos

Para os economistas, as empresas têm alguns custos que variam com a quantidade produzida e outros que são fixos. O custo fixo será gasto independentemente do nível de produção. Estes costumam incluir dispêndios, tais como aluguéis, equipamentos, manutenção da fábrica e seguro, entre outros. Adiante, os custos fixos são representados pela constante F , não apresentando relação alguma com o nível de produção. Já os custos variáveis CV são custos que variam proporcionalmente com o nível de produção. Os principais custos variáveis são matéria-prima e eletricidade, entre outros. Existem alguns custos que são conhecidos como custos irrecuperáveis (*sunk costs*). Por refletir uma relação complexa entre custo e tempo, este conceito de custos irrecuperáveis será mais bem definido e estudado especificamente no Capítulo 12.

Como vimos no Capítulo 2, é importante notar que a diferenciação entre custo fixo e variável só faz sentido quando se trata do curto prazo. Neste caso, alguns fatores de produção não podem variar sem que se incorra em perdas. Já no longo prazo, todos os custos são ajustados, pois há tempo suficiente para que todos os fatores de produção sejam ajustados sem custo algum.

Na medida em que os custos variáveis mudam com o aumento do nível de produção, eles são considerados uma função da quantidade produzida $CV(q)$. Os custos totais são a soma de todos os custos variáveis e fixos, isto é, $Ct(q) = F + CV(q)$. Quando a quantidade produzida cresce, o aumento dos custos totais corresponde somente (e é idêntico) ao aumento dos custos variáveis necessário para produzir a maior quantidade de produto.

Existem alguns outros conceitos importantes de custo, tais como custo médio (CMe), custo variável médio ($CVMe$) e custo fixo médio ($CFMe$). O CMe é definido como o custo total dividido pela quantidade produzida ($CMe = \frac{C(q)}{q}$). O $CVMe$ é o resultado da divisão entre o CV e a quantidade produzida ($CVMe = \frac{CV(q)}{q}$). O $CFMe$ corresponde ao custo fixo (F) dividido pela quantidade produzida, isto é, $CFMe = \frac{F}{q}$.

O custo médio é a soma do custo variável médio e do custo fixo médio ($CMe = CFMe + CVMe$). Dessa forma, o custo variável médio e o custo fixo médio não podem ser maiores do que o custo médio.

O custo marginal (CMg) é o incremento ou adição ao custo que decorre da produção de uma unidade a mais.¹ Nesse caso, o CMg é independente do custo fixo (F). Ele está relacionado apenas à variação do CV .

3.3 Os Custos de Curto Prazo

Supondo uma empresa no curto prazo com um custo fixo (F) de \$100, é possível observar, por meio do exemplo apresentado na Tabela 3.1, as principais relações entre CMg , CMe e $CVMe$. As relações geométricas entre os diversos custos e suas propriedades serão demonstradas em seguida.

O Gráfico 3.1 nos mostra o conjunto de curvas de $CFMe$, $CVMe$, CMe e CMg . Observamos inicialmente que a curva de $CFMe$ apresenta uma queda contínua, assumindo o formato da hipérbole correspondente a F/q .

A curva de $CVMe$ é inicialmente decrescente em resposta ao aumento da produtividade do fator variável, atingindo um ponto de mínimo quando então a planta opera com uma combinação ótima dos fatores fixo e variável, e aumenta posteriormente, como resposta à queda da produtividade do fator variável.

A curva de CMe corresponde à soma das curvas de $CFMe$ e de $CVMe$. Assim como a curva de $CVMe$, a curva de CMe assume um formato em U, decrescendo inicialmente, atingindo um ponto mínimo ao nível ótimo de operação da planta, e crescendo em seguida. Esse formato em U tanto da curva de $CVMe$, como da curva de CMe , reflete a lei dos rendimentos decrescentes, cuja dinâmica é descrita no Quadro 3.1. É válido ressaltar ainda que o ponto de mínimo da curva de CMe (ponto q_2) ocorre à direita do ponto de mínimo da curva de $CVMe$ (ponto q_1). Isso ocorre porque o CMe inclui o $CFMe$, que decresce continuamente com o aumento do nível de produção. Após o $CVMe$ ter atingido seu ponto mínimo, ele passa então a aumentar, sendo este aumento acompanhado pela queda do $CFMe$, o que leva, assim, o CMe a continuar decrescendo, embora o $CVMe$ esteja crescendo. Entretanto, esse crescimento do $CVMe$ passa a ser maior do que a queda do $CFMe$, e, portanto, o CMe passa a crescer. A curva de $CVMe$ então se aproxima assintoticamente do CMe à medida que o nível de produto cresce.

TABELA 3.1 Relações entre Custos

Produto	Custo fixo	Custo variável	Custo fixo médio	Custo variável médio	Custo total	Custo médio	Custo marginal
0	100	0			100		
1	100	10	100	10	110	110	10
2	100	19	50	9,5	119	59,5	9
3	100	25	33,3	8,3	125	41,7	6
4	100	32	25	8,0	132	33,0	7
5	100	40	20	8,0	140	28,0	8
6	100	49	16,7	8,2	149	24,8	9
7	100	60	14,2	8,6	160	22,9	11
8	100	73	12,5	9,1	173	21,6	13
9	100	88	11,1	9,8	188	20,9	15
10	100	108	10	10,8	208	20,8	20

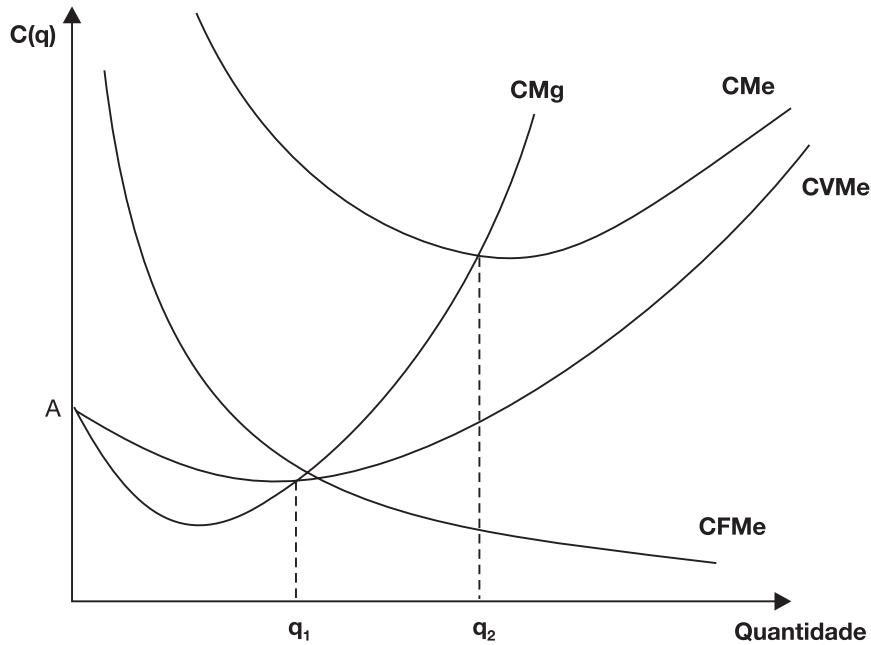


Gráfico 3.1 Curvas de Custo. A curva de custo marginal (CMg) corta as curvas de custo variável médio ($CVMe$) e custo médio (CMe) em seus pontos de mínimo.

QUADRO 3.1 LEI DOS RENDIMENTOS MARGINAIS DECRESCENTES

Consideremos uma empresa que contrata o fator de produção variável, a mão de obra (L), a uma remuneração fixa (w). O $CVMe$ corresponde ao custo variável por unidade de produto, $CVMe = \frac{CV(q)}{q}$. Assumindo que L unidades de mão de obra são utilizadas no processo produtivo, o custo variável é então definido como wL . Desse modo, $CVMe = \frac{wL}{q}$. Lembrando que o produto médio do trabalho ($PMel$) é definido como a produção por unidade de insumo, isto é, $PMel = \frac{q}{L}$, podemos denotar a seguinte expressão: $CVMe = \frac{w}{PMel}$. Ocorre então uma relação inversa entre o $CVMe$ e o $PMel$.

É razoável imaginar que o aumento inicial no nível de produção seja feito de forma eficiente para que a produtividade marginal do trabalhador seja crescente, o que acarreta, portanto, uma necessidade menor do fator variável trabalho (L) para produzir a mesma quantidade de produto. Dessa forma, o $PMel$ seria, por conseguinte, crescente, ocasionando, assim, a redução do $CVMe$. Como se trata de uma análise de curto prazo, a existência do fator fixo capital (K) restringe o processo de produção a partir de um determinado ponto, quando então a lei dos rendimentos marginais decrescentes começa a agir: a produtividade marginal do trabalhador diminui. Passa então a ser necessária uma quantidade maior do fator variável trabalho (L) para produzir a mesma quantidade de produto. Desse modo, o $PMel$ será, portanto, decrescente, ocasionando, dessa forma, o aumento do $CVMe$.

A curva de CMg assume um formato em U, o que também se explica pela lei dos rendimentos marginais decrescentes.² É importante observar que as curvas de CMg e $CVMe$ partem do mesmo ponto (ponto A). Se escrevermos a função de CMg em termos da função de $CVMe$, como descrito a seguir,

$$CMg(q) = \frac{\Delta CV(q)}{\Delta q} = \frac{CV(q + \Delta q) - CV(q)}{\Delta q}$$

então, a igualdade entre o CMg e o $CVMe$ no ponto A se explica pelo fato de que para a primeira unidade produzida é válido que:

$$CMg(q = 1) = \frac{CV(q = 1) + F - CV(q = 0) - F}{1_{\Delta q=1}} = \frac{CV(q = 1)}{1} = CVMe(q = 1)$$

Desse modo, o custo marginal de se produzir a primeira unidade de produção se iguala ao $CVMe$ de se produzir uma unidade.

Uma importante relação a ser destacada no diagrama de custos de curto prazo, representado anteriormente, é aquela que se estabelece entre as curvas de CMg e CMe . Para entendê-la melhor consideremos q^* como o ponto de mínimo do CMe ; assim, para qualquer nível de produção à esquerda de q^* o CMe deve ser decrescente, de forma que para $q < q^*$ é válido que:

$$\frac{d}{dq} \left(\frac{C(q)}{q} \right) \leq 0$$

Por meio do cálculo da derivada concluímos que

$$\frac{qC'(q) - C(q)}{q^2} \leq 0 \quad \text{para } q \leq q^*,$$

o que implica

$$C'(q) \leq \frac{C(q)}{q}, \quad \text{para } q \leq q^*.$$

Essa desigualdade nos indica que o CMg é menor do que o CMe para níveis de produção menores do que o ponto que minimiza o CMe . Analogamente, é possível demonstrar que:

$$C'(q) \geq \frac{C(q)}{q}, \quad \text{para } q \geq q^*.$$

Sendo válidas as duas desigualdades, então para $q = q^*$ a seguinte expressão é verdadeira

$$C'(q) = \frac{C(q)}{q},$$

ou seja, o CMg se iguala ao CMe no ponto de mínimo do CMe (ponto C). É possível notar que a relação entre o CMg e o $CVMe$ é semelhante, bastando apenas modificar a notação na demonstração acima para verificar que o CMg é inferior ao $CVMe$ quando este é decrescente, e é superior ao $CVMe$ quando este é crescente. De forma análoga, verificamos que o $CVMe$ passa pelo CMe no ponto de mínimo do CMe .

3.4 Análise dos Custos de Longo Prazo

No longo prazo, por definição, a empresa pode escolher a quantidade de todos os fatores que são utilizados. A curva de custo de longo prazo pode ser considerada uma curva de planejamento, na medida em que ela é um guia para o empreendedor planejar as decisões de expansão de produção no futuro.

Os custos de longo prazo refletem as escolhas da empresa quando as quantidades de todos os fatores podem variar. Dessa forma, no longo prazo, o que importa é o exame do comportamento global do custo diante da possibilidade de variação dos fatores de produção e, portanto, em relação à variação do nível de produto. Logo, é importante analisar o comportamento do custo médio de longo prazo ($CMeLP$). À medida que o nível de produção aumenta, os $CMeLP$ de uma empresa podem permanecer constantes, aumentar ou diminuir. Se o $CMeLP$ da empresa é reduzido quando a produção é elevada, a empresa possui economias de escala. Se o $CMeLP$ da empresa permanece constante na medida em que se produz maiores quantidades de produto, a empresa tem, então, retornos constantes de escala. Se o $CMeLP$ da empresa cresce quando a produção é elevada, a empresa possui deseconomias de escala.

Para poder compreender formalmente o comportamento do $CMeLP$ seria interessante analisar a derivação da curva de $CMeLP$ a partir da técnica produtiva utilizada pela empresa. A produção da empresa pode apresentar uma das três propriedades apresentadas a seguir.

1. Se $f(tK, tL) < tf(K, L)$, os retornos de escala são decrescentes (há deseconomias de escala).
2. Se $f(tK, tL) = tf(K, L)$, os retornos de escala são constantes.
3. Se $f(tK, tL) > tf(K, L)$, os retornos de escala são crescentes (há economias de escala).

Suponha que a função esteja sujeita a retornos crescentes de escala. Considere então $q_1 = f(K_1, L_1)$, onde os insumos K e L são comprados aos preços r e w , respectivamente.

Desse modo, $CMe_1LP = \frac{rK_1 + wL_1}{q_1}$. Se há uma variação positiva, ($t > 0$), de igual magnitude nos insumos K e L , de forma que $q_2 = f(tK_1, tL_1)$.

Sendo assim, o custo médio de longo prazo para o novo nível de produção será:

$$CMe_2LP = \frac{r.tK_1 + w.tL_1}{q_2} = \frac{t.(rK_1 + wL_1)}{q_2} = \frac{t.CMe_1LP.q_1}{q_2}$$

Mas $q_1 = f(K_1, L_1)$, e $q_2 = f(tK_1, tL_1)$, então

$$\begin{aligned} CMe_2LP &= \frac{t \cdot CMe_1LP \cdot f(K_1, L_1)}{f(tK_1, tL_1)}, \quad \text{daí:} \\ CMe_2LP &= CMe_1LP \cdot \left(\frac{t \cdot f(K_1, L_1)}{f(tK_1, tL_1)} \right) \end{aligned}$$

Como a função está sujeita a retornos **crescentes**, é válido que:

$$q_2 = f(tK_1, tL_1) > t f(K_1, L_1)$$

Desse modo,

$$\left(\frac{t \cdot f(K_1, L_1)}{f(tK_1, tL_1)} \right) < 1$$

Isto implica que $CMe_2LP = CMe_1LP \cdot \alpha$, onde $\alpha < 1$. Logo, conclui-se que estando a função sujeita a retornos crescentes, os $CMeLP$ serão decrescentes. De certa forma, a existência de economias de escala decorre da propriedade de subaditividade de custos, cujo caso geral pode ser formalizado como:

$$\sum_{i=1}^n C(q_i) > C\left(\sum_{i=1}^n q_i\right)$$

sendo q_1, \dots, q_n um vetor de produção. Isto significa dizer que é mais barato produzir vários produtos juntamente do que produzi-los separadamente.

De forma análoga, é possível demonstrar que se a função está sujeita a retornos decrescentes, os $CMeLP$ serão crescentes e, por fim, uma função sujeita a retornos constantes gera $CMeLP$ constantes.

Admitindo uma tecnologia perfeitamente divisível,³ o $CMeLP$ possui o formato como mostra o Gráfico 3.2.

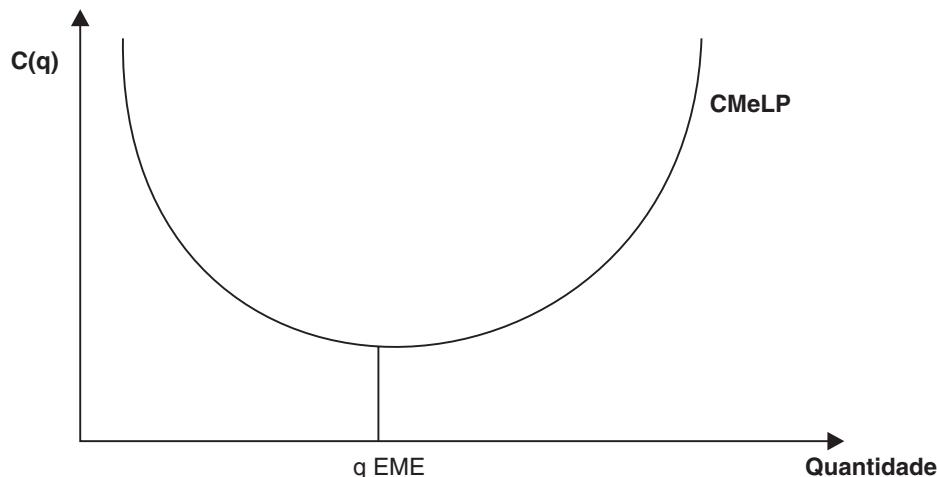


Gráfico 3.2 Curva de $CMeLP$ com formato em U.

A teoria tradicional dos custos, que admite ser a curva de $CMeLP$ a envoltória inferior das curvas de $CMeCP$,⁴ utiliza esse formato em U da curva de $CMeLP$ assumindo que as economias de escala existem até um determinado tamanho da planta produtiva, que é conhecido como tamanho ótimo da planta, no qual, então, todas as possíveis economias de escala são exploradas. Se a planta produtiva cresce acima desse tamanho ótimo, passam, então, a existir deseconomias de escala, geralmente decorrentes das ineficiências gerenciais e administrativas. Ao postular uma curva de $CMeLP$ em formato de U, a teoria tradicional dos custos considera implicitamente a forte hipótese de que a planta produtiva é completamente inflexível, uma vez que qualquer aumento da produção acima do tamanho ótimo, por menor que seja, acarreta crescimento substancial dos custos.

Para muitos autores, a curva de $CMeLP$ em formato de U é apenas uma curva teórica, não sendo verificada na prática, pelo menos de forma frequente. De fato, como veremos ao fim deste capítulo, existe uma crescente evidência empírica que sustenta o formato da curva de $CMeLP$ em L, como mostrado no Gráfico 3.3, e, sobre isso, houve então um grande debate para questionar a validade da curva em formato de U. Argumentou-se que as deseconomias de escala derivadas das ineficiências gerenciais, apontadas pela teoria tradicional como as responsáveis pelo aumento dos custos após o tamanho ótimo da planta, poderiam ser evitadas a partir da implementação de modernos métodos de gerência. Além disso, foi apontado também que mesmo se as deseconomias gerenciais de fato aparecessem (para elevadas escalas produtivas), elas seriam insignificantes em relação às economias de escala – que poderiam derivar de fontes, como veremos a seguir, no nível da planta, da multiplanta, ou da multiprodução – de forma que os custos totais por unidade produzida se manteriam constantes. Na defesa de uma curva de $CMeLP$ em formato de L, muitos autores argumentaram ainda que as deseconomias de escala podem até ser significativas para elevadas escalas produtivas, o que resultaria, de fato, em aumento dos $CMeLP$; mas isso ocorreria num nível de produção tão elevado que estaria fora da área relevante de produção, podendo, portanto, ser ignorado. As fontes de economias e deseconomias de escala, citadas neste debate empírico como responsáveis pelo aumento e redução dos $CMeLP$, serão discutidas nos itens seguintes deste capítulo.

Se aceitamos a possibilidade da curva de $CMeLP$ em formato de L, como representada no Gráfico 3.3, é importante notar que, nesse caso, existe uma escala mínima eficiente (EME) da planta, em vez de um único tamanho ótimo da planta como apresentado na curva de $CMeLP$ em formato em U. A EME corresponde ao nível da planta onde todas as economias de escala possíveis são exauridas e representa, portanto, a menor quantidade de produto (q^*) possível de ser obtida, de forma que o $CMeLP$ seja minimizado. Ou seja, para uma curva de $CMeLP$ com formato em L, para $q < EME$, a curva de $CMeLP$ é decrescente; para $q > EME$ a curva de $CMeLP$ é constante.

Alguns autores propuseram uma mescla entre as curvas de $CMeLP$ em formatos de U e L, como representado no Gráfico 3.4. Nesse caso, existiriam economias e deseconomias de escala, respondendo, respectivamente, pelas fases decrescentes e crescentes da curva de $CMeLP$. No segmento plano da curva para uma determinada faixa de produção (entre os níveis q_1 e q_2), as economias e deseconomias de escala praticamente se igualariam, determinando, pois, uma faixa de custos constantes por unidade produzida. Essa faixa de custos constantes corresponderia, então, à existência de uma reserva de capacidade que seria planejada pela empresa, de forma a lhe conceder o máximo de flexibilidade na operação, sem incorrer em aumento de custos.

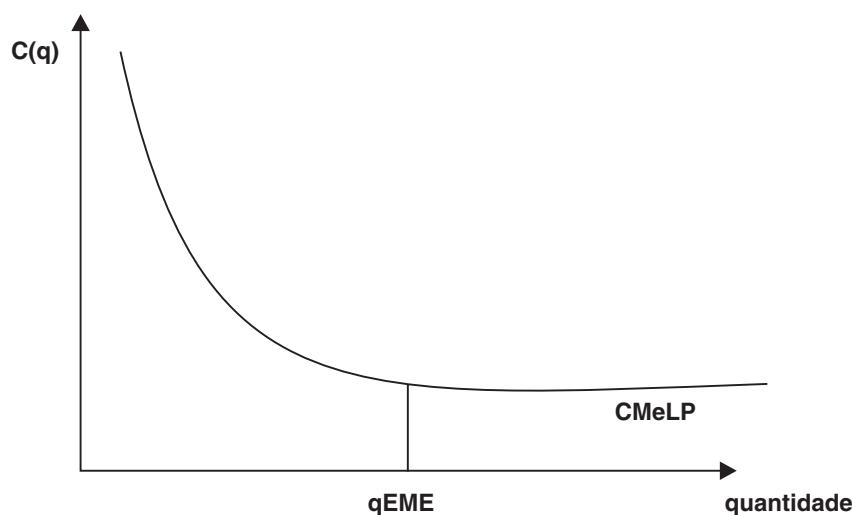


Gráfico 3.3 Curva de $CMeLP$ em formato de L.

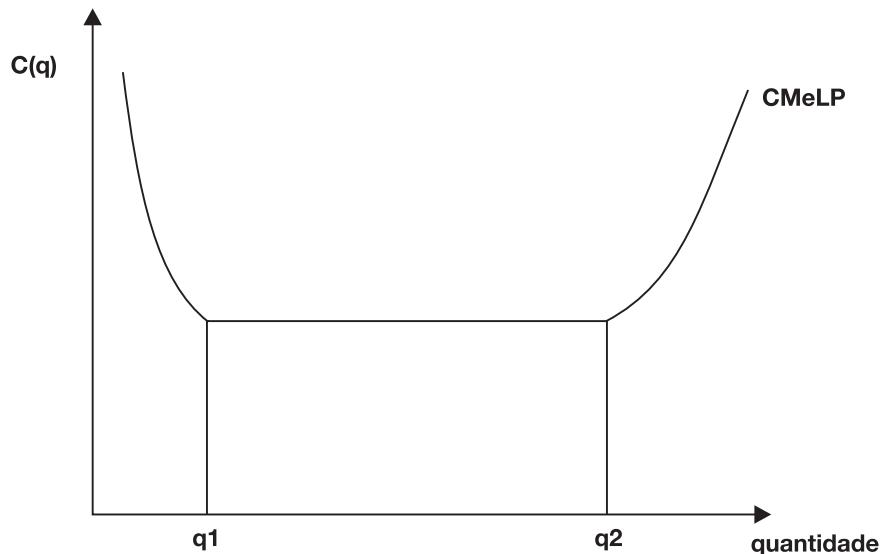


Gráfico 3.4 Curva de $CMeLP$ com segmento horizontal.

QUADRO 3.2 MONOPÓLIO NATURAL

A estrutura de custos determina em grande medida a estrutura de mercado. A *EME* da planta produtiva, quando comparada com o tamanho do mercado, é uma importante medida para verificarmos quantas empresas podem operar na indústria.

O monopólio natural ocorre quando é eficiente para apenas uma empresa suprir a demanda do mercado. Nesse caso, os custos totais de produção aumentariam se duas ou mais empresas operassem, em vez de uma.

No caso do monopólio natural, a estrutura de custos é caracterizada por economias de escala em todos os níveis da faixa relevante de produção (região até o ponto onde a curva de demanda de mercado cruza com a curva de $CMeLP$). Dessa maneira, a *EME* da planta é tão grande que supre toda a demanda do mercado.

As indústrias de infraestrutura são bons exemplos de monopólio natural. No Brasil, os setores de energia elétrica e telecomunicações foram desenvolvidos com uma estrutura de monopólio natural, de propriedade estatal (monopólios estatais). Isso ocorreu em função do consenso de que esta seria a estrutura mais eficiente para desenvolver tais setores no Brasil. A propriedade estatal era justificada em função da necessidade dos altos investimentos que o setor privado não poderia fazer. A operação de mais de uma empresa seria, assim, ineficiente.

Na média, a empresa desejaria operar entre os níveis mínimo (q_1) e máximo (q_2) da região onde os $CMeLP$ são minimizados e, sendo assim, quaisquer níveis de operação abaixo ou acima dos níveis mínimo e máximo, respectivamente, seriam tamanhos subótimos. Leia o Quadro 3.2.

3.4.1 FONTES DE ECONOMIAS DE ESCALA

O debate empírico anteriormente relatado demonstra não haver um consenso sobre o formato da curva de $CMeLP$. Entretanto, como pudemos observar, em todas as possibilidades de curva anteriormente demonstradas havia duas características comuns: a existência de um segmento decrescente, indicando, portanto, a presença de economias de escala, e a existência de uma escala mínima eficiente (*EME*), onde então as economias de escala se esgotam. Essas características tornam-se, por conseguinte, propriedades que devem ser mais profundamente examinadas.

De fato, existem muitas razões empíricas para acreditarmos que os $CMeLP$ sejam decrescentes, pelo menos inicialmente, à medida que a quantidade produzida no longo prazo se eleva, e que há um ponto em que as mesmas economias de escala se esgotam. Com relação a isso, a discussão sobre as estruturas de custo de longo prazo consiste, sobretudo, no exame empírico das fontes de economias de escala e deseconomias de escala.

Iniciaremos pelas economias de escala. As economias de escala estão, de um modo geral, associadas a dois tipos de fontes, quais sejam, as *economias de escala reais* e as *economias de escala pecuniárias*.

Consideremos o custo de produção representado pela seguinte expressão:

$$C(q) = \sum_{i=1}^n x_i p_i,$$

onde x_i é a quantidade do fator i , e p_i é o preço do fator i .

As economias de escala são ditas *reais* se o fator que as explica é o aumento proporcionalmente menor na quantidade média de fatores produtivos utilizados quando há um aumento da produção. Sendo assim, quando ocorre:

$$C(tq) < \sum_{i=1}^n t \cdot x_i p_i, \text{ sendo } t > 0,$$

temos, então, *economias de escala reais*, pois muito embora a produção esteja crescendo t vezes, a quantidade de insumos utilizados não cresce na mesma proporção, e sim em uma proporção inferior.

As economias de escala são ditas *pecuniárias* se o fator que as explica é uma redução no preço pago pelo insumo (p_i). Nesse caso, os custos da empresa se reduzem, mas não em resposta a mudanças reais no método de produção. Vale ressaltar que a redução de preços do insumo (p_i) e, portanto, a obtenção de ganhos de economias de escala pecuniárias por parte da empresa, reflete em geral ganhos de economias de escala reais por parte do fornecedor, uma vez que quanto maior é a demanda da empresa por fatores produtivos, menores serão os custos para o fornecedor produzi-los e fornecê-los. O nosso exame a respeito das fontes de economias de escala a ser feito aqui se deterá na análise das fontes de economias *reais*.

Existem quatro fontes principais de economias de escala reais.

1. Ganhos de especialização

A redução dos *CMeLP* com o aumento da produção pode ser primeiramente explicada pelos ganhos de especialização.

Essa fonte, que se observa no nível do produto, é uma das mais tradicionais, tendo sido ressaltada como ponto central por Adam Smith em seu importante trabalho *A Riqueza das Nações*, de 1776. Com uma maior quantidade de produto, maior poderá ser a divisão do trabalho, e mais os trabalhadores e máquinas poderão se especializar. Os trabalhadores adquirirão, pois, maior habilidade em suas funções e, com máquinas especializadas, maior será a sua produtividade, gerando, por conseguinte, menores custos. O Quadro 3.3 disposto a seguir nos traz exemplos a respeito dos ganhos de especialização produtiva e sua relação com comportamento do *CMeLP*.

2. Indivisibilidade técnica

A segunda fonte de economia de escala se relaciona com o tamanho dos equipamentos industriais, sendo, portanto, observável no nível da planta produtiva. Muito embora seja possível aumentar a quantidade do fator K , não é viável dividí-lo, uma vez que suas unidades estão definidas discretamente. Desse modo, como nem sempre é possível comprar equipamentos com um tamanho exato para produzir a quantidade de produto exatamente necessária, possíveis subutilizações do equipamento podem servir para uma futura expansão produtiva. Dessa forma, haveria, então, uma expansão produtiva a taxas constantes, ocasionando, portanto, *CMe* decrescentes. Assim, para cada tamanho de equipamento industrial é provável encontrar retornos crescentes decorrentes da maior utilização deste equipamento até o esgotamento de sua capacidade.

QUADRO 3.3 ESPECIALIZAÇÃO E REDUÇÃO DOS CMeLP

Dois exemplos ilustram como a especialização pode contribuir para a redução dos *CMeLP*. O primeiro é dado por Adam Smith, que analisa as vantagens da especialização na manufatura de alfinetes. Adam Smith afirmava que um trabalhador não qualificado para tal negócio, nem familiarizado com a utilização dos equipamentos usados, poderia fazer poucos alfinetes por dia. No entanto, na época em que o autor estava escrevendo *A Riqueza das Nações*, verificou-se uma tendência a uma divisão do trabalho para a produção do alfinete. Nessa divisão, o processo de produção foi repartido em “pequenas tarefas” (aproximadamente 18 etapas), onde cada trabalhador especializado passou a ser responsável por uma das etapas, resultando num aumento da produtividade dos trabalhadores e, consequentemente, na redução do *CMeLP*. No caso que ele descreve, dez trabalhadores especializados (alguns desempenhando duas ou três funções), trabalhando na produção do alfinete, produzem aproximadamente 48 mil alfinetes por dia.

O segundo exemplo é dado a partir de Henry Ford, que se tornou o maior produtor de automóveis (e provavelmente o mais bem-sucedido) no início do século XX, a partir do desenvolvimento da produção em massa. Ele adaptou a linha de montagem para carros padronizados e baratos numa série de funções desempenhadas por trabalhadores especializados. Foi obtida uma economia significativa a partir da redução do custo por unidade produzida apesar dos salários pagos terem sido consideravelmente altos em relação à média da época. Dessa maneira, ele criou, em 1914, um plano de participação nos lucros que distribuiu mais de US\$ 30 milhões anualmente para seus empregados.

3. Economias geométricas

A terceira fonte de economia de escala é também relacionada com o tamanho do equipamento industrial, sendo, portanto, observado também no nível da planta produtiva. Essa fonte diz respeito às propriedades geométricas da unidade processadora. Em casos como das indústrias de processo químico e metalúrgico (como, por exemplo, o refino de petróleo, cimento, indústria química e geração de energia elétrica), a mais importante fonte de economias de escala no nível planta específico decorre da expansão do tamanho individual das unidades processadoras. O produto dessas unidades tende a ser proporcional ao volume da unidade, ao passo que o custo associado à produção é proporcional à área da superfície das unidades processadoras.⁵

Dois autores, John Haldi e David Whitcomb, realizaram um conhecido trabalho em 1967, em que procuraram estimar a relação entre o custo de capital (associado à compra de equipamentos ou à construção de plantas) e a sua capacidade (isto é, o limite superior de produção que pode ser obtido). Para realizar esse teste, os autores partiram de uma amostra de 687 tipos de equipamentos industriais, aplicando a seguinte equação exponencial:

$$C = aq^b,$$

onde C é o custo, q é a capacidade produtiva, a é uma constante, e b é o fator de escala. Se $b < 1$, existem economias de escala; se $b = 1$, existem retornos constantes de escala; e sendo $b > 1$, existem deseconomias de escala. Como resultado, Haldi e Whitcomb concluíram que 618 das máquinas da amostra (cerca de 90% do total) apresentaram retornos crescentes ao nível dos equipamentos, isto é, os custos não cresceram na mesma proporção que o crescimento da capacidade dos equipamentos, mesmo para capacidades elevadas. Foi possível então concluir que quanto maior a capacidade produtiva do equipamento (ou da planta), menores seriam os custos de aquisição associados. É importante ressaltar que este resultado não implica que os *CMeLP* declinam com a quantidade produzida, mas sim se reduzem com a compra de equipamentos (e construção de plantas). Assim, o teste verificou a existência de economias de escala no nível dos equipamentos e, portanto, da planta produtiva, onde se incluem fontes derivadas de indivisibilidades técnicas e as economias geométricas.

4. Economias relacionadas à lei dos grandes números

Uma outra relevante fonte de economia de escala, que se apresenta no nível da planta, está associada à lei dos grandes números: quanto maior for o tamanho da planta produtiva, sendo, portanto, maior o número de máquinas utilizadas, proporcionalmente menores deverão ser, por exemplo, o *staff* de manutenção e o número de peças de reposição necessário. Ou seja, a equipe e os materiais utilizados para reposição e conserto de peças destinados a manter qualquer nível de atividade produtiva diante da possibilidade de problemas técnicos cresce menos do que proporcionalmente em relação ao número de máquinas em operação.

Por exemplo, uma empresa pequena que utiliza apenas uma única máquina deverá manter duas máquinas para se precaver de possíveis defeitos. Uma empresa maior que utiliza um grande número de máquinas deve manter, como capacidade de reserva, apenas uma proporção das máquinas utilizadas. De forma análoga, o número de trabalhadores necessário para a realização de reparos não cresce proporcionalmente com o tamanho da escala.

Uma demonstração mais geral a respeito das economias de escala relacionadas à Lei dos Grandes Números está descrita no Quadro 3.4. O Quadro 3.5, por sua vez, apresenta uma forma prática de verificar a existência de economias de escala a partir do comportamento das funções custo.

QUADRO 3.4 ECONOMIAS DE ESCALA RELACIONADAS À LEI DOS GRANDES NÚMEROS

Consideremos inicialmente que: (1) p é a probabilidade de uma máquina apresentar defeito num período de produção; (2) n é o número de máquinas; e (3) cada funcionário da equipe de manutenção cuida de uma máquina quebrada em cada período. Se n é grande, o número esperado de defeitos nas n máquinas num dado período pode ser descrito por uma distribuição binomial com média np e variância $np(1-p)$. Para um $n \geq 30$, essa distribuição se aproxima de uma normal. Suponha que a empresa deseja providenciar uma equipe de manutenção grande o suficiente para resolver de forma imediata um defeito que venha a aparecer em uma das máquinas, excluindo a possibilidade de que um grande número de máquinas apresente defeitos simultaneamente. Considere que a chance de a empresa não obter sucesso nessa estratégia seja de apenas 5%. Dessa forma, a quantidade máxima do fator trabalho necessária será:

$L = p + 1,96\sqrt{np(1-p)}$. Então, a quantidade de trabalhadores de manutenção por máquinas será dada por: $L = np + 1,96\sqrt{\frac{p(1-p)}{n}}$, o que é uma função decrescente em n .

QUADRO 3.5 UMA MEDIDA DE ECONOMIAS DE ESCALA

Sabemos que para uma dada função de custos, a existência de economias de escala se observa quando os custos diminuem à medida que o nível de produto aumenta, quando então há custos médios decrescentes. Sendo assim, utilizando o conceito de elasticidade-custo poderíamos ter uma forma útil de medir economias de escala:

$$\varepsilon_c = \frac{(\Delta C / C)}{(\Delta Q / Q)}$$

Se $\varepsilon_c = 1$, então os custos aumentam proporcionalmente ao aumento no nível de produção; se $\varepsilon_c > 1$, então os custos aumentam proporcionalmente mais do que o aumento no nível de produto; finalmente se $\varepsilon_c < 1$, os custos aumentam menos rapidamente do que uma elevação do nível de produção. Dessa forma, é possível definir um *índice de economias de escala* do seguinte modo:

$$IEE = 1 - \varepsilon_c$$

se $\varepsilon_c = 1 \rightarrow IEE = 0$ não existem economias de escala

se $\varepsilon_c > 1 \rightarrow IEE < 0$ existem deseconomias de escala

se $\varepsilon_c < 1 \rightarrow IEE > 0$ existem economias de escala

3.4.2 ECONOMIAS DINÂMICAS X ECONOMIAS ESTÁTICAS

As fontes de economias de escala destacadas até aqui são *estáticas*, uma vez que a nossa análise procurou relacionar o comportamento do *CMeLP* com a quantidade produzida, não tendo sido feita referência alguma ao tempo de produção despendido. Quando passamos a incorporar o tempo, isto é, a variação da quantidade com o passar do tempo, o relevante passam a ser as fontes de economias de escala *dinâmicas*. São duas as principais fontes de economias de escala dinâmicas:

1. Economias de reinício (set up)

Os custos de reinício estão envolvidos na operacionalização de equipamentos “multitarefas” para o desempenho de uma função ou produto específico. Como exemplo de empresas que utilizam equipamentos com várias funções podemos citar uma empresa produtora de automóveis. Tal empresa deve usar um equipamento de estamparia de metal que produz tanto o chassi quanto vários componentes, tais como portas, rodas, carroceria etc. O equipamento deve ser reinicializado toda vez que uma parte específica do produto final é produzida, uma vez que há a necessidade de regular a máquina para a execução de outra tarefa. O processo de ajuste do equipamento a um novo reinício envolve custos e tempo, visto que a regulagem da máquina implica que ela esteja fora de funcionamento. Nesse caso, quanto maior a produção, por mais tempo a máquina poderá operar com a mesma regulagem, o que irá reduzir os custos associados à perda de tempo necessária ao reinício da operação.

2. Economias de aprendizado

Quando um novo processo ou produto é iniciado, origina-se também um processo de aprendizado. Em geral, o processo de produção das primeiras unidades envolve um certo nível de tentativa e erro uma vez que novos métodos são estabelecidos e as pessoas são treinadas. Assim, os custos iniciais por unidade são geralmente altos. À medida que a produção aumenta, os trabalhadores se tornam mais rápidos e precisos em suas funções; as máquinas são adaptadas e os ajustamentos são feitos na direção do melhor sistema de produção. Recentemente, os economistas têm chamado atenção para o fato de que a primeira empresa a se mover (a primeira que entra no mercado e adquire tamanho substancial e experiência com um produto) terá vantagens sobre as empresas seguidoras em função das economias de aprendizado.

O resultado do processo descrito acima é uma curva de aprendizado, que apresenta um declínio no custo médio corrente na medida em que a produção total aumenta. A inclinação da curva de aprendizado depende de cada produto ou processo de produção específico, e um possível exemplo pode ser visto no Gráfico 3.5. Observe no gráfico que a curva de aprendizado está relacionada à produção acumulada, uma vez que, nesse caso, a nossa análise deve considerar o aspecto dinâmico da produção, isto é, a variação da quantidade produzida no tempo.

3.4.3 ECONOMIAS DE ESCOPO

Até agora consideramos que a planta produtiva da empresa produzia apenas um tipo de produto. Vamos agora analisar a possibilidade de uma produção conjunta de duas ou mais mercadorias. Na prática, como ressaltado no Capítulo 2, a maior

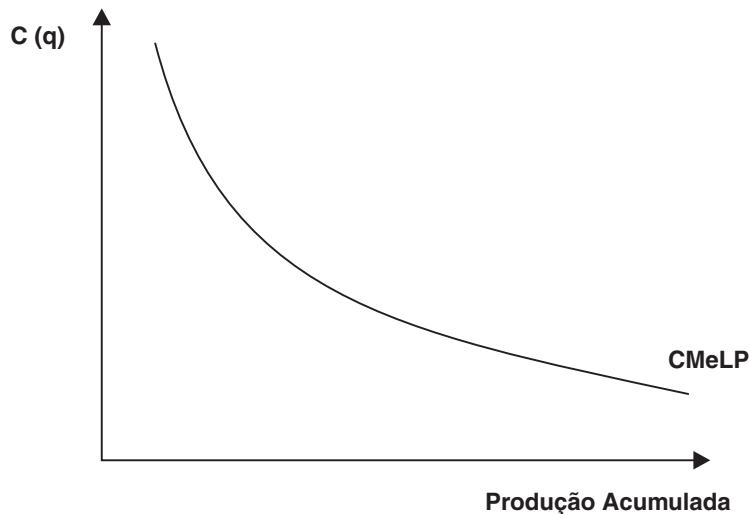


Gráfico 3.5 Curva de Aprendizado.

parte das plantas industriais produz vários produtos, cada um deles com sua própria estrutura de custos. Nesse caso, o custo de produção de um produto em particular depende não somente do seu próprio volume de produção, mas também do tamanho da planta onde o produto é feito. Uma possível razão para a produção conjunta, isto é, produção de mais de um produto numa mesma planta, é a existência de economias de escopo, que podem ser formalmente definidas como:

$$C(q_a, q_b) < C(q_a, 0) + C(0, q_b).$$

Isso significa dizer que o custo de produzir os produtos q_a e q_b conjuntamente é menor do que o custo de produzi-los separadamente. Ou seja, algumas empresas conseguem reduzir seus custos médios com a diversificação de produtos. O aumento da variedade de produtos no portfólio provoca uma redução em seu custo médio. É interessante observar que esta definição de economias de escopo decorre do conceito de subaditividade de custos, o que nos permite concluir que a existência de economias de escopo depende em grande medida das economias de escala.

Na prática, é possível identificar três fontes de economias de escopo, conforme descrito a seguir.

1. Existência de fatores comuns

A primeira fonte de economia de escopo é verificada quando, para a produção de um bem, é necessária a aquisição de um fator de produção comum, isto é, que se adquire uma única vez. Uma vez que tal fator tenha sido comprado, sua posterior utilização na produção de outro bem é praticamente gratuita.

Como exemplo, podemos citar a compra, por uma empresa, de um gerador de eletricidade para permitir que os picos de demanda por energia para a produção de um determinado bem sejam atendidos. Nesse caso, a produção de mais de um bem não demandará um novo gerador. De fato, a diversificação da produção implicará a redução do custo médio de cada produto, dado que a capacidade de geração de energia já foi instalada.

2. Existência de reserva de capacidade

A segunda fonte de economia de escopo ocorre quando um insumo (ou alguns insumos) pode(m) ser compartilhado(s) para produzir vários produtos em função de seu processo produtivo. Se existe capacidade ociosa⁶ na planta instalada para a produção da principal linha de produto, a empresa tem um incentivo para procurar outros produtos que possam utilizar a reserva de capacidade. Esse tipo de economia de escopo se diferencia da anterior, já que a capacidade utilizada para a produção de um determinado bem não pode ser compartilhada para a produção de um outro, diferentemente do bem comum adquirido, como visto anteriormente. Somente a capacidade ociosa pode ser usada na produção de um bem que tenha um processo produtivo relativamente similar ao principal produto da empresa. Assim, esse tipo de economia de escopo estaria associado a uma das fontes de economias de escala reais, seja qual for a propriedade de indivisibilidade técnica anteriormente descrita. Vale ressaltar ainda que o entendimento da existência de capacidade ociosa como um tipo de economia de escopo requer o exame das condições de mercado bem como dos custos de produção.

3. Complementaridades tecnológicas e comerciais

A terceira fonte de economias de escopo surge de complementaridades tecnológicas e comerciais na produção de alguns bens, constituindo-se na fonte mais importante em termos de desdobramentos econômicos. Esse tipo de economia de escopo

pode gerar sinergias na produção de alguns bens e ocorre quando os produtos apresentam similaridades em termos de base técnica e/ou de mercado. A utilização de insumos comuns e a propaganda com os produtos são importantes fontes desse tipo de economias de escopo. Esta última ocorre porque, na medida em que uma empresa realiza custos com propaganda de um determinado produto e este passa a ser reconhecido no mercado por sua qualidade, a empresa poderá incorrer em menores custos com propagandas de outros produtos.

3.4.4 ECONOMIAS NO NÍVEL DA MULTIPLANTA

O exame das economias de escala feito até aqui considerou a empresa limitada no nível de uma única planta. Desse modo, nossa análise terminou por omitir um aspecto da realidade de operação das grandes empresas, qual seja o aspecto das operações multiplanta. Em um importante trabalho, F. Scherer e outros pesquisadores apresentaram em 1975 evidências a respeito da relevância e existência de empresas multiplantas, tendo concluído para uma amostra de 155 setores da indústria manufatureira (a quatro dígitos) norte-americana, que em mais da metade deles as empresas líderes possuíam mais de quatro plantas produtivas. A empresa transnacional é um exemplo particularmente relevante das operações multiplantas, o que será tratado de forma detalhada no Capítulo 17.

Se na prática é comum a muitos setores da indústria a existência de empresas que operam com múltiplas plantas, então é razoável supor que as empresas operem dessa forma porque podem tirar proveito de economias de escala que não estão disponíveis no nível de uma planta apenas. Quais seriam então as razões para a existência de operações multiplantas? É possível identificar quatro fatores principais que poderiam responder a essa questão, todos eles relacionados com o custo das empresas.

3.4.4.1 *Economias da duplicação*

A primeira fonte de economias no nível das multiplantas decorre da possibilidade de adição de nova capacidade produtiva ao longo do tempo. Considerando um mercado que cresce a uma taxa incremental ao longo do tempo, a empresa deve então planejar seus investimentos para se ajustar à demanda. Cada planta uma vez construída tem uma dada escala e custos associados. Supondo que a propriedade das economias geométricas implica que o custo de capital específico à planta cresce menos que proporcionalmente com o tamanho, então a empresa deve lidar com diferentes custos. Se a empresa tem que acompanhar o crescimento da demanda, ela pode escolher entre duas opções: realizar frequentes adições de capacidade em pequena escala ou adições menos frequentes numa escala maior. A primeira opção reduz o grau de excesso de capacidade que pode vir a se verificar na empresa, o que, entretanto, é compensado pelo maior custo unitário do capital. Por outro lado, adições menos frequentes implicariam maior capacidade ociosa.

3.4.4.2 *Custo de transporte*

A segunda fonte de economias de escala no nível multiplanta decorre dos custos de reunião dos insumos e dos custos de distribuição dos produtos. A existência de mercados geográficos dispersos e de centros fornecedores de insumos longínquos acaba acarretando significativos custos de transporte. Desse modo, a operação multiplanta seria uma forma de minimizar os elevados custos de transporte associados à operação nesses mercados.

3.4.4.3 *Alcance de especialização no nível das multiplantas*

Outra razão para a operação no nível das multiplantas é a possibilidade de alcançar especialização da produção em diferentes plantas. Assim, por exemplo, se a empresa multiplanta opera em mercados cujos padrões de comportamento da demanda são inversamente correlacionados, ou não correlacionados, a estratégia de especialização produtiva em direção às várias plantas daria à empresa uma maior segurança, reduzindo, assim, a variância do comportamento das suas receitas, diminuindo, portanto, o risco de operação em mercados diferentes. Além disso, ao alcançar a especialização no nível das várias plantas produtivas, a empresa pode reduzir os custos de reinício, já descritos anteriormente, uma vez que existe a possibilidade de que cada planta opere mais tempo com a mesma regulagem em suas máquinas.

3.4.4.4 *Flexibilização da operação*

Empresas que operam no nível multiplanta possuem maior flexibilidade na operação, o que pode contribuir para a redução dos custos quando comparados com os custos referentes a uma única planta. Isso poderia ser explicado pela possibilidade de compensação produtiva (fluxo produtivo) entre as múltiplas plantas de uma mesma empresa no caso de ocorrência de flutuações produtivas entre elas, ou no caso de redução de demanda. Nesse último caso, haveria então a possibilidade de a empresa fechar

a planta produtiva com maior custo e operar somente com a(s) outra(s) planta(s), utilizando a capacidade instalada de modo mais eficiente.

3.4.5 DESECONOMIAS DE ESCALA

Como visto anteriormente, é claro o fato de que as economias de escala existem e que o $CMeLP$ cai com o aumento do volume produzido, do tamanho da planta (e/ou equipamentos), das operações multiplantas e multiprodutos das empresas. Entretanto, existem também razões para acreditar que o $CMeLP$ não se reduz indefinidamente, existindo, portanto, deseconomias de escala a partir de um determinado volume de produção.

Em geral, existem dois fatores principais que podem gerar deseconomias de escala:

1. Custos de transporte

O custo de entrega do produto (ou o custo de levar os consumidores até o local onde o serviço/produto será oferecido) pode limitar as economias de escala no nível de uma única planta produtiva (ou no nível de um complexo de plantas), na medida em que quanto maior o nível de produto, maior deverá ser a venda e, portanto, maior a necessidade de alcançar consumidores. Isso leva ao crescimento dos custos de transporte por unidade vendida, e a magnitude desse crescimento depende da influência de alguns fatores adicionais. Dentre eles, destacamos três: o primeiro diz respeito ao tamanho da planta produtiva em relação ao tamanho do mercado. No caso de a oferta da planta representar apenas uma pequena parcela da demanda, será possível que esta planta aumente suas vendas sem necessariamente expandir seus custos de transporte. O segundo dos fatores tem a ver com a natureza do sistema de preços. Os custos de transporte absorvidos por uma empresa aumentam com o nível de produto quando os preços são uniformes em todos os mercados onde ela opera, ou então quando o preço num mercado mais distante é estabelecido por um concorrente local que possua melhores condições competitivas. O terceiro fator diz respeito à possibilidade de a empresa transferir os custos de transporte para os consumidores. Se isso ocorre, os custos de transporte crescem lentamente com o nível de produto.

A combinação desses fatores resulta, em geral, em custos médios de transporte ($CMeT$) crescentes.

2. Deseconomias gerenciais

Como havíamos visto no início do capítulo, a teoria tradicional da empresa fornece argumentos de que as deseconomias gerenciais constituiriam o fator responsável pela elevação dos $CMeLP$ logo após a exaustão das economias de escala, o que explicaria a curva em formato de U. De acordo com a teoria tradicional da empresa, isso se explicaria porque após um determinado ponto um aumento na equipe de gerência acarretaria aumento menos que proporcional na produção, o que causaria, então, o crescimento dos custos por unidade produzida no longo prazo. Esse decréscimo na eficiência gerencial poderia ser explicado primeiramente pelo fato de que após a empresa ter ultrapassado um tamanho ótimo, a equipe de gerência perderia o controle sobre o processo de decisão.

Em segundo lugar, a queda na eficiência gerencial seria decorrente da maior incerteza inerente ao comportamento da demanda e do processo de competição enfrentado pela empresa de grande porte.

Esses argumentos utilizados pela teoria tradicional para explicar que as deseconomias gerenciais seriam as responsáveis pelo crescimento do $CMeLP$ foram aos poucos sendo contestados. Alguns autores procuraram demonstrar que a elevação da escala da planta produtiva não é necessariamente acompanhada por uma inficiência gerencial. A descentralização do processo de tomada de decisão, a mecanização de várias funções gerenciais, o sistema de relatórios regulares entre os vários níveis hierárquicos, o uso de computadores e similares para o processamento de informação, tenderiam a amenizar as deficiências de uma complexa organização, de modo que as deseconomias gerenciais não seriam um problema no mundo industrial moderno. Ademais, o surgimento da empresa multidivisional (ver o Capítulo 2) como uma inovação organizacional também teria contribuído para “adiar” o ponto em que as deseconomias se manifestariam.

Como vemos, existe uma controvérsia a respeito de os fatores gerenciais constituírem ou não uma fonte de deseconomias de escala. Entretanto, como será visto na próxima seção, a evidência empírica demonstra que a curva de $CMeLp$ tende a considerar, na prática, um formato em L. Isso nos fornece algum suporte para acreditar que esses fatores não parecem, na prática, representar uma fonte de deseconomias de escala ao ponto de elevar os custos unitários de produção no longo prazo.

3.4.6 DEBATE EMPÍRICO SOBRE AS CURVAS DE $CMeLP$

Após a discussão a respeito das fontes de economias e deseconomias de escala, é interessante examinar como são feitos os exames empíricos sobre o formato da curva de $CMeLp$.

Até hoje foram vários os estudos empíricos realizados sobre custos, tendo sido possível identificar, de acordo com a metodologia utilizada, os seguintes grupos: estudos baseados em custos de engenharia, estudos baseados em análise estatística e estudos baseados nas técnicas do “sobrevivente”.

1. Estudos baseados nos custos de engenharia

Esse método procura examinar as relações técnicas entre insumos e nível de produto dispostas na função de produção, utilizando para tanto as informações obtidas via aplicações de questionários e entrevistas entre engenheiros e pessoal, responsáveis pelo planejamento e projeto de novas plantas e unidades de produção. De posse dessas informações, os pesquisadores adeptos dessa metodologia procuram estimar as relações de custo para cada equipamento individual ou processo, integrando-os posteriormente por intermédio de um modelo econométrico.

Vários estudos utilizaram esse tipo de metodologia para o exame das economias de escala curvas de custo. Dentre eles, um estudo importante realizado por Scherer e outros pesquisadores em 1975 teve como objetivo estimar as economias de escala no nível da planta em 12 setores da indústria nos EUA e em outros cinco países (Canadá, Alemanha, França, Suécia e Inglaterra). Os dois principais resultados encontrados foram os seguintes: a *EME* é pequena em relação ao tamanho do mercado; a curva de *CMeLP*, na maior parte dos setores, possui uma reduzida inclinação na sua parte decrescente.

Se, por um lado, essa metodologia possui pontos positivos, como, por exemplo, o fato de considerar a melhor aproximação do que os economistas apontam como custos de produção, por outro, muitos ressaltam alguns problemas. O primeiro deles é o fato de que os custos de engenharia não consideram os fatores gerenciais, não sendo, portanto, possível a essa metodologia a mensuração das deseconomias de escala que poderiam advir dos problemas gerenciais. O segundo problema se refere ao elevado custo associado à aplicação dessa metodologia, uma vez que o método requer um elevado esforço de entrevistas.

2. Estudos baseados em análises estatísticas

Esse método de estudo sobre custos consiste na utilização de técnicas de regressão múltipla – numa análise *cross section* ou de séries temporais, para relacionar custos, como variável dependente, com variáveis independentes que traduzam aspectos relevantes do processo de produção da empresa. Em geral, as variáveis independentes utilizadas são as seguintes: volume de produção; razão da capacidade utilizada; diferenças entre os preços dos insumos; número de produtos diferentes ofertados (o que refletiria as economias e deseconomias de escopo); o volume acumulado de produção; diferenças em termos da “idade” do estoque de capital (refletindo, assim, a tecnologia utilizada) e outros mais. Como exemplo, Johnston, por meio de um estudo desse tipo em 1960, procurou estimar a curva de custos e a *EME* da escala produtiva para empresas de diferentes setores da manufatura e setores de utilidade pública nos EUA, como telecomunicações e energia elétrica.

Como resultado, para os setores de utilidade pública Johnston encontrou elevadas *EME*, o que era de se esperar, uma vez que esses setores eram considerados monopólios naturais. Para os setores da indústria de transformação, o pesquisador encontrou como resultado uma curva de *CMeLP* que tende a um formato em L, na qual para um reduzido nível de produção existem economias de escala, mas para um nível de produção maior essas economias se esgotam e o custo por unidade produzida torna-se constante.

As críticas a esse tipo de metodologia se referem à qualidade dos dados utilizados, já que, na maior parte das vezes, utilizam-se dados contábeis que não refletem os custos de oportunidade tão importantes para os economistas, como visto na introdução deste capítulo.

3. Estudos baseados na técnica do “sobrevivente”

Esse tipo de técnica de estudo foi desenvolvido por George Stigler em 1958 e tomou como base o princípio de que os tamanhos de empresas e plantas que “sobrevivem” e contribuem de forma crescente ao longo do tempo com a oferta da indústria são tamanhos eficientes. Desse modo, esse tipo de estudo procura observar somente a evolução do tamanho das empresas na indústria em diferentes pontos do tempo. A partir desse tipo de observação, os pesquisadores dessa linha de estudo procuram estimar o formato das curvas de custo na indústria em questão.

Muitos pesquisadores aplicaram essa “técnica da sobrevivência” para examinar uma série de indústrias. Como resultado, esses testes indicaram que a curva de *CMeLP* é horizontal para elevados níveis de produção (e tamanho da planta). Da mesma forma, os estudos indicaram que, na maior parte das vezes, a *EME* das plantas produtivas representa, em média, uma parcela de 2% do tamanho do mercado.

As principais críticas a esse tipo de metodologia se referem primeiramente à qualidade da observação feita, uma vez que muitos argumentam que essa técnica ignora as medidas de custos reais da empresa, ou qualquer outra medida que traduza a eficiência da empresa. Ademais, é possível argumentar que os critérios para distinção entre empresas sobrevidentes e não sobrevidentes são um tanto arbitrários e, além disso, os padrões de sobrevivência nem sempre são estáveis ao longo do tempo.

Como se viu, mesmo possuindo problemas metodológicos, os mais variados métodos de pesquisa empírica apontam para uma mesma direção ao sugerir que a curva de *CMeLP* não possui o formato em U, como postulada pela teoria tradicional. Ao contrário, os estudos tendem a mostrar que no longo prazo as curvas de custo por unidade produzida diminuem com reduzidos níveis de produção e permanecem constantes para escalas de produção elevadas; o que significa que, na prática, a curva de *CMeLp* tende a possuir um formato em L.

3.5 Conclusão

Este capítulo procurou apresentar os conceitos básicos a respeito da estrutura de custos, bem como discussões em perspectivas de curto e de longo prazos. A compreensão teórica desse tema é de suma importância para a análise dos comportamentos decisórios da empresa.

Vejamos alguns exemplos. A estrutura de custos constitui-se em um dos principais determinantes dos preços nas várias estruturas de mercado, além de determinar, em grande medida, a própria estrutura de mercado, uma vez que quanto maiores as economias de escala num dado tamanho do mercado, menor será o número de empresas na indústria em questão. Na forma de vantagem absoluta de custos ou na forma de *EME*, os custos determinam, de maneira significativa, a magnitude das barreiras à entrada. Além disso, um conhecimento da estrutura de custos de uma indústria é de extrema importância para a política de regulação governamental e defesa da concorrência.

Esses exemplos demonstram, portanto, a importância do conhecimento teórico a respeito dos custos para compreender e analisar, na prática, a dinâmica industrial.

Notas

1. Se $C(q)$ é o custo total de produzir q unidades, o custo marginal é calculado da seguinte forma: $CMg = dC(q)/dq$.
2. Assumindo que o $CMg = \frac{\Delta CV(q)}{\Delta q}$ – uma vez que a variação do custo total em relação à variação do nível de produto independe do custo fixo – é possível então reescrevê-lo como $CMg = \frac{w\Delta L}{\Delta q}$. Admitindo que $PMgL = \frac{\Delta Q}{\Delta L}$, então $CMg = \frac{w}{PMgL}$. Desse modo, o CMg apresenta uma relação inversa com a PMgL. O aumento inicial no nível de produção é acompanhado do crescimento da PMgL, ocasionando, portanto, uma redução do CMg. Como se trata de uma análise de curto prazo, por conta da existência do fator fixo (K), à medida que se utiliza uma quantidade cada vez maior do fator variável (L), a PMgL passa a diminuir. A lei dos rendimentos marginais decrescentes faz então o CMg crescer.
3. Uma tecnologia é dita perfeitamente divisível quando é possível variar continuamente a proporção dos fatores produtivos, acarretando, assim, a possibilidade de variações contínuas dos níveis de produto.
4. De acordo com a teoria tradicional dos custos, a curva de *CMeLP* é derivada das curvas de *CMeCP*; cada ponto da curva de *CMeLP* corresponde a um ponto da curva de *CMeCP*, que é tangente à curva de *CMeLP* naquele ponto. Dada a flexibilidade dos fatores de produção no longo prazo, o *CMeLP* é sempre pelo menos tão baixo quanto o *CMeCP*; ou seja, é possível que a empresa produza determinada quantidade no longo prazo com menor custo do que no curto prazo.
5. Considere r como o raio da unidade processadora. Desse modo, as economias geométricas podem ser assim traduzidas: o custo de fabricação da unidade aumenta em r^2 enquanto o produto dela resultante (volume) aumenta em r^3 .
6. A existência de capacidade ociosa pode ocorrer em função do tamanho do mercado ser menor do que a capacidade de produção de uma planta indivisível. Alternativamente, ela pode surgir como resultado da competição imperfeita, onde a maximização de lucros ocorre num nível de produto abaixo da capacidade instalada.

Bibliografia

- Haldi, J. e Whitcomb, D. (1967) Economies of Scale in Industrial Plants, *Journal of Political Economy*, pp. 373-385.
- Hay, D. e Morris, D. (1991) *Industrial Economics: Theory and Evidence*, 2^a ed. Oxford: Oxford University Press.
- Johnston, J. (1960) *Statistical Cost Analysis*, 1^a ed. Nova York: McGraw-Hill.
- Koutsoyiannis, A. (1979) *Modern Microeconomics*, 2^a ed. Londres: MacMillan Education.
- Scherer, F. et alii (1975) *The Economics of Multi-plant Operations: An International Comparisons Study*, 1^a ed. Cambridge: Harvard University Press.
- Scherer, F. e Ross, D. (1990) *Industrial Market Structure and Economic Performance*, 3^a ed. Boston: Houghton Mifflin.
- Sheperd, W. (1996) *The Economics of Industrial Organization*, 4^a ed. New Jersey: Prentice-Hall.
- Stigler, G. (1958) The Economies of Scale, *The Journal of Law and Economics*, pp. 54-71.

O Modelo Estrutura, Conduta e Desempenho e seus Desdobramentos

Lia Hasen clever e Ricardo Torres

4

4.1 Introdução

O nascimento da disciplina Economia Industrial é contemporâneo ao nascimento dos primeiros sistemas industriais na segunda metade do século XIX. Posteriormente, a disciplina passa a ser denominada, nos Estados Unidos, de organização industrial, como visto na Introdução do livro. Ela se inscreve na tradição de análise econômica anglo-saxônica e, em seguida, esta disciplina conheceu um enorme sucesso nos meios econômicos europeus e americanos. São ilustrativos os desenvolvimentos aportados por J. B. Say, na França; F. List, na Alemanha; T. Veblen e J. M. Clark, nos Estados Unidos; e Alfred Marshall, na Inglaterra.

Estes economistas se interessavam por problemas reais do funcionamento do sistema capitalista, tendo a indústria como principal motor da economia, e focaram seus esforços na explicação para o aparecimento de novos setores industriais, a formação dos cartéis e o impacto da concentração sobre o funcionamento dos mercados, a história dos trustes nos Estados Unidos, a formação dos preços e a decisão de localização das empresas. Segundo esses autores, esses assuntos são tratados de forma insatisfatória pelas teorias neoclássicas de equilíbrio parcial e geral para explicarem qual a natureza e o funcionamento real das empresas, e quais os mecanismos de coordenação de suas atividades, incluindo o funcionamento dos mercados. Alguns destes questionamentos foram apresentados na conclusão do Capítulo 1 e desdobrados nos Capítulos 2 e 3 deste livro.

Em seus primórdios, a disciplina era essencialmente empírica. Estudava várias empresas e indústrias para tentar explicar as causas dos comportamentos desviantes das previsões do modelo neoclássico tradicional. O seu foco é principalmente o estudo da estrutura das indústrias. Como visto no Capítulo 2, por indústria entende-se o conjunto de empresas produzindo bens e serviços substitutos entre si e, portanto, em concorrência em um mercado. Por estrutura entende-se não somente as características morfológicas das indústrias (número e tamanho das empresas, grau de concentração da indústria, entre outras características), mas também os princípios de seu funcionamento (o padrão de concorrência vigente, por exemplo).

Progressivamente, os estudos foram se estruturando em torno de um arcabouço de análise empírica comum que procurava prever como as atividades produtivas são trazidas em harmonia com a demanda por bens e serviços, por meio de alguns mecanismos organizacionais, tais como o mercado, e como as variações e imperfeições no mecanismo organizativo afetavam o sucesso das indústrias em satisfazer o bem-estar econômico.

Nesta evolução, pode-se reconhecer a contribuição de vários autores na busca de conjuntos de atributos ou variáveis que influenciam o desempenho econômico e de construção de teorias detalhando as ligações entre esses atributos e o desempenho de cada indústria. O paradigma fundador da disciplina economia industrial, denominado Modelo Estrutura, Conduta e Desempenho (ECD), foi concebido por Edward E. Mansfield, da Universidade de Harvard, na década de 1930. Posteriormente, recebeu contribuições de diversos seguidores. Entre os mais relevantes estão Joe Bain e Frederic M. Scherer.

Este capítulo tem por objetivo expor esse paradigma fundador da disciplina economia industrial, sublinhando as potencialidades intelectuais que ele abre para o estudo das empresas e indústrias, suas contribuições empíricas e os seus limites.

4.2 Os Antecedentes, o Escopo e o Método

Duas constatações empíricas estão na base da disciplina de economia industrial. A primeira é que a estrutura dos mercados não corresponde àquela hipotetizada no modelo neoclássico de concorrência perfeita, também denominada de “mercado natural”. Segundo a teoria tradicional, apresentada no Capítulo 1, o número de empresas em um mercado deve ser igual ao tamanho do mercado dividido pelo tamanho ótimo que permite produzir ao custo total médio mínimo. Entretanto, os estudos empíricos realizados mostravam que o mais comum é se observar um grau de concentração das empresas bem superior ao da estrutura ótima do mercado em concorrência perfeita. Surge a questão de explicar porque isso ocorre e quais as consequências sobre o desempenho das indústrias para atingirem o bem-estar econômico.

A segunda constatação empírica é que a taxa de lucro varia entre os diversos setores da economia. Assim, ao contrário do que previa a teoria econômica neoclássica, que estipulava que haveria uma equalização das taxas de lucros em todos os mercados, caso houvesse livre mobilidade dos fatores de produção, na realidade, há uma divergência das taxas de lucros. Portanto, ou a predição da teoria neoclássica estaria incorreta ou uma de suas hipóteses não seria válida. A segunda opção foi explorada pela economia industrial, que revelou haver barreiras à mobilidade dos fatores de produção e à entrada de novas empresas nos mercados.

Embora economia industrial trate de temas comuns à teoria microeconômica, ela apresenta especificidades quanto aos objetivos e à metodologia adotados. Apesar disso, não chega a romper completamente com a tradição neoclássica, continuando a utilizar uma série de conceitos da microeconomia.

De fato, os autores responsáveis pela concepção do, assim denominado, novo paradigma de análise, adotaram uma filiação teórica distinta dos princípios metodológicos marginalistas e constituíram uma abordagem alternativa à teoria neoclássica ao reconhecerem os dois fatos estilizados acima como relevantes no desempenho econômico. Dessa forma, os fenômenos da concentração dos mercados e de barreiras à mobilidade dos fatores tornaram-se os principais objetos de estudo deste novo paradigma.

Edward Mason, um dos fundadores do paradigma alternativo, em artigo publicado em 1939 na *American Economic Review*, foi explícito a respeito da metodologia: a abordagem utilizada pode ser denominada de institucional, no sentido, que investiga os fatores institucionais e históricos responsáveis pela concentração dos mercados e impedimento à mobilidade dos fatores produtivos. Seus fundamentos se aproximam do tipo das abordagens utilizadas por economistas como T. Veblen, J. R. Commons, J. M. Clark e W. Sombart. O interesse principal é aplicação das teorias no controle econômico do ambiente moderno. Este ambiente traz mudanças econômicas profundas que podem ser tanto a fonte de aumento de bem estar coletivo quanto da insuficiência ou insegurança econômica (Mason, 1939).

Antes de seu trabalho, deve-se, incontestavelmente, a Alfred Marshall os fundamentos da organização industrial. Em seus escritos Marshall apresenta uma análise das fases de desenvolvimento das indústrias e se esforça para integrar as abordagens das escolas históricas em um quadro de análise clássico e neoclássico. Convencido de que as análises utilitaristas contemporâneas, realizadas pela escola neoclássica, não eram capazes de explicar os fenômenos reais, o autor levou em conta o papel das estruturas produtivas no desenvolvimento, criando um conjunto de instrumentos para compreender as realidades industriais. Chamou atenção sob o peso das novas tecnologias, a organização do capital financeiro em escala mundial, a importância da estratégia das empresas sob situação de monopólio, desenvolveu os conceitos de economias internas e externas, aprofundou a análise dos rendimentos decrescentes, introduziu a noção de obstáculos à entrada, e colocou ênfase na importância da análise das indústrias como o lugar onde se organiza a concorrência e se estrutura as relações entre os agentes.

Portanto, contrárias ao objetivo principal de modelar e simplificar a observação praticada pela teoria neoclássica, as novas ideias ganham lugar na organização industrial adicionando novas variáveis nos modelos com a preocupação em fazer previsões e *explanar* o mundo real. Os economistas da organização industrial são mais inclinados a explicações ricas em detalhes institucionais e quantitativos, sem descuidar da teoria que demanda a construção de fatos estilizados abstraídos da realidade.

Concluindo, pode-se dizer que eles seguem à risca a definição de ciência econômica formulada por Joseph Schumpeter: qualquer domínio de conhecimento que desenvolveu técnicas especializadas de busca de fatos, interpretação ou análise pode ser considerado uma ciência. No domínio econômico, segundo o mesmo autor, o exercício da análise econômica depende da habilidade simultânea em três técnicas: história, estatística e teoria econômica. Assim os economistas industriais precisam dominar a teoria econômica para criar hipóteses sobre pressupostos fundamentais e suas consequências comportamentais. Eles precisam usar modernos métodos estatísticos para extrair generalizações apropriadas dos dados sobre estrutura e desempenho

industrial. É necessária também alguma familiaridade com os métodos e os resultados da pesquisa histórica tanto para perceber o fluxo amplo de eventos ao longo do tempo quanto para extrair do emaranhado de detalhes institucionais as causas iniciais de desvio da norma.

As motivações dos fundadores desse paradigma alternativo para o estudo da economia industrial justificam-se por dois aspectos principais. Inicialmente, os estudos nesta área exercem uma influência contínua na formulação e na execução de políticas públicas em diversas áreas, como escolha entre empresas públicas e privadas, regulação ou desregulação dos serviços de utilidade pública, promoção da competição por meio de regras antitruste e políticas de livre-comércio, estímulo à inovação por meio de patentes, financiamento a fundo perdido e subsídios. A segunda razão é contribuir para um paradigma em construção. A teoria, os dados e as metodologias estão gradualmente tornando-se disponíveis, mas ainda existe muito espaço para contribuições no avanço da fronteira do conhecimento.

Historicamente, há um pêndulo no interesse dos economistas pela disciplina economia industrial que, em princípio, bate ao contrário do funcionamento eficiente do mercado. Ou seja, à medida que a história mostra as deficiências do mercado natural, renova-se o interesse pela disciplina. Os economistas que acreditam ser necessário complementar o funcionamento do mercado com políticas atuaram fortemente no período conhecido como a Grande Depressão (1887-1915), quando as leis antitrustes e as primeiras agências regulatórias federais estavam se formando, e entre 1933 e 1940, quando a teoria econômica estava envolvida com a problemática de restabelecer o funcionamento do mercado após a crise de 1933.¹ Após a Segunda Guerra as atenções se voltaram para a questão de estabilização macroeconômica e o problema das nações em desenvolvimento, diminuindo o interesse pela área. Em 1970, o interesse foi fortemente reabilitado devido às mudanças econômicas profundas trazidas pelos novos paradigmas tecnológicos da informática e da biotecnologia, o esgotamento do uso de recursos naturais não renováveis como fontes de energia e as preocupações ambientais crescentes.

Entretanto, o interesse na disciplina ressurge neste período com novas orientações em relação à intervenção na economia. Quatro pressupostos mais condizentes com uma intervenção mais branda direcionam esse novo interesse: (1) ceticismo na efetividade da regulação governamental, (2) reconhecimento de que a organização dos mercados afeta os fluxos de comércio internacional, (3) dúvidas crescentes sobre a capacidade de adaptação e resposta das empresas industriais e (4) um debate teórico sobre as relações de determinação propugnadas pelo Modelo ECD e suas implicações para as políticas antitrustes.

Ao longo da primeira década do século XXI, várias das novas orientações começam a ser questionadas e, com a ocorrência de uma nova crise econômica global, em setembro de 2008, reascende-se o debate sobre a necessidade eminente do restabelecimento das políticas para o bom funcionamento do mercado, em particular, no mercado financeiro. No caso brasileiro, recomenda-se, como estratégia de crescimento e de desenvolvimento da economia, uma política de expansão da capacidade produtiva e de superação dos seus gargalos decorrentes de uma industrialização incipiente, característica dos países retardatários na construção de sua estrutura industrial. Entre os conjuntos de políticas propostos, estão aqueles voltados para a administração dos riscos provindos do setor externo e as políticas industriais (Lopez e Cardim, 2009).

A seguir, apresenta-se o Modelo ECD, paradigma que se contrapõe à teoria neoclássica para o estudo do comportamento das empresas e do mercado, vista no Capítulo 1, considerado também como paradigma fundador da disciplina de economia industrial.

4.3 O Modelo Estrutura-Conduta-Desempenho (ECD)

O Modelo ECD será apresentado aqui a partir da elaboração feita por Frederic M. Scherer, em 1970, em seu livro *Industrial Market Structure and Economic Performance*, em sua terceira edição, revista e atualizada em conjunto com David Ross, em 1990.

Como já mencionado na seção anterior, deve-se a Edward Mason os esforços mais importantes na direção de oferecer um método de análise geral das realidades econômicas industriais. Ele unificou as abordagens de observações históricas e de reflexões teóricas críticas contemporâneas (Joan Robinson, Edward Chamberlin, e Piero Sraffa, por exemplo) e o apresentou o Modelo ECD como o *quadro unificador*, capaz de permitir autonomia ao campo de economia industrial como uma disciplina independente. O autor elegeu quatro temas como foco de sua tarefa. Em primeiro lugar, os conceitos da microeconomia são muito abstratos, difíceis de estimar, e, em especial a concorrência pura e perfeita é concebida de maneira estática e sem considerar as decisões empresariais que prestam dinamismo ao sistema econômico.

Em segundo lugar, é necessário que a metodologia seja capaz de misturar fatos empíricos e teoria para reconhecer os fatores determinantes, rejeitando abordagens puramente normativas e preferir abordagens mais indutivas que levem em conta as realidades históricas e institucionais. Em terceiro lugar é preciso repensar o conceito de mercado ligado a produtos homogêneos e não diferenciados, adotando-se o conceito de indústria definido por A. Marshall – conjunto de tamanho variável de produtores

com produções heterogêneas – muito mais realista. Finalmente, deve-se adotar a hipótese que os comportamentos de uma empresa de uma dada atividade econômica são fortemente determinados pelas estruturas dominantes destas atividades.

Assim, pergunta-se inicialmente o que a sociedade deseja dos produtores de bens e serviços. A resposta, em geral, é um bom desempenho. O problema, então passa a ser como definir desempenho, já que ele pode ser entendido a partir de várias dimensões e, portanto, significar diferentes objetivos e metas como os ilustrados a seguir:

1. O que, quanto e como produzir de forma eficiente, ou seja, sem desperdiçar recursos escassos e levando em conta qualitativa e quantitativamente a demanda do mercado?
2. Como obter aumentos contínuos de produtividade por meio da incorporação de oportunidades científicas e tecnológicas em novos produtos, processos, matérias-primas, formas de organização?
3. Como favorecer o pleno emprego dos recursos, especialmente dos recursos humanos?
4. Como distribuir equitativamente a renda, impedindo o ganho excessivo de produtores ou apropriação indevida do excedente dos consumidores e/ou de determinados elos da cadeia produtiva?

A obtenção desses objetivos é envolta de conflitos políticos e a sua realização nunca poderá ser maximizada por cada produtor ou consumidor, mas deve-se buscar uma solução de maior satisfação possível de cada um por meio da monitoração permanente do desempenho de cada indústria.

Apesar das dificuldades de mensuração do desempenho industrial, dado que os indicadores refletem sempre apenas parte da realidade, existem alguns indicadores clássicos na literatura econômica que podem ser adotados: magnitude da margem preço-custo; taxas de mudança na relação entre produto/empregado; níveis de preço; distância entre custo atual e possível custo mínimo; variabilidade do emprego ao longo do ciclo de negócios; lançamento de novos produtos etc.

A ideia básica do Modelo ECD consiste, portanto, em identificar que variáveis ou conjunto de atributos são capazes de explicar as diferenças de desempenho observadas a partir do monitoramento das indústrias. As condutas das empresas são diferenciadas e motivadas, principalmente, pelo tipo de estrutura da indústria. A estrutura da indústria, por sua vez, depende de certo número de condições básicas que são de naturezas bastante diversas: técnicas, institucionais e relevância da demanda.

A Figura 4.1 ilustra esquematicamente o Modelo.

Pressupõe-se que o desempenho em indústrias ou mercados particulares seja dependente da conduta ou das estratégias dos vendedores (produtores) e dos compradores (consumidores) em diferentes assuntos, como a política e a prática de preços, cooperação explícita ou tácita entre as empresas, estratégias de linhas de produtos e propaganda, esforços de pesquisa e desenvolvimento (P&D), investimento em plantas produtivas, táticas legais (como, por exemplo, *enforcement* dos direitos de patentes), e assim por diante. A conduta, por sua vez, depende da estrutura do mercado relevante,² caracterizada pelo número e tamanho dos vendedores e compradores, o grau da diferenciação física ou subjetiva dos produtos e serviços, da presença ou ausência de barreiras à entrada de novas empresas, do formato das curvas de custo, do grau de integração vertical das empresas e da extensão de diversificação das empresas para outros mercados.

Finalmente, a estrutura de mercado é determinada por uma série de condições básicas. Por exemplo, do lado da oferta, os fatores determinantes da estrutura são: a localização e a propriedade da matéria prima, a natureza das tecnologias relevantes (processos contínuos ou em batelada, ou alta ou baixa elasticidade de substituição dos insumos), o grau de sindicalização da força de trabalho, a durabilidade do produto, o padrão de entrega da produção (pronta entrega ou encomenda), relação entre valor e peso do produto e o ambiente econômico informado pelas atitudes empresariais e pelo quadro legal. Do lado da demanda devem-se incluir entre outros fatores as condições de elasticidade-preço da demanda, a disponibilidade de produtos substitutos, a taxa de crescimento e flutuação da demanda ao longo do tempo, os métodos de compra utilizados pelos compradores (à vista ou a crédito). As condições básicas são fortemente influenciadas pelas leis vigentes e pelos valores socioeconômicos predominantes da comunidade de negócios.

Como sugerem as setas na Figura 4.1, o fluxo causal ou a relação de determinação inicia-se nas condições básicas de demanda e de oferta, seguindo para a estrutura do mercado, para as condutas ou estratégias empresariais e para o desempenho. Nesta sequência, poderia dizer-se que as condições estruturais representadas pelas condições básicas e estrutura do mercado são exógenas, ou não determinadas pelo modelo, o que lhe empresta um caráter estático.

Porém, podem-se observar também importantes efeitos de retroalimentação ou de encadeamento representados pelas setas pontilhadas na Figura 4.1, prevendo-se efeitos sistêmicos de influência entre as variáveis. Por exemplo, um esforço intenso em P&D (conduta) pode alterar o paradigma tecnológico dominante na indústria (condição básica de oferta) e, portanto suas condições de estrutura de custo e de diferenciação de produto (atributos da estrutura). Outro exemplo seria a política de preços (conduta) praticada por algumas empresas já estabelecidas na indústria, estimulando a entrada ou a saída de, respectivamente, novas empresas ou a saída de empresas estabelecidas. Dessa forma, teria havido uma influência da conduta para a estrutura

representada pela mudança no número de vendedores/produtores. A partir desses e outros exemplos pode-se pressupor que tanto as condições básicas de oferta e de demanda quanto a estrutura do mercado são determinadas endogenamente por meio do conjunto de relações do sistema e não fixadas por forças exógenas, como se discutirá na próxima seção.

4.3.1 O PAPEL DAS POLÍTICAS E DA REGULAÇÃO PÚBLICAS

Se o mercado natural falha em organizar a indústria, conduzindo-a a um desempenho não eficiente, então, o governo precisa intervir. Sua intervenção pode se dirigir às variáveis da estrutura e/ou da conduta, conforme indicam as setas tracejadas da Figura 4.1.

O arsenal de instrumentos de política e de regulação é bastante amplo e na Parte VI deste livro eles serão discutidos em cinco capítulos, aqui citamos algumas possibilidades. Custos, investimentos e preços podem ser influenciados por taxas ou subsídios. A estrutura de mercado pode ser influenciada pelo estabelecimento de tarifas de comércio, quotas de importação e outras políticas voltadas para o comércio internacional ou atração de investimentos externos diretos. Outra forma de intervenção mais branda seria melhorar o provimento de informações tanto para consumidores quanto para produtores, reduzindo as

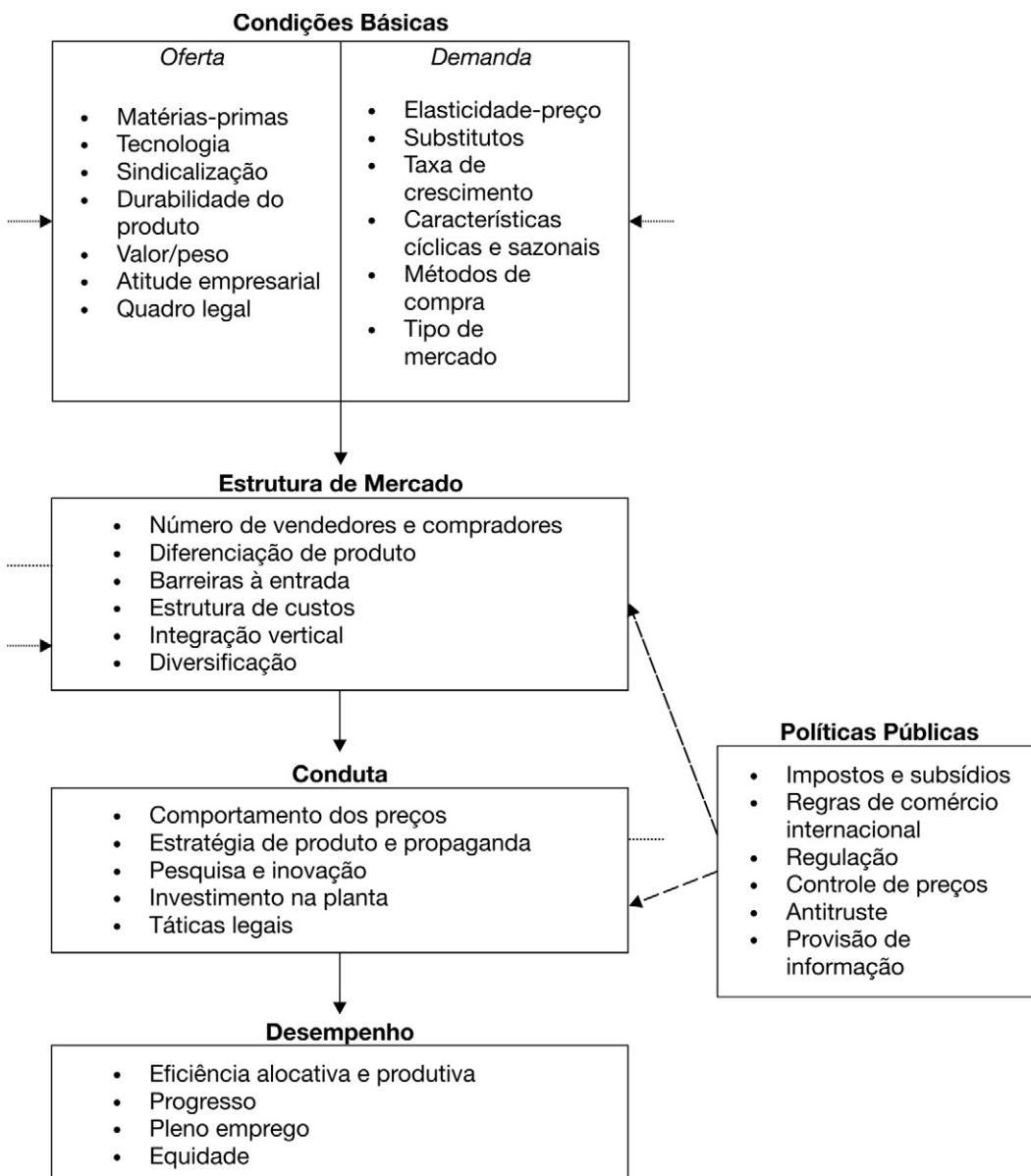


Figura 4.1 O Paradigma Estrutura-Conduta-Desempenho (Modelo ECD) – ver Scherer, p. 5

assimetrias de informação. A regulação dos mercados é outra forma possível de intervenção determinando condições prévias de operação e circulação de bens e serviços ou controlando preços de entrada. As políticas antitrustes ou as políticas que incentivam a concorrência, bem como as políticas ambientais são estratégias mais recentes e cada vez mais valorizadas como guias de bom desempenho industrial e formas indiretas ou passivas de intervir no mercado. Em casos extremos o governo também pode decidir por prover, ele mesmo, por meio da produção pública, bens e serviços. Alternativa é criar empresas públicas com funções reguladoras de custos, preços e padrões de qualidade capazes de emular uma concorrência entre os rivais privados que elevem o desempenho geral da indústria.

Como já desenvolvido anteriormente, a intensidade da intervenção praticada pelos governos varia em sentido inverso ao grau de bom funcionamento do mercado. Assim, em períodos nos quais o desempenho industrial se mostra pior, crescem as políticas intervencionistas. Há também diferenças importantes entre os países líderes na crença sobre a capacidade do mercado natural organizar o desempenho industrial de forma eficiente. Os Estados Unidos, por exemplo, sempre foram crédulos, ao passo que a Europa e o Japão têm sido bem mais céticos.

A regulação pode ter um caráter genérico, quando é voltada para a economia como um todo, ou específico para algumas indústrias apenas. Um exemplo do primeiro tipo é o controle de preços generalizado para a economia.

Entretanto, a justificativa econômica clássica, utilizada desde os escritos de Adam Smith para o Estado regular a economia, era garantir que as indústrias denominadas de bens e serviços públicos atendessem a todos os cidadãos de forma equitativa. Em geral, os setores que atuam nessa área (eletricidade, distribuição de gás, serviços telefônicos, estradas de ferro, dutos para transportar petróleo e gás natural) atuam por intermédio de grandes empresas, devido aos elevados custos fixos envolvidos nestas atividades e à necessidade de diluí-los por meio de grandes escala de produção. Portanto, uma única empresa ocupa o mercado todo, formando o que se denomina o monopólio natural. Daí a necessidade de regulação de custos, preços e qualidade dos bens e serviços.

Na prática, existem poucas empresas no que se denomina indústria de utilidade pública. Entretanto, várias outras indústrias são reguladas por outras razões. Uma justificativa clássica para a intervenção é a presença de assimetrias de informação ou de externalidades. O primeiro caso é característico de indústrias com tecnologias e padrões de produção complexos bem como produtos com características tecnológicas fora do padrão do senso comum. Nesse caso, justifica-se a regulação porque nem sempre a indústria costuma fornecer as informações suficientes para o consumo desses produtos, havendo necessidade de se arbitrar um padrão, por meio do estabelecimento de normas mínimas de produção e características do produto. O segundo caso é ilustrado pela poluição, uma conduta que afeta a todos e não somente os produtores que a geraram. Ou seja, a decisão de produção de uma empresa gera custos para a sociedade, sem que ela tenha decidido produzir mais poluição. Há que se arbitrar por meio de taxas, quotas de emissão e outros instrumentos, multas às empresas para desestimular as condutas agressivas ao meio ambiente.

A regulação também pode se justificar para impedir que lucros extraordinários sejam apropriados por determinados grupos que não estão contribuindo para isso. Exemplos são mudanças conjunturais dos preços e abuso do poder de monopólio obtido por concessões do Estado (patentes e exploração exclusiva de recursos naturais).

Outra hipótese considerada para a regulação é a existência de interesses organizados para a obtenção de benefícios. Nesses casos o regulador deve estar muito atento para não ser “capturado” pelo interesse dos regulados. Esses, em geral, têm mais conhecimento sobre os objetos de regulação do que os próprios reguladores e utilizam influências políticas para defenderem seus interesses de forma sistemática.

Enfim, a questão da regulação é extremamente complexa. Encontrar o ponto certo entre regular ou desregular é um desafio permanente. Entretanto, não regular pode significar deixar que o funcionamento do mercado natural, muitas vezes inadequado, distorça os resultados desejados pela sociedade do desempenho industrial.

A legislação antitruste, também conhecida como “defesa da concorrência”, é uma forma específica de regulação. As principais diferenças em relação a outros tipos de regulação são o caráter episódico da regulação antitruste e o escopo mais bem definido das intervenções. Entre as principais intervenções estão: proibição de contratos, combinações e conspirações para restringirem o comércio; proibição de monopolização; proibição de discriminação de preços; e, em nível multinacional, o Tratado de Roma, de 1957, contém uma política de competição explícita. Um maior detalhamento das particularidades das políticas e da regulação públicas, mas também das práticas brasileiras a respeito, será discutido na Parte VI desse livro.

4.4 Evidências Empíricas e Limitações do Modelo ECD

O verdadeiro entendimento da racionalidade de intervenção regulatória ou de políticas antitruste depende da compreensão do funcionamento do Modelo ECD, ou seja, das relações entre desempenho do mercado, estrutura e conduta das empresas. A compreensão dessas relações, por sua vez, é ilustrada por casos reais de estudos industriais. Muitas propostas e tipos de

verificações empíricas foram conduzidos para enriquecer ou verificar a veracidade das determinações causais propostas pelo quadro teórico que se acabou de apresentar. Elas podem ser resumidas por dois tipos de estudos: estudos de caso e estudos econôméticos.

Os estudos de caso realizados durante os anos 1950 foram devotados às atividades de base das indústrias de aço, petróleo e automóveis. Levavam-se em conta todos os aspectos qualitativos capazes de elucidar melhor a realidade industrial desses setores. Ainda que as informações quantitativas dos principais indicadores de concentração, de rentabilidade e outros tenham fornecido informações fecundas sobre o funcionamento do mercado, elas não foram suficientes para permitirem generalizações sobre o seu funcionamento. Um exemplo desse tipo de estudo realizado para analisar modificações ocorridas na indústria brasileira de massas alimentícias encontra-se no **Quadro 4.1**.

Os estudos econôméticos, conduzidos, sobretudo nos anos 1960 e 70, dedicaram-se a encontrar ligações específicas significativas entre certas estruturas e diversas medidas de desempenho. As análises de regressão foram realizadas a partir de amostras de atividade e verificações de hipóteses simples tais como qual é a influência do grau de concentração das atividades sobre suas margens de lucro (Bain, 1941)? Sobre sua produtividade (Stigler, 1968)? Do tamanho das empresas sobre o grau de inovação (Scherer, 1970)? Ou testando relações mais complexas tais como que papéis específicos exercem as variáveis características das estruturas das atividades sobre suas taxas de lucro (Bain, 1956)? Ou ainda qual é o impacto das despesas de publicidade sobre a lucratividade (Comanor e Wilson, 1979)? Apesar dos avanços metodológicos dos estudos econôméticos, que possibilitavam a generalização das conclusões, sua análise baseava-se em regressões lineares simples entre duas variáveis, como as taxas de lucros e a estrutura de mercado (medidos em número de produtores ou índices de concentração). A análise das condutas das empresas em diferentes estruturas de mercado era, portanto, perdida.

Várias observações devem contribuir para marcar os limites destas abordagens metodológicas:

1. Os argumentos teóricos utilizados para incluírem este ou aquele aspecto estrutural da lista de variáveis explicativas eram frequentemente bem pobres e escolhidos sem referência a um modelo subjacente;
2. Às vezes os métodos se perdiam nos problemas infundáveis da causalidade e da simultaneidade do conjunto de variáveis da estrutura, conduta e desempenho, sem conseguir saber o que determina o quê;
3. Enfim, a maior parte das relações estaria caracterizada pelas condições históricas nas quais elas foram estabelecidas e dificilmente sobreviveriam a mudanças na economia.

QUADRO 4.1 A ANÁLISE E-C-D NAS INDÚSTRIAS DE MASSAS ALIMENTÍCIAS NO BRASIL

O estudo de Ferreira Jr. e Gomes (2006) aplicou a análise E-C-D para avaliar as modificações nas indústrias de massas alimentícias após a abertura comercial no Brasil, a partir de 1990. O trabalho partiu do pressuposto de que a abertura comercial modificou as condições básicas da oferta, ao expor as empresas nacionais à concorrência internacional. Os autores buscaram explorar, de maneira objetiva, dois componentes do modelo: a estrutura, medida pelo grau de concentração da produção, e o desempenho, medido por uma série de indicadores, como o nível de produção, exportação, importação, faturamento, lucratividade e custos de produção. A **Tabela 4.1**, a seguir, resume os principais resultados encontrados para a indústria de massas alimentícias.

TABELA 4.1 Indicadores de estrutura e desempenho na indústria de massas alimentícias: 1995 e 2001.

Ano	CR ⁸	Produção (mil ton.)	Exportação (mil ton.)	Importação (mil ton.)	Faturamento (R\$ milhões)	L/F ²	C/F ²
1995	32,64	850	0,86	21,71	2.140,67	4,95	72,15
2001	48,52	935	1,78	14,52	1.899,15	0,59	72,81

Fonte: Ferreira Jr. e Gomes (2006).

Notas: ¹Participação nas vendas das 8 maiores empresas; ²Lucro sobre faturamento; ³Custos de produção sobre faturamento.

Como se pode observar, houve uma tendência de concentração da produção entre as oito maiores companhias. Por outro lado, a produção e a exportação aumentaram, enquanto as importações reduziram-se. Analisando o faturamento, em preços constantes de 2001, verifica-se uma queda no seu valor e, consequentemente, uma redução da lucratividade. Segundo os autores, boa parte da queda do faturamento se deu pela redução do preço das massas alimentícias ao longo do período analisado. Isso sugere que, apesar da concentração na estrutura, a abertura comercial promoveu um aumento da competição entre as empresas, o que pode ser observado na melhora do indicador de desempenho lucro sobre faturamento. A relação dos custos de produção e faturamento, por outro lado, apresentou um acréscimo, em parte porque o faturamento caiu, e em outro, porque o aumento da competição exigiu melhoria na qualidade dos produtos fabricados, o que pode ter implicado em aumento de custos na produção.

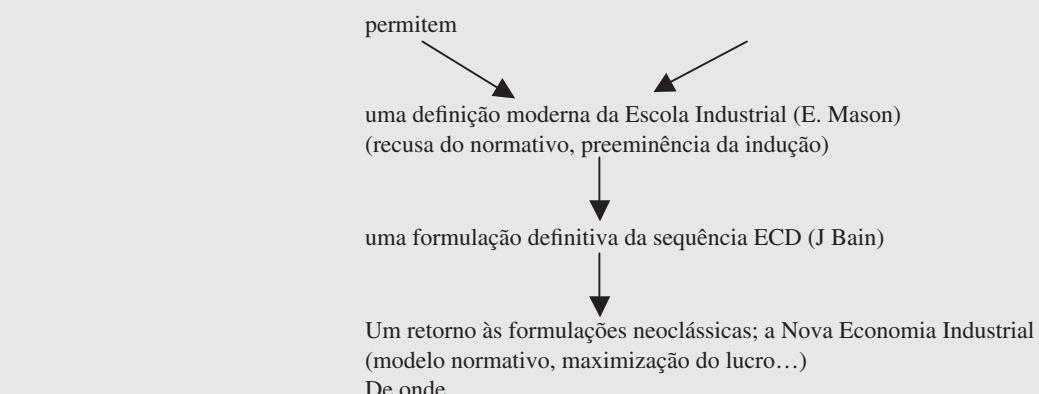
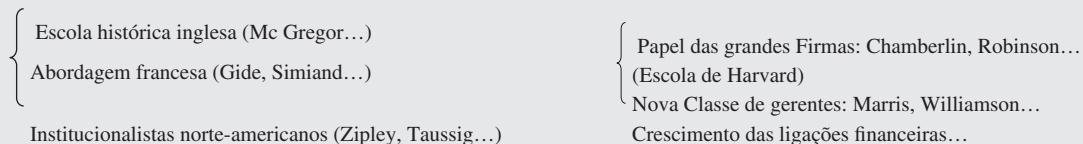
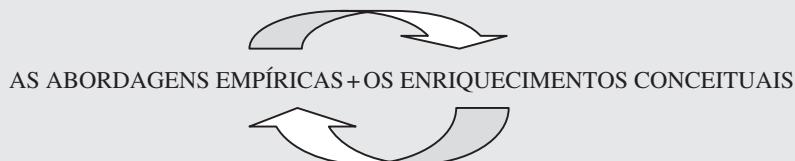
4.5 Desdobramentos e Contribuições

As principais limitações da disciplina economia industrial nascem do seu não rompimento inicial com a teoria neoclássica tradicional. A nova disciplina pretendia apenas confrontar as previsões da economia pura com as evidências empíricas de funcionamento corrente dos mercados. Desde Alfred Marshall já se podia prever os seus desdobramentos modernos. De um lado, os fenômenos produtivos passaram a ser abordados a partir de uma ótica indutiva, dando um grande espaço às realidades institucionais. De outro lado, eles passaram a ser analisados a partir de uma ótica mais hipotética dedutiva, inicialmente apoiada em uma perspectiva da microeconomia evolutiva (inspirada no paradigma da biologia) e, posteriormente, cada vez mais, racionalizada. A segunda corrente hoje é dominante. Para compreensão sucinta dessa evolução e desdobramentos consultar o Quadro 4.2.

Uma discussão introduzida nos anos 1970 sobre as relações de causalidade do modelo e já adiantada na introdução deste livro pode resumir as contribuições e os possíveis desdobramentos do Modelo ECD. Imaginar o funcionamento

QUADRO 4.2 RESUMO SUCINTO DE EVOLUÇÕES DAS ABORDAGENS DE ECONOMIA INDUSTRIAL

- A partir das abordagens iniciais de J.B. Say, de List, de W. Ashley, de T. Veblen.
- Filho dos clássicos e pai dos neoclássicos, A. Marshall estabelece as bases da economia industrial (noção de “*indústria*” em vez de mercado, importância das “*economias de escala*”, das *elasticidades*,...)
- Em seguida,



1. Críticas
 - Os “desvios” da abordagem inicial: os pós-marshallianos
(S. Labini, Clark, P. Andrews, E. Brenner...)
 - Os fundamentos da abordagem (A Escola Austríaca);
2. Emergência de uma Nova Economia Industrial: os neo-neoclássicos
 - Teoria do oligopólio e teoria dos jogos,
 - Teoria dos “mercados contestáveis”,
 - Teoria dos “custos de transação”;
3. Desenvolvimento da mesoeconomia: além do mercado único, levando em conta as estruturas socioeconômicas; as interações entre estruturas e comportamentos; o jogo coletivo (grupos, cadeias produtivas, complexos industriais...)

Fonte: Adaptado de Morvan (1991), p. 33.

do modelo supondo uma determinação endógena das condições básicas da demanda e da oferta permite que a análise paradigmática alternativa possa englobar tanto a tradição estática quanto a tradição dinâmica de concorrência da análise de economia industrial.

De fato, pode-se perceber claramente que o conceito de concorrência na análise econômica pode ter dois significados diferentes. Um primeiro que enfatiza a conduta dos vendedores/produtores e outro que enfatiza a estrutura do mercado. Além disso, enfatiza-se que esse conceito evoluiu ao longo do tempo.

Adam Smith considerava a essência da competição a rivalidade estabelecida de forma independente por cada vendedor/produtor para obter clientela para si em detrimento dos demais. Quanto mais vendedores, mais competitivo seria o mercado. A regulação do mercado se daria de forma automática (“mão invisível”) por meio da atração de mais vendedores para o mercado que permitisse a realização de maiores taxas de lucro, as quais tenderiam a se reduzir pela divisão dos ganhos com um número cada vez maior de produtores. Dessa forma, os preços cairiam no nível dos custos (esforço necessário para produzir) garantindo uma remuneração justa e equilibrada no longo prazo. É importante ressaltar que este movimento de redução das taxas de lucro no nível dos custos no longo prazo, não é incompatível com lucros extraordinários no curto prazo. Além disso, para que o mecanismo de alocação de investimentos via o mercado funcionasse era preciso não haver barreiras à mobilidade perfeita dos fatores de produção.

A teoria econômica neoclássica, ao introduzir o raciocínio da disciplina de cálculo para formalizar a teoria econômica, redefiniu o conceito de competição imaginado por Adam Smith. Um mercado é considerado perfeitamente competitivo quando o número de empresas vendedoras, vendendo um produto homogêneo, é tão grande e a parcela individual de cada uma delas tão pequena que nenhuma empresa é capaz de influenciar significativamente o preço dos produtos por meio da variação da quantidade produzida por ela. Em outras palavras o preço é um dado para a empresa. Ele é determinado pelas forças de oferta e demanda globais no mercado e foge ao controle delas. Esta definição de competição, diferentemente da adotada por Adam Smith, não segue o senso comum: uma batalha entre preços e outras estratégias para conquistar parcelas já ocupadas do mercado. Se o mercado é amplo o suficiente para não ser influenciado pelas decisões individuais dos produtores, então, pode haver competição sem rivalidade entre as empresas. Dessa forma, os produtores se restringem a decidirem sobre a quantidade a ser produzida.

Assim, considerando uma visão de concorrência estática, que toma o mercado de competição perfeita, tal como hipotetizado pela teoria neoclássica, a análise das condutas ou das estratégias das empresas é supérflua, dado que as informações fornecidas pela estrutura são suficientes para prever ou induzir o comportamento adequado das empresas. Logo, em um mercado de competição perfeita, as empresas não têm condições individuais de influenciar a concorrência e se comportarão de acordo com os sinais emitidos pelo mercado (preços), ajustando seus custos para não ter prejuízo. Em um mercado do tipo oligopolizado, elas se comportarão de forma colusiva para determinar seus preços de modo a extraír o máximo de excedente. Esta é considerada uma visão estática da concorrência porque as empresas apenas reagem às estruturas dadas dos mercados, não rivalizando diretamente com as demais empresas.

Esta é também a principal contribuição de Joe Bain que retomara a abordagem de E. Mason, no início dos anos 1950, formalizando-a a partir de observações estatísticas sobre as relações causais entre estrutura, conduta e desempenho. O autor demonstrou que a taxa de lucro dos setores (indicador de desempenho) é estatisticamente correlacionada com o grau de concentração e com o nível de barreiras à entrada, afirmando existir uma relação indireta entre os desempenhos e as estruturas de mercado. O autor marca de certa forma uma ruptura com a abordagem de E. Mason e um retorno à tradição neoclássica conforme ilustra o Quadro 4.3, ou seja, os desempenhos podem ser diretamente deduzidos das características das estruturas.

A grande contribuição de Bain foi estudar as condições de entrada de concorrentes. Assim, a concorrência que um produtor enfrenta em um mercado não se resume apenas àquelas empresas já estabelecidas, mas também deve considerar os potenciais entrantes. As condições de entradas poderiam ser facilitadas, dificultadas (intermediárias) ou bloqueadas, de acordo com as barreiras impostas pelas empresas já estabelecidas no mercado. Bain estabeleceu três tipos de barreiras à entrada: diferenciação de produto, vantagens absolutas de custos e economias de escala. A presença de um ou mais desses elementos poderia dificultar a entrada dos competidores potenciais. Assim, sua teoria fornecia uma explicação alternativa para a inexistência, na vida real, de mercados perfeitamente competitivos.

Em outra visão dinâmica da concorrência, ilustrada pela Figura 4.2, mais condizente com a visão seminal de E. Mason pressupõe-se que as empresas busquem permanentemente alterar as condições básicas de oferta e demanda e de estrutura do mercado, por meio de suas estratégias, a fim de obterem vantagens competitivas em seu favor. Os exemplos das estratégias de investimentos em P&D e de preço, apresentadas anteriormente, ilustram bem como as empresas são capazes de alterar, respectivamente, a condição básica de oferta e a estrutura da indústria. A evidência empírica, ao contrário do que supõe a visão estática de concorrência, tem mostrado condutas ou estratégias diferenciadas em ambos os tipos de estruturas. Tanto as pequenas empresas, atuando em mercados competitivos, podem introduzir inovações que provocam disrupturas em suas condições básicas de oferta, quanto grandes empresas, atuando em estruturas oligopolizadas, podem não se comportar colusivamente, abrindo guerras de preços desenfreadas

QUADRO 4.3 CONCORRÊNCIA PERFEITA (MERCADO 1) E OLIGOPÓLIO (MERCADO 2) NA VISÃO ESTÁTICA E TRADICIONAL DO MODELO ECD

COMPARAÇÃO DE DOIS CASOS POSSÍVEIS

	MERCADO 1 (concorrência perfeita)	MERCADO 2 (oligopólio)
Características estruturais	<ul style="list-style-type: none"> • Muitos vendedores e compradores • Produto homogêneo • Ausência de barreiras à entrada 	<ul style="list-style-type: none"> • Poucas firmas com concentração possível de compradores • Produto bastante diferenciado • Existência de barreiras à entrada (e à saída)
Desempenho	<ul style="list-style-type: none"> • Ausência de lucros extraordinários • Alocação ótima dos fatores • Ausência de inovação 	<ul style="list-style-type: none"> • Existência de lucros extraordinários • Alocação não ótima dos fatores • Inovação provável

Fonte: Adaptado de Morvan (1991), p. 33. Fonte: Adaptado de Morvan (1991), p. 10.

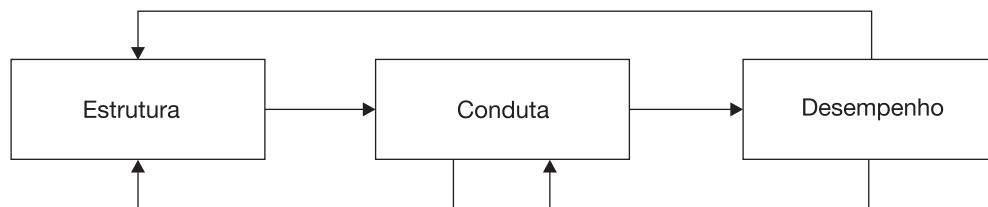


Figura 4.2 A abordagem dinâmica e alternativa do Modelo ECD

ou adotando atitudes inovadoras relevantes para a acumulação do conhecimento. Essas evidências mostram que as estratégias empresariais devem ser objeto de estudo e não apenas serem deduzidas das estruturas existentes, como pressupõe a visão tradicional de concorrência. As estratégias das empresas são, portanto, responsáveis pelo dinamismo dos mercados.

Alguns autores chegam a falar do conceito de barreiras à entrada estáticas e dinâmicas. As primeiras são decorrentes de condições estruturais das indústrias e dos mercados e as segundas resultam de estratégias deliberadas das empresas estabelecidas para deterem a entrada das empresas potenciais entrantes. Neste sentido, as estratégias empresariais são utilizadas como mecanismos de prevenção à entrada.

Outros chamam a atenção para as barreiras à saída. Qualquer custo que uma empresa tenha que incorrer para deixar uma indústria é uma barreira à saída. Um exemplo bastante comum são as multas contratuais.

Em geral, as barreiras à saída estão associadas aos custos irrecuperáveis. Estes custos estão associados quase sempre a fatores fixos tangíveis e intangíveis e específicos à empresa ou ao produto. Exemplos incluem patentes combinadas com investimentos físicos e em especialização dos recursos humanos e investimentos em propaganda e marketing.

Em mercados nos quais os custos de capital não puderem ser facilmente repassados a terceiros devido a sua especificidade, as empresas estabelecidas estarão fadadas a permanecerem em seus mercados ou perderem todo ou quase todo o investimento inicial realizado. Um ótimo exemplo é a indústria ferroviária, na qual os trilhos utilizados não são vendáveis, fazendo com que as empresas ferroviárias permaneçam no mercado mesmo com prejuízo.

Mais recentemente, o conceito de competição foi ampliado com a noção de “mercado contestável”. Em mercados nos quais a entrada e a saída de empresas não são dificultadas, ainda que as condições de homogeneidade dos produtos e número de produtores não sejam compatíveis com as condições de concorrência perfeita, há potencial para a competição. Dessa forma, nasceu o conceito de “competição potencial”. Enfim, várias dessas contribuições e também os seus limites serão abordados no desenvolvimento deste livro, conforme já apresentado na Introdução.

Notas

1. O *Economic of National Resources Committee* é um órgão americano de planejamento, criado em 1933, responsável pela eficiência de todo o sistema econômico com o fim de garantir os maiores ganhos possíveis do ponto de vista social, em oposição aos pontos de vista dos empresários e da indústria. Sua ênfase de atuação era na coordenação das atividades econômicas, deixando a sua normatização específica a cargo de cada atividade econômica (Gruchy, 1939).
2. O conceito de mercado relevante (ver Capítulo 22) identifica a amplitude ou escopo de produtores e compradores que concorrem para a produção de produtos substitutos e os consomem indiferentemente e, portanto, devem ser incluídos entre os atores relevantes na dinâmica competitiva dos mercados.

Bibliografia

- Bain, J. (1941) The Profit Rate as a Measure of Monopoly Power. *The Quarterly Journal of Economics*, v. 55, n. 2, p. 271-293.
- Bain, J. (1942) Market Classifications in Modern Price Theory. *The Quarterly Journal of Economics*, v. 56, n. 4, p. 560-574.
- Bain, J. (1956) *Barriers to New Competition: their character and consequences in manufacturing industries*. Cambridge: Harvard University Press.
- Bain, J. (1959) *Industrial Organization*. New York: John Wiley& Sons, Inc.
- Comanor, W. S.; Wilson, T. A. (1979) The Effect on Competition: a survey. *Journal of Economic Literature*, v. 17, n. 2, p. 453-476.
- Ferguson, P. R. (1988) *Industrial Economics: issues and perspectives*. London: Macmillan Education Ltda.
- Ferreira Júnior, S.; Gomes, M. F. M. (2006) Ajustamentos nas agroindústrias de biscoitos e massas alimentícias no Brasil, 1995 a 2001. *Revista de Economia e Sociologia Rural*, Rio de Janeiro, v. 44, n. 1, p. 79-98, jan./mar.
- Gruchy, A. G. (1939) The Economic of National Resources Committee. *The American Economic Review*, v. 29, n. 1, p. 60-73.
- Lopez, J.; Cardim, F. (2009) Uma Estratégia de Desenvolvimento na Tradição Keynesiana-Estruturalista. In: Sicsú, J.; Castelar, A. (orgs). *Sociedade e Economia: Estratégias de Crescimento e Desenvolvimento*. Brasília: Ipea, cap. 6, p. 51-58.
- Mason, E. S. (1939) Price and production policies of large-scale enterprise. *American Economic Review*, v. 29, p. 61-74.
- Mason, E. S. (1949) The current state of the monopoly problem in the United States. *Harvard Law Review*, v. 62, p. 1265-1285.
- Morvan, Y. (1991) *Fondements d'Economie Industrielle*. Paris: Econômica.
- Scherer, F. M. (1970) *Industrial Market Structure and Economic Performance*. Chicago: Rand mc Nally College Publishing.
- Scherer, F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*. Boston: Houghton Mifflin Company.
- Stigler, G. (1968) *The Organization of Industry*. Chicago: The University of Chicago Press.



PARTE



ANÁLISE ESTRUTURAL DOS MERCADOS



Concentração Industrial

Marcelo Resende e Hugo Boff

5.1 Introdução

A noção de estrutura de mercado desempenha um papel fundamental dentro do influente paradigma da estrutura-conduta-desempenho. Dessa forma, a quantificação do componente estrutural, em termos de medidas sintéticas ainda encontra ampla utilização em Economia Industrial.

As aplicações empíricas frequentemente se amparam em medidas de concentração deficientes, o que serve de motivação para uma apresentação comparativa das principais medidas empregadas em estudos na área. As medidas de concentração pretendem captar de que forma agentes econômicos apresentam um *comportamento dominante* em determinado mercado e, dessa forma, os diferentes indicadores consideram as participações no mercado dos agentes (por exemplo, a participação de cada empresa no total das vendas do setor), segundo diferentes critérios de ponderação.

Medidas de concentração industrial são úteis para indicar preliminarmente os setores para os quais se espera que o *poder de mercado* seja significativo. Contudo, existem pelo menos três razões para que esses indicadores, construídos a partir de participações de mercado, não sejam completos nesse tocante:¹

1. Se a entrada em um mercado for fácil, nenhuma empresa poderá exercer poder de mercado, não importando o quanto ampla seja sua participação nesse mercado;
2. Uma empresa pode ter uma parcela de mercado elevada não decorrente de poder de mercado, mas advinda de custos reduzidos ou de produtos de qualidade superior;
3. O cálculo de medidas de concentração pressupõe a delimitação de mercado e implica ignorar a disciplina exercida por substitutos próximos, comercializados em outros mercados.

Este capítulo pretende fornecer uma introdução às principais medidas de concentração usualmente utilizadas, indicando os méritos relativos desses diferentes indicadores. O capítulo está organizado da seguinte forma: a segunda seção apresenta três medidas de concentração e discute suas propriedades. A terceira seção apresenta uma ilustração empírica das medidas de concentração para setores industriais selecionados na economia brasileira recente. Finalmente, a quarta seção compara os méritos relativos das diferentes medidas segundo critérios levantados na literatura.

5.2 Medidas de Concentração

Índices de concentração pretendem fornecer um indicador sintético da *concorrência* existente em um determinado mercado. Quanto maior o valor do índice de concentração, menor é o grau de concorrência entre as empresas e mais concentrado (em uma

ou poucas empresas) estará o poder de mercado virtual da indústria. O *padrão concorrencial* vigente, contudo, é o resultado da ação dos produtores individuais (conduta), ao escolherem os níveis de preço ou as quantidades ofertadas (variáveis estratégicas), dadas as características específicas dos produtos fabricados (substituição ou diferenciação existente entre eles, níveis de qualidade etc.), as preferências dos consumidores e as condições de acesso (existência ou não de barreiras de mercado à entrada de novas empresas). As taxas de *preferências intertemporais* dos agentes, seus graus de *informação* e seus coeficientes de *aversão ao risco* (incerteza) são fatores que influenciam as tomadas de decisão. Enfim, o padrão concorrencial pode, eventualmente, ser balizado também por fatores de ordem *institucional*, tais como os arcabouços jurídico e tarifário (fiscal) que limitam ou disciplinam a atuação das empresas ou do mercado.

O padrão concorrencial contribui para dar uma *estrutura* particular à indústria, como consequência do *desempenho* das empresas e dos *resultados* obtidos, vale dizer: dados os recursos empregados, como consequência da maior ou menor eficiência produtiva alcançada (menores ou maiores custos) e da maior ou menor eficiência gerencial obtida (maiores ou menores lucros). Os resultados obtidos pelas empresas lhes conferem, pelo seu lado, um determinado “poder de mercado” individual no seio da indústria, que o índice de concentração tentará justamente capturar, de uma forma sintética.

O *poder de mercado* virtual de uma empresa individual está relacionado com sua capacidade de *controlar* o preço de venda do produto. Por isso, empresas mais eficientes, que produzem com custos de produção mais baixos, têm mais facilidade que as demais de competir em preço e de ocupar parcelas crescentes do mercado por meio de reduções progressivas no preço. Mais particularmente, o poder de mercado de uma empresa se manifesta pela sua capacidade de fixar e sustentar o preço de venda em um nível *acima* daquele fixado pelas concorrentes, sem prejuízo para sua participação no mercado.

O poder de mercado assume forma *aparente* na participação no mercado (*market share*) da empresa, a razão entre sua oferta (ou suas vendas) e a oferta total da indústria (ou vendas totais). Outras medidas de tamanho são também ocasionalmente usadas, tais como o patrimônio líquido, a capacidade produtiva instalada ou o número de empregados, muito embora as parcelas obtidas nesses casos não reflitam necessariamente o poder de mercado exercido pela empresa sobre o seu produto. Pelo seu lado, a distribuição resultante das parcelas de mercado entre as empresas cristaliza uma forma aparente da *estrutura do mercado*, de maneira que o índice de concentração deverá levar em conta não apenas o *nível* das parcelas de mercado individuais, como também a sua *distribuição* (que poderá ser mais ou menos desigual).

5.2.1 CONCENTRAÇÃO E DESIGUALDADE

Uma maior *concentração* industrial implica maior *desigualdade* na repartição do mercado entre as empresas, isto não significa que o inverso seja verdadeiro, isto é, que maior desigualdade implica maior concentração. Apesar de próximos, os dois conceitos não são equivalentes. Por exemplo, uma indústria composta de duas empresas que dividem o mercado em partes iguais possui graus de desigualdade e de concentração mínimos. Entretanto, a entrada de uma terceira empresa para atender 1% do mercado em detrimento das empresas estabelecidas (que preservam 49,5% cada uma) aumentará consideravelmente o grau de desigualdade, mas não o grau de concentração, já que o poder de mercado das empresas instaladas não será significativamente afetado com a presença da empresa entrante.

5.2.2 MEDIDAS POSITIVAS E NORMATIVAS

Do ponto de vista classificatório, podemos definir medidas de concentração como *parciais* ou *sumárias*, *positivas* ou *normativas*. Medidas de concentração *parciais* não utilizam os dados da totalidade das empresas em operação na indústria em consideração, mas apenas de uma parte delas. As chamadas *razões de concentração* (*concentration ratios*) constituem o principal exemplo dessa categoria. As medidas sumárias, por outro lado, requerem dados sobre todas as empresas em operação. Exemplos influentes são os índices de concentração de *Hirschman-Herfindahl*² e *entropia* a serem apresentados adiante.

Medidas de concentração *positivas* são unicamente função da estrutura aparente do mercado industrial (o nível e a distribuição das parcelas de mercado) e não dependem de qualquer parâmetro comportamental, seja ele relativo aos produtores (variações conjecturais, coeficientes de aversão à incerteza etc.), seja ele relativo aos consumidores (elasticidades diretas ou coeficientes de substituição). Medidas positivas resumem melhor os aspectos estatísticos presentes no fenômeno da concentração, mas nem sempre são adequadas para uma avaliação econômica (necessariamente normativa) do desempenho industrial.

Medidas de concentração *normativas* levam em conta, além da estrutura aparente, os parâmetros comportamentais mencionados anteriormente, que estão relacionados com as preferências dos produtores ou dos consumidores. Essas medidas visam avaliar normativamente o desempenho industrial, seja do ponto de vista dos primeiros (maiores ou menores lucros), seja do ponto de vista dos segundos (maior ou menor excedente do consumidor). Medidas normativas que visam a uma avaliação

social (global) também podem ser construídas, levando em conta os interesses de todos os agentes (maior ou menor excedente total). No primeiro e no último casos, as medidas de concentração também dependerão dos parâmetros relacionados com os custos incorridos pelas empresas (tecnologia).

Neste capítulo, trataremos unicamente das medidas *positivas* de maior uso na literatura econômica. Vale lembrar que os indicadores positivos de concentração englobam dois componentes fundamentais: o número de empresas e a desigualdade no tamanho destas. Conforme veremos a seguir, um ou outro desses componentes nem sempre são adequadamente contemplados pelas medidas utilizadas com frequência em análises empíricas da estrutura industrial.

Introduziremos agora algumas notações. Seja X_i ($X_i > 0$) a informação disponível sobre a empresa i (quantidades produzidas ou vendas totais, por exemplo), que opera em uma indústria, compreendendo n empresas: $i = 1, 2, \dots, n$. Suponha que as unidades de medida usadas para quantificar as informações sejam comuns às várias empresas (por exemplo, unidades ou toneladas para as quantidades; reais ou dólares para os valores monetários), de maneira que a agregação das informações das várias empresas seja possível. Assim, faz sentido calcularmos a informação agregada disponível para a indústria $X = \sum_{i=1}^n X_i$ e as parcelas de mercado de cada empresa $s_i = \frac{X_i}{X}$.

Suponha que as empresas são classificadas em ordem decrescente, de acordo com sua posição no mercado: $X_1 \geq X_2 \geq \dots \geq X_n$, de modo que a empresa 1 é a maior do mercado ($s_1 \geq s_2 \geq \dots \geq s_n$).

5.2.3 RAZÕES DE CONCENTRAÇÃO

A razão de concentração de ordem k é um índice positivo que fornece a parcela de mercado das k maiores empresas da indústria ($k = 1, 2, \dots, n$). Assim,

$$CR(k) = \sum_{i=1}^k s_i \quad (1)$$

Quanto maior o valor do índice, maior é o poder de mercado exercido pelas k maiores empresas. Nas aplicações empíricas, toma-se comumente $k=4$ ou $k=8$, isto é, considera-se apenas a participação das quatro ou das oito maiores empresas. As respectivas razões de concentração são conhecidas como $CR(4)$ e $CR(8)$. Contudo, observamos algumas deficiências imediatas dos índices CR :

1. Eles ignoram a presença das $n-k$ empresas menores da indústria. Desse modo, fusões horizontais ou transferências de mercado que ocorrem entre elas não alterarão o valor do índice, se a participação de mercado da nova empresa (resultante da fusão) ou das empresas beneficiárias (das transferências) se mantiver abaixo da k -ésima posição;
2. Esses índices não levam em conta a participação relativa de cada empresa no grupo das k maiores. Assim, importantes transferências de mercado que ocorrerem no interior do grupo (sem exclusão de nenhuma delas) não afetarão a concentração medida pelo índice.

Essas omissões dificultam o uso do $CR(k)$ como uma medida do poder de mercado (ou do grau de competição) existente na indústria. Considera-se também que o uso desse índice para acompanhar a evolução da estrutura industrial ao longo do tempo poderá acarretar inconsistências de natureza, uma vez que as k empresas de referência podem não ser as mesmas entre dois períodos.

Tais dificuldades servem de motivação para a consideração de medidas *sumárias* que possuam propriedades mais atraentes. Nas seções seguintes discute-se duas dessas medidas e considera-se, do ponto de vista comparativo, a adequação das razões de concentração.

5.2.4 ÍNDICE DE HIRSCHMAN-HERFINDAHL (HH)

Trata-se do índice positivo definido por:

$$HH = \sum_{i=1}^n s_i^2. \quad (2)$$

Tal expressão pode ser reescrita como $\sum_{i=1}^n s_i(s_i)$, o que evidencia a estrutura de pesos implícita no índice HH . Elevar cada parcela de mercado ao quadrado implica atribuir um peso maior às empresas relativamente maiores. Assim, quanto maior for HH , mais elevada será a concentração e, portanto, menor a concorrência entre os produtores.

O índice HH varia entre $1/n$ e 1. O limite superior do índice está associado ao caso extremo de *monopólio* no qual uma única empresa opera no mercado. O limite inferior decorre de que HH é uma função convexa definida no simplex $S_{n-1} = \left\{ S \in [0,1]^n : \sum_{i=1}^n S_i = 1 \right\}$.

Assim, o índice assume o valor mínimo $HH = 1/n$ para $s_1 = s_2 = \dots = s_n$, isto é, quando todas as empresas têm o mesmo tamanho ($s_i = 1/n$). Temos então³: $1/n \leq HH \leq 1$.

Observa-se que o limite inferior de HH decresce à medida que aumenta o número de empresas e que no limite (para $n \rightarrow \infty$) ele tende para zero.

Entretanto, isto não significa que HH sempre decresça com o aumento do número de empresas. A entrada de uma empresa adicional na indústria é compatível tanto com um aumento quanto com uma redução na concentração medida por HH . Isso poderá ser visualizado expressando-se o valor do índice em termos do *coeficiente de variação* das parcelas de mercado. Com efeito, como a média destas é $\bar{s} = 1/n$, a variância é $V(s) = (1/n)HH - (1/n)^2$. Lembrando que o coeficiente de variação empírico CV de uma variável é definido como a razão entre seu desvio-padrão e a sua média, obtemos a partir da expressão anterior: $(CV_s)^2 = V(s)/(1/n)^2 = nHH - 1$, ou, ainda,

$$HH = \frac{1}{n} [CV_s^2 + 1]. \quad (2a)$$

Nessa expressão, vemos que HH depende tanto do número de empresas quanto da dispersão relativa da repartição do mercado entre elas (CV_n). Como o valor da dispersão também depende de n , o efeito da entrada de uma nova empresa sobre a concentração é assinalado de maneira ambígua pela fórmula (2a). Tudo depende de como as empresas estabelecidas “acomodarão” a entrada da nova empresa. Se a acomodação não aumentar significativamente ou mesmo reduzir a dispersão preexistente entre as parcelas de mercado, então a entrada diminuirá a concentração. Ao contrário, se a acomodação aumentar significativamente a dispersão das participações no mercado, a entrada aumentará a concentração. Nesse caso, a entrada da nova empresa (dotada de eficiência muito acima ou muito abaixo da eficiência das empresas já instaladas) não será favorável para a concorrência industrial.

5.2.4.1 Interpretação teórica

Ao considerarmos o equilíbrio de um oligopólio homogêneo em competição de Cournot, podemos obter uma relação entre o índice HH e o grau de lucratividade da indústria, o que permite atribuir ao índice uma interpretação teórica. Esse aspecto torna o HH mais atraente que os demais, pois sua fórmula deixa de ser *ad hoc* (arbitrária), uma vez que está relacionada a um processo de otimização reconhecido em teoria econômica (maximização do lucro pelas empresas).

Com efeito, sob a hipótese de Cournot, cada empresa i escolhe as quantidades q_i que maximizam seu lucro individual $\pi_i(q_1, q_2, \dots, q_i, \dots, q_n) = q_i P \left(\sum_{i=1}^n q_i \right) - C_i(q_i)$, não antecipando qualquer reação dos competidores a uma mudança nas suas quantidades escolhidas (conjecturas nulas). Então, se as funções inversas de demanda indireta (P) e de custo (C_i) forem contínuas e diferenciáveis e tais que π_i seja côncava em q_i , a condição de primeira ordem $P + q_i P' - C'_i = 0$ é necessária e suficiente para determinar a solução ótima. Usando-se as definições $\lambda_i = (P - C'_i)/P$ para o índice de lucratividade de Lerner e $\varepsilon = -(\partial Q/\partial P)(P/Q)$ para a elasticidade-preço da demanda, a condição citada pode ser escrita como: $\lambda_i = s_i/\varepsilon$. Logo, $\sum_{i=1}^n \lambda_i^2 = HH/\varepsilon^2$ e, como $\sum_{i=1}^n \lambda_i = 1/\varepsilon$, vem:

$$HH = \frac{\sum_{i=1}^n \lambda_i^2}{\sum_{i=1}^n \lambda_i^2 + \sum_{j,k(j \neq k)} \lambda_j \lambda_k}.$$

Quanto maior é o grau de competição entre as empresas na indústria, maior é o quociente $(\sum_{j,k(j \neq k)} \lambda_j \lambda_k) / \sum_{i=1}^n \lambda_i^2$ e, em consequência, menor o índice HH . Usando o critério da otimização condicionada de uma função de bem-estar côncava para o produto industrial, Dansby e Willig (1979) mostram que $\sum_{i=1}^n \lambda_i^2$ é uma medida inversa da performance industrial para o caso homogêneo: quanto mais elevada a soma, mais afastada da posição ótima se acha a posição atual da indústria (q_1, q_2, \dots, q_n), o que (dada a proporcionalidade existente entre essa soma e o índice) permite considerar o HH também como um índice normativo.

5.2.4.2 Uso normativo: Regulação e defesa da concorrência⁴

Desde o início de 1980, a Federal Trade Commission dos Estados Unidos tem defendido a utilização do índice Hirschman-Herfindahl em substituição ao índice $CR(4)$ para fins de política antitruste. As orientações emitidas em 1992 a respeito de processos

de fusões (Merger Guidelines) instituem bandas referenciais com o objetivo de balizar a análise das fusões potenciais entre empresas.

Com efeito, como HH é uma função convexa das parcelas de mercado, o efeito potencial das fusões horizontais entre duas ou mais empresas (sem alteração nas parcelas de mercado das outras) sempre resultará em aumento na concentração medida pelo índice. Suponha que as duas primeiras empresas (1 e 2) realizem uma fusão; a contribuição da nova entidade para a concentração é dada por $(s_1 + s_2)^2 = s_1^2 + s_2^2 + 2s_1s_2$, a qual é maior que $s_1^2 + s_2^2$, a contribuição dada anteriormente à concentração pelas duas empresas separadamente.

As agências antitruste trabalham com índices HH calculados a partir de participações de mercado medidas com base em 100 (percentuais) e não com base em 1 (razões decimais), como fizemos aqui. Nesse caso, o índice pode potencialmente variar entre 0 e 10.000. Três são as faixas propostas no Mergers Guidelines para balizar as análises preliminares de processos de fusões, considerando-se os valores potenciais do índice após a fusão entre duas empresas:

1. $0 \leq HH < 1.000$: não existe preocupação quanto à competição na indústria, caso a fusão se concretize.
2. $1.000 \leq HH \leq 1.800$: existe preocupação quanto à competição se o aumento do índice for maior ou igual a 100 pontos, com relação ao índice pré-fusão.
3. $HH > 1.800$: existe preocupação quanto à competição se o aumento do índice for maior ou igual a 50 pontos, com relação à situação inicial (pré-fusão).

Um exemplo ilustrará o uso desses critérios. Considere uma indústria composta de quatro empresas com as participações de mercado (percentuais) de 40, 30, 20 e 10. Nesse caso, o índice inicial (pré-fusão) é $HH_{ante} = 40^2 + 30^2 + 20^2 + 10^2 = 3.000$. Suponha agora que a terceira e a quarta empresas abram processo de fusão na agência reguladora. Ao calcular o valor potencial do índice pós-fusão, o regulador obterá: $HH_{pós} = 40^2 + 30^2 + 30^2 = 3.400$. Esse valor inclui o processo na terceira faixa. Como a variação no índice $HH_{pós} - HH_{ante} = 400$ supera 50 pontos, o regulador obterá com isto um sinal de alerta contra tal fusão.

Note que ao calcular o valor potencial $HH_{pós}$, o regulador está adotando duas hipóteses simplificadoras: (i) a fusão entre as empresas não gera qualquer sinergia produtiva ou mercadológica para a nova empresa, uma vez que a posição atribuída a esta resulta da soma das posições das empresas originais; (ii) a fusão não afeta as posições de mercado das empresas não participantes, as quais se supõem permanecerem neutras e, conjuntamente, não “acomodarem” a eliminação de duas concorrentes e o surgimento de uma concorrente maior.

Apesar dessas limitações, as faixas propostas no Mergers Guidelines fornecem critérios diretores muito úteis para uma análise preliminar das propostas de fusão horizontal. Seu emprego não dispensa, todavia, detalhamentos posteriores.

5.2.4.3 Números equivalentes

Outro aspecto de interesse para a interpretação de um índice de concentração consiste em saber se este possui ou não a propriedade dos *números equivalentes*. Seja s o vetor das parcelas de mercado das n empresas em operação na indústria e C o valor da concentração industrial calculado pela fórmula: $C = C(s)$

O *número equivalente* associado ao índice $C(\cdot)$ define-se como o número de empresas (v_C) que torna consistente o valor C com o valor do índice obtido para uma indústria composta de v_C empresas de *igual tamanho*.

Assim, o número equivalente v_C resolve em v a equação: $C(\frac{1}{v}, \frac{1}{v}, \dots, \frac{1}{v}) = C$. Como bons índices de concentração apresentam valores mínimos no caso simétrico (empresas de mesmo tamanho), temos $C \geq 1/n$. Usualmente, se requer que $C(\frac{1}{v}, \frac{1}{v}, \dots, \frac{1}{v})$ seja decrescente em v . Então, se esse requerimento for satisfeito, podemos afirmar que $v_C \leq n$. Assim, o número equivalente de empresas nunca será superior ao número efetivo.

Obviamente, v_C pode não ser um número inteiro, como seria desejável. Entretanto, seu valor serve como um indicador qualitativo do desempenho aparente da estrutura industrial. Quanto maior o valor de v_C (mais próximo de n), menos a estrutura do mercado restringe a competição na indústria, isto é, menos entraves existem para as empresas individuais exercerem plenamente seu poder concorrencial.

No caso do índice Hirschman-Herfindahl obtemos

$$HH = \sum_{i=1}^v (1/v)^2 = v(1/v)^2 = 1/v$$

de maneira que o número equivalente associado é: $v_H = 1/HH$.

5.2.4.4 Padronização

Ao longo do tempo, a variação do índice de concentração pode ser significativa. Por isso, alguns autores propuseram ajustes na sua fórmula para facilitar análises comparativas. Por exemplo, ajustes na amplitude de variação HH foram propostos, definindo-se o índice transformado:

$$HH' = \frac{1}{n-1} [nHH - 1]; \quad n > 1.$$

Substituindo os valores extremos de HH na fórmula ajustada obtemos os limites 0 e 1, que são independentes de n .⁵

Versões ajustadas do índice HH podem ser úteis em algumas comparações intertemporais, quando se deseja (por alguma razão) controlar a concentração pelo número de empresas. Entretanto, é bom lembrar que n é um dos fatores determinantes do grau de concorrência existente na indústria (quanto maior o número de empresas, maior será esse grau), de modo que o uso de coeficientes ajustados (como o HH' anterior) somente se justifica quando a análise comparativa efetivamente demanda indicadores padronizados para a concentração.

5.2.5 ÍNDICE DE ENTROPIA DE THEIL (ET)

Esse índice foi proposto por Theil (1967) no contexto da Teoria da Informação, mas ele pode ser aplicado em *Economia Industrial*, como veremos adiante. Seja A um evento genérico e p a sua probabilidade de ocorrência. A uma mensagem confirmando a ocorrência de A vem associado o que chamamos *grau de surpresa*, que varia na proporção inversa de p . Pelo seu lado, o *conteúdo informacional* da mensagem, notado h , é considerado uma função crescente e derivável do grau de surpresa associado à ocorrência de A , de modo que podemos expressá-lo como uma função decrescente de p , a probabilidade do evento relatado na mensagem: $h(p)$, com $h' < 0$. Theil particulariza a função h definindo-a por meio da função logaritmo: $h(p) = \ln(1/p) = -\ln(p)$.

O argumento anterior pode ser estendido para o caso de n eventos A_1, A_2, \dots, A_n com probabilidades p_1, p_2, \dots, p_n . Essas probabilidades somam 1 se pelo menos um desses eventos certamente ocorrer. Pode-se calcular o *conteúdo informacional esperado* da mensagem relatando a ocorrência de um destes eventos tomando-se a esperança matemática de $h(p)$ com relação à distribuição de probabilidade p_1, p_2, \dots, p_n :

$$E(h) = \sum_{i=1}^n p_i h(p_i).$$

Assim fica definido o *índice de entropia*, notado T . Usando-se a especificação de Theil para o conteúdo informacional obtemos:

$$ET = \sum_{i=1}^n p_i \ln(1/p_i) \tag{3}$$

O índice de entropia pode ser interpretado como uma medida *inversa* da concentração. Com relação a isso, Braga e Mascolo (1982) particularizam a interpretação da medida no contexto da *Economia Industrial* substituindo p_i pela parcela de mercado s_i da firma i . Reescrevendo a expressão (3) em termos de s_i tem-se uma medida inversa da concentração:

$$ET = -\sum_{i=1}^n s_i \ln(s_i). \tag{3a}$$

Assim, dada a ocorrência de uma venda no mercado industrial, a probabilidade que essa venda tenha sido efetuada pela empresa i é s_i , e a contribuição desta para o conteúdo informacional presente na mensagem é $-s_i \ln(s_i)$, de modo que o índice ET estará indicando o conteúdo informacional esperado da ocorrência, calculado sobre todas as empresas da indústria. Trata-se, portanto, da informação prestada pela “empresa média”. Quanto maior for esta, menor será o grau de surpresa associado à mensagem e menor o índice de entropia; em consequência, maior será a concentração na indústria.

Note que o limite inferior do índice (que corresponde à concentração máxima) é igual a zero. Esse valor ocorre em situação de monopólio, quando existe apenas um produtor ($s=1$).

Para encontrarmos o limite superior de ET , usamos o fato de que $-ET$ é uma função convexa definida no simplex S_{n-1} . Assim, $-ET$ assume o valor mínimo quando todas as empresas são iguais ($s_i = 1/n$). Nesse caso, o índice valerá

$$-\sum_{i=1}^n \frac{1}{n} \ln\left(\frac{1}{n}\right) = \ln(n),$$

e teremos então:

$$0 \leq ET \leq \ln(n).$$

Podemos ainda determinar o *número equivalente* de empresas de igual tamanho consistente com o índice de entropia. Como acabamos de verificar, participações de mercado idênticas nos levam a $ET = \ln(v)$. Assim, o número equivalente é dado por $v_T = e^{ET}$. Quando ET for mínimo, esse iguala 1; quando for máximo, iguala n :

$$1 \leq v_T \leq n.$$

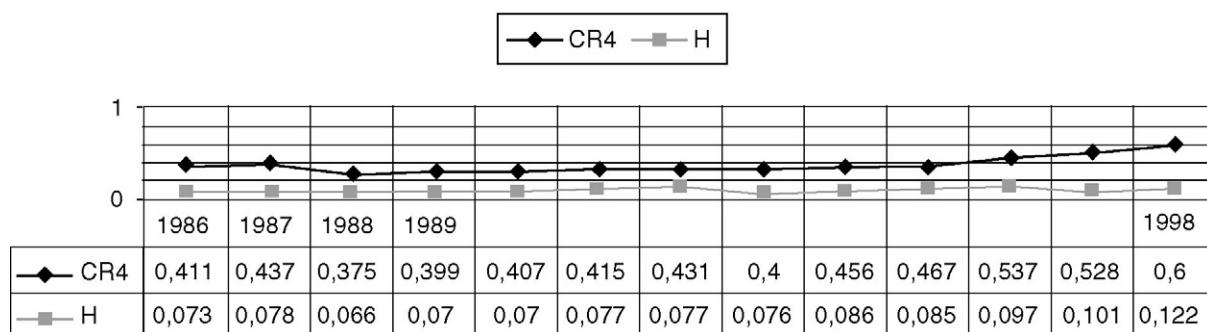
Também, nesse caso, pode-se conceber uma versão ajustada do índice de entropia para que sua amplitude de variação seja unitária, com $0 \leq T' \leq 1$. Para tanto, basta definir índice ajustado como:

$$T' = -\frac{1}{\ln(n)} \sum_{i=1}^n s_i \ln(s_i)$$

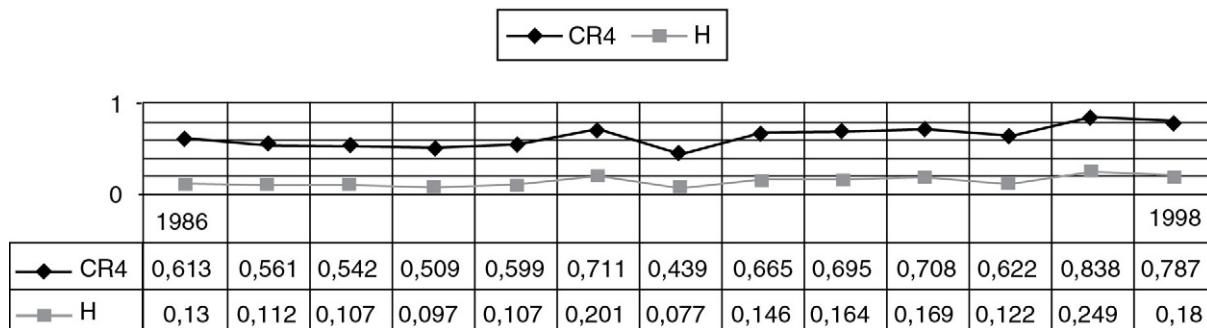
5.3 Uma Ilustração

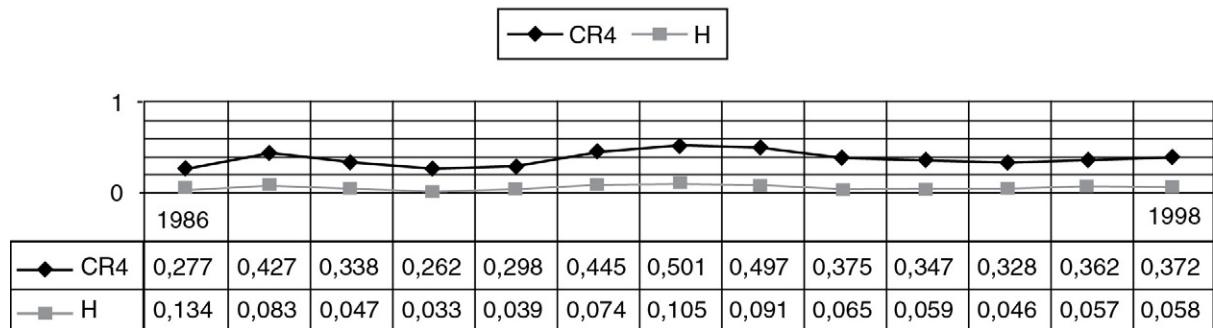
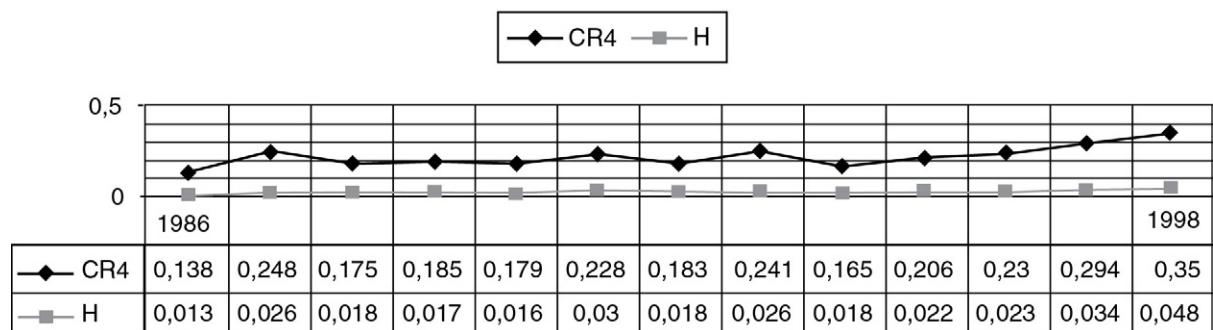
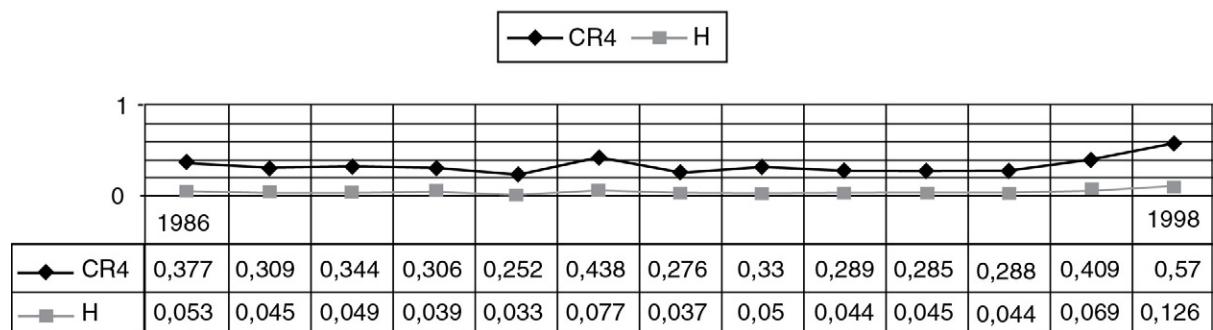
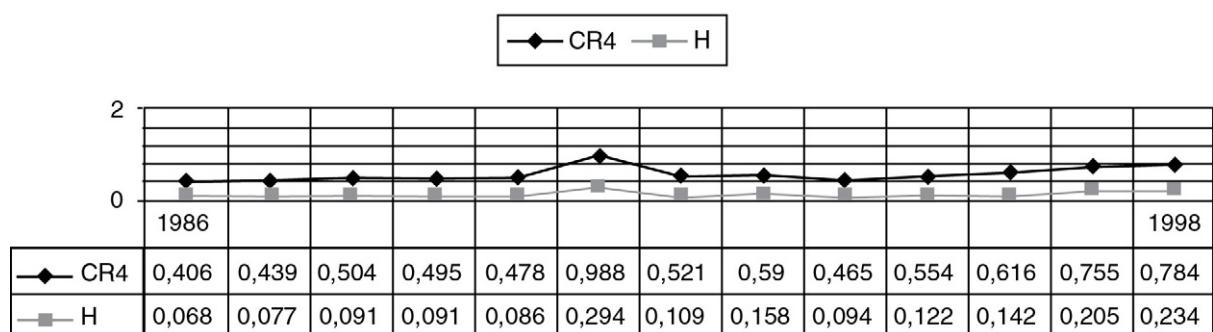
Nesta seção apresentamos valores dos índices anteriormente discutidos para setores industriais selecionados da economia brasileira. Uma dificuldade fundamental associada ao cálculo desses índices refere-se à disponibilidade de dados em um nível suficientemente desagregado. Idealmente, o cálculo das participações de mercado embutido nesses índices requer dados para setores no nível de quatro dígitos, de acordo com a Classificação Nacional das Atividades Econômicas do IBGE. Estes raramente se encontram disponíveis para os segmentos mais significativos da indústria brasileira. Isso explica a escassez de estudos detalhados para o Brasil, excetuando-se o trabalho de Braga e Mascolo (1982) que analisou o ano de 1978. A base de dados disponível apresenta duas deficiências principais: em primeiro lugar, o nível de classificação é o de três dígitos; em segundo lugar, os dados não refletem o desempenho da totalidade da indústria, mas apenas uma parte dela, mais particularmente, a das mil maiores empresas do Brasil, de acordo com as informações coletadas pelo *Centro de Estudos Empresariais* da Fundação Getúlio Vargas (período 1986-98). Contudo, calculamos os índices para os segmentos industriais mais homogêneos para obter valores mais consistentes. Os setores representados a seguir são: adubos e fertilizantes; tintas e vernizes; celulose, papel e papelão; têxtil; plásticos; e material cerâmico.

Adubos & Fertilizantes



Tintas & Vernizes



Celulose, Papel & Papelão**Têxtil****Plásticos****Material Cerâmico**

Pelos gráficos apresentados percebemos que os movimentos de concentração ou desconcentração ocorridos no seio das quatro maiores empresas refletem, de um modo geral, a dinâmica da concentração de todo o segmento industrial considerado. Notemos também que, com exceção do segmento celulose e derivados, todos os demais mostram tendência geral para aumento da concentração a partir de 1994, seja ela medida por $CR(4)$, ou por HH . Tal resultado é, todavia, uma casualidade, pois em outras amostras podem exibir trajetórias bastante distintas.

5.4 Escolha da Medida de Concentração

A discussão das seções anteriores evidencia a precariedade de medidas mais simplificadas, como as razões de concentração, e motiva o uso de medidas sumárias como o índice de Hirschman-Herfindahl. Exploraremos aqui, mais detalhadamente, a relação entre razões de concentração e o índice HH e apresentaremos alguns critérios axiomáticos encontrados na literatura. De um ponto de vista meramente empírico, existe certa evidência de que os diferentes índices de concentração poderiam estar captando dimensões comuns nos dados. Bailey e Boyle (1971) defenderam a utilização de medidas mais simplificadas como o $CR(4)$ diante da elevada correlação obtida com medidas mais complexas para o caso norte-americano. Resende (1995) empregou análise de componentes principais para detectar dimensões comuns em um conjunto de medidas de concentração calculadas para o caso norte-americano. Uma discussão fundamentada em critérios aceitáveis permitirá uma melhor avaliação comparativa dos diferentes índices.

5.4.1 RAZÕES DE CONCENTRAÇÃO E ÍNDICE HIRSCHMAN-HERFINDAHL

Em uma primeira abordagem, Boyes e Smyth (1979) chamam a atenção para o fato de que uma perfeita correlação entre índices de concentração não implica perfeita equivalência entre eles. Considere-se uma sequência de valores C_t de uma medida de concentração C crescendo a uma taxa constante g por período, isto é, $C_{t+1} = (1+g) C_t ; t=1,2,3,\dots$

Para que a medida de concentração alternativa C^* seja considerada equivalente à medida C , ela deverá apresentar o mesmo padrão de crescimento intertemporal, isto é, deve também crescer à taxa g por período. Suponha agora que C e C^* sejam perfeitamente correlacionados ao longo da reta $C_t = a + bC_t^*, t=1,2,3,\dots$ Combinando as duas equações, obtemos: $C_{t+1}^* - C_t^* = (g/b)C_t$, ou ainda: $C_{t+1}^* = (ga/b) + (1+g)C_t^*$. Esta última expressão mostra que, a menos que as duas medidas de concentração sejam proporcionais ($a=0$), a taxa de crescimento intertemporal de C^* será variável e diferente de g . Assim, apesar das medidas C e C^* apresentarem correlação perfeita, elas não são equivalentes.

Mostraremos em seguida que os índices $CR(4)$ e HH não exibem proporcionalidade e que, portanto, não são equivalentes entre si. Com efeito, podemos escrever:

$$CR(4) = 4\bar{s}_4, \text{ onde } \bar{s}_4 = \frac{1}{4} \sum_{i=1}^4 s_i.$$

Por outro lado, o índice HH pode ser escrito como $4\bar{s}_4^2 + \sum_{i=5}^n s_i^2$, de modo que temos:

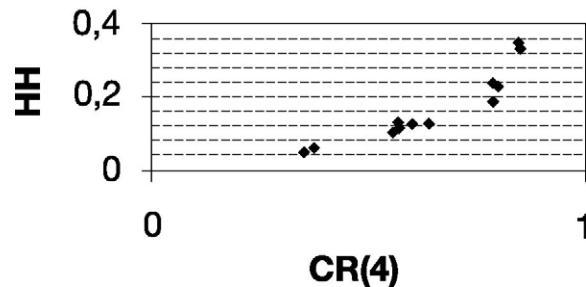
$$HH = \frac{CR^2(4)}{4} + \sum_{i=5}^n s_i^2.$$

Logo, não existe proporcionalidade entre esses dois índices.

A equação anterior, expressando HH como função de $CR(4)$, sugere a existência de uma relação convexa entre esses dois índices, muito embora esta não seja biunívoca. Com efeito, pode-se facilmente verificar, a partir daquela equação, que um mesmo valor de $CR(4)$ é compatível com uma infinidade de distribuições do mercado e, em consequência, com uma infinidade de valores de HH . Tal fato foi explorado por Sleuwaegen *et alii* (1989). Focalizando a indústria europeia, eles conseguem reproduzir empiricamente a relação convexa descrita anteriormente (*horn relation*).

A figura a seguir mostra, no plano $CR(4) \times HH$, os pontos calculados para 13 segmentos da indústria brasileira em 1998. Apesar da deficiência da base de dados já mencionada (agregação em nível três dígitos, mil maiores empresas), a relação convexa esperada entre esses índices aparece claramente no gráfico representado.

BRASIL (1998): CR(4) x HH



5.4.2 CRITÉRIOS AXIOMÁTICOS

Considerando um índice de concentração genérico C como uma função positiva das distribuições do mercado $s = (s_1, s_2, \dots, s_n)$ pertencentes ao simplex unitário $(n-1)$ -dimensional (S_{n-1}), Encaoua e Jacquemin (1980) propuseram cinco propriedades básicas que bons índices de concentração (positivos) devem atender. As três primeiras aplicam-se a indústrias fechadas, com número (n) fixo de empresas; as duas últimas aplicam-se a indústrias abertas, nas quais existem entrada e saída de empresas.

P1. Princípio da transferência: esse princípio estabelece que se uma empresa maior ocupar parte do mercado atendido por uma empresa menor, o índice de concentração não poderá diminuir. Formalmente:

$$\text{se } s_2 < s_1 \text{ e } 0 < \Delta < s_2, \text{ então } C(s_1 + \Delta, s_2 - \Delta, s_3, \dots, s_n) \geq C(s_1, s_2, s_3, \dots, s_n);$$

P2. Minimalidade em simetria: um bom índice de concentração deve apresentar valor mínimo no caso simétrico, quando todas as empresas detêm parcelas de mercado iguais. Formalmente:

$$C\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = \min_{s \in S_{n-1}} C(s)$$

P3. Critério de Lorenz: se em duas indústrias com o mesmo número de empresas e parcelas de mercado $s = (s_1, s_2, \dots, s_n)$ e $r = (r_1, r_2, r_3, \dots, r_n)$, respectivamente, a primeira distribuição dominar a segunda, a concentração na primeira deverá ser maior do que na segunda. Formalmente: se $C_s(k)$ é a razão de concentração de ordem K na indústria com parcelas s e $C_s(k) \geq C_r(k)$; $k = 1, 2, \dots, n$ então $C(s) \geq C(r)$;

P4. Não decrescimento em fusões horizontais: por esse princípio, fusões horizontais nunca são benéficas para a concorrência: a concentração medida *ex-post* é maior que ou igual à concentração *ex-ante*. Formalmente: se s'_1 é a parcela de mercado da nova empresa resultante da fusão das empresas 1 e 2, então: $C_{n-1}(s'_1, s_2, \dots, s_n) \geq C_n(s_1, s_2, s_3, \dots, s_n)$;

P5. Não crescimento em simetria: em uma indústria com empresas de igual tamanho na qual existe acomodação perfeita à entrada de uma nova empresa, o índice de concentração não aumenta. Formalmente $C\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$ é uma função não crescente de n .

Encaoua e Jacquemin (1980) mostraram que se a propriedade P1 for satisfeita, então P2 e P3 também serão satisfeitas. Mostram também que se P2 e P4 forem conjuntamente atendidas, P5 também o será. Se h for uma função definida no intervalo $[0,1]$, os autores estabelecem a convexidade da função $sh(s)$ como uma condição suficiente para que o índice $C = \sum_{i=1}^n s_i h(s_i)$ atenda à propriedade P1. Além disso, o não decrescimento de h aparece como uma propriedade suficiente para que o índice C (anterior) também verifique P4.

Quais dos índices aqui apresentados podem ser considerados boas medidas da concentração industrial, à luz das cinco propriedades apresentadas?

Dada a convexidade das funções $sh(s) = s^2$ e $sh(s) = \ln(s)$, os índices HH e $-ET$ verificam P1 e, em consequência, também P2 e P3. Por outro lado, como as funções h implícitas na definição dos dois índices são não decrescentes, eles também verificam P4. Logo, P5 também é satisfeita. Assim, os índices Hirschman-Herfindahl e o negativo do índice de Entropia (Theil) atendem os cinco requisitos propostos. Não é difícil verificar que as razões de concentração $CR(k)$ não atendem, necessariamente, os requisitos P1 e P4 (para $k = 1, 2, \dots, n-2$).

Notas

1. Ver Baker e Bresnahan (1992) a este respeito.
2. Embora o índice HH venha comumente associado à Herfindahl (1950), sua paternidade pode ser atribuída a Hirschman que, alguns anos antes (1945), utilizou a norma euclidiana do vetor das parcelas de mercado ($= \sqrt{HH}$) para medir a concentração industrial norte-americana (Hirschman, 1964).
3. Para se obter o limite inferior do índice considere o seguinte problema de minimização condicionada: $\sum_{i=1}^n s_i^2$ sujeito a $\sum_{i=1}^n s_i = 1$. A este problema de otimização vem associado o seguinte lagrangeano: $L = \sum_{i=1}^n s_i^2 - \lambda \left[\sum_{i=1}^n s_i - 1 \right]$. Em razão da convexidade do índice, as condições de primeira ordem sobre L são necessárias e suficientes para que obtenhamos um mínimo. Assim, igualando a zero as derivadas de L com relação a s_i e com relação ao multiplicador de Lagrange λ obtemos: $\frac{\partial L}{\partial s_i} = 2s_i - \lambda = 0$ e $\frac{\partial L}{\partial \lambda} = 1 - \sum_{i=1}^n s_i = 0$. A primeira equação implica $s_i = \lambda/2$ constante, de modo que, combinando $s_1 = s_2 = \dots = s_n$ com a segunda equação veín: $s_i = 1/n$.
4. Para o caso brasileiro, consulte o Capítulo 22.
5. Ver a este respeito, Srivastava e Aggarwal (1979).

Bibliografia

- Bailey, D. e Boyle, S.E. (1971) The Optimal Measure of Concentration, *Journal of the American Statistical Society*, 66, 702-706.
- Bain, J. (1956) *Industrial Organization*, Nova York: John Wiley e Sons.
- Baker, J.B. e Bresnahan, T.F. (1992) Empirical Methods of Identifying and Measuring Market Power, *Antitrust Law Journal*, 61, 3-16.
- Barbosa, F.H. (1981) Medidas de Concentração, *Revista de Econometria*, 1, 31-53.
- Blackorby, C., Donaldson, D. e Weymark, J.A. (1982) A normative approach to industrial-performance evaluation and concentration indices, *European Economic Review*, 19, 89-121.
- Boyce, W.J. e Smyth, D.F. (1979) The Optimal Concentration Measure: Theory and Evidence for the Canadian Manufacturing Industry, *Applied Economics*, 11, 289-302.
- Braga, H.C. e Mascolo, J.L. (1982) Mensuração da Concentração Industrial no Brasil, *Pesquisa e Planejamento Econômico*, 12.
- Carlton, D.W. e Perloff, J.M. (1994) *Modern Industrial Organization*, Nova York: Harper-Collins.
- Dansby, R.E. e Willig, E.D. (1979) Industry performance gradient indices, *The American Economic Review* 69(3), 249-260.
- Davies, S. (1988) Concentration, In S. Davies e B. Lyons (orgs.), *Economics of Industrial Organization*, Essex: Longman, 73-126.
- Encaoua, D. e Jacquemin, A. (1980) Degree of monopoly, Indices of Concentration and Threat of Entry, *International Economic Review*, 21, 87-105.
- Filkenstein, M.O. e Frieberg, R.M. (1967) The Application of an Entropy Theory of Concentration to the Clayton Act, *Yale Law Journal*.
- Herfindahl, O.C. (1950) Concentration in the Steel Industry, Columbia University, dissertação de Ph.D não publicada.
- Hoffman, R. (1998) *Estatística para Economistas*, São Paulo: Editora Pioneira.
- Resende, M. (1994) Medidas de Concentração Industrial: uma Resenha, *Análise Econômica*, 12, 24-33.
- Resende, M. (1995) On the Characterization of Market Structure in U.S. Industry: a Principal Components Analysis, *Economia*, 19, 87-94.
- Sleuwaegen, L.E., De Bondt, R.R. e Dehandschutter, W.V. (1989) The Herfindahl Index and Concentration Ratios Revisited, *Antitrust Bulletin*, 34, 625-640.
- Srivastava, D.K. e Aggarwal, P.K. (1979) Measuring Tax Revenue Centralization Federal Systems: a Case Study of India, *Public Finance*, 34, 414-433.
- Sutton, J. (1991) *Sunk Costs and Market Structure*, Cambridge: MIT Press.
- Theil, H. (1967) *Economics and Information Theory*, Rand McNally.

Diferenciação de Produtos

Luciano Losekann e Margarida Gutierrez

6.1 Introdução

Como vimos no Capítulo 1, a hipótese de homogeneidade dos produtos é essencial para a definição da concorrência perfeita. Neste arcabouço analítico, como os consumidores consideram os produtos idênticos (isto é, substitutos perfeitos), a ocorrência de preços distintos não permite à empresa com preço superior realizar vendas, pois os consumidores optam por adquirir apenas da empresa com menor preço. Por isso, quando são considerados mercados sem diferenciação de produtos, respeitando as demais condições competitivas, o preço é único, sendo definido pelo mercado. Nesse caso, as empresas são *tomadoras de preços (price takers)*, já que não podem influenciar o preço de mercado e, na verdade, se deparam com curvas de demandas perfeitamente elásticas.

Na prática, dificilmente as mercadorias que competem entre si são idênticas. Quando vamos a um supermercado, nos deparamos com produtos similares, mas não idênticos, que apresentam preços distintos. Contrariando o modelo de concorrência perfeita, muitas vezes os consumidores se dispõem a pagar preços superiores. Por exemplo, preferem abastecer o carro em postos de gasolina próximos à sua residência ou consumir seu refrigerante preferido, mesmo que seus preços sejam mais elevados.

O que explica esse comportamento é a diferenciação do produto. Os produtos são diferenciados segundo diversos aspectos como: local da oferta, qualidade do produto ou percepção da marca. Rigorosamente, basta que os consumidores percebam os produtos como diferentes, isto é, que tenham preferências subjetivas distintas, para ocorrer diferenciação de produto. Muitos produtos que apresentam características físicas idênticas são percebidos como diferentes pelos consumidores em função da marca. As empresas orientam parte relevante de seus recursos para consolidarem as marcas perante os consumidores.

Como os produtos diferenciados são substitutos imperfeitos, uma empresa pode fixar preços acima das demais e realizar vendas. Incorporar a diferenciação do produto como uma estratégia possível torna-se importante, na medida em que a maior parte das empresas produz uma grande variedade de bens/serviços e que uma grande parte das empresas que utilizam a diferenciação como uma estratégia de mercado tem algum poder de mercado. Dessa forma, empresas que atuam em mercados onde os produtos são diferenciados se defrontam com uma demanda residual inclinada,¹ havendo espaço para *fixação de preços*.

Em função de sua inadequação às situações reais, a hipótese de produto homogêneo é bastante criticada desde o início do século. Como vimos no Capítulo 1, o modelo de competição monopolística desenvolvido por Chamberlin, em 1933, considerado um marco no surgimento da Economia Industrial, incorpora a diferenciação de produtos. Posteriormente, muitos outros modelos, que utilizam vários aspectos do arcabouço da microeconomia tradicional, foram desenvolvidos para analisar esse aspecto.

As abordagens heterodoxas também dedicam parte relevante de sua análise para o processo de diferenciação. Na corrente evolucionista, seguindo a analogia com a diversidade de espécies, o processo de diferenciação resulta na criação de diversidades

que são expostas ao ambiente seletivo. Essa abordagem analisa o resultado do processo de diferenciação em termos da geração de variedades e, principalmente, o seu efeito sobre a dinâmica da estrutura da indústria.

Este capítulo contém quatro seções, além desta introdução. Primeiramente são apresentados os fatores e os tipos de diferenciação de produtos. Na terceira seção, é desenvolvida a formalização do modelo de concorrência monopolística. Na quarta seção, é apresentada a classe de modelos para analisar o processo de diferenciação mais enfatizada pela literatura de Economia Industrial. Dois modelos locacionais são abordados: a cidade linear, desenvolvido por Hotteling, e a cidade circular, elaborado por Salop. Dada a característica do livro, os modelos foram expostos em versões simplificadas, mas que captam sua essência. Na última seção, são enfocadas as críticas aos modelos e a contribuição da corrente heterodoxa.

6.2 Fatores e Tipos de Diferenciação

Como a diferenciação de produtos decorre de fatores subjetivos, qualquer listagem das possibilidades de diferenciação é incompleta. Em linhas gerais, os produtos são diferenciados conforme os seguintes atributos: especificações técnicas; desempenho ou confiabilidade; durabilidade; ergonomia e *design*; estética; custo de utilização do produto; imagem e marca; formas de comercialização; assistência técnica e suporte ao usuário; financiamento aos usuários.

É importante destacar que há setores que apresentam uma *vocação* maior para a diferenciação, condicionada às características do produto (tecnológicas e possibilidades de uso) e/ou às dos seus consumidores. Isso porque a diferenciação implica um produto *novo* e para que essa estratégia tenha sucesso é preciso que o produto possa ser modificado e que, de fato, os consumidores o considerem um produto melhor que os existentes. Na verdade, a introdução de um produto *novo* depende, antes de tudo, dos critérios de avaliação dos consumidores. Há produtos que são avaliados por um único critério, como, por exemplo, por seu poder calorífico, e há outros que são avaliados por vários critérios (conforto, segurança, economia etc.). Evidentemente, existe uma maior possibilidade de competir por diferenciação quando se tratam de produtos avaliados em múltiplas dimensões. No Quadro 6.1 apresentamos um exemplo de multidimensionalidade do produto *software*.

Existem dois tipos de diferenciação de produto: a horizontal e a vertical. Duas variedades de um produto são consideradas verticalmente diferenciadas quando um dos produtos apresenta atributos mais desejáveis que o outro. Dessa forma, em condições de preços iguais, todos os consumidores escolhem apenas o melhor produto. Em uma definição mais rigorosa, a

QUADRO 6.1 MULTIDIMENSIONALIDADE E TIPOS DE USUÁRIOS DE SOFTWARES

Os programas para computadores (*softwares*) são um exemplo de produto que possui várias dimensões para a realização de diferenciação. A Tabela 6.1 relaciona as dimensões de programas sujeitos à diferenciação e os tipos de usuários que lhes associam ganhos de valor.

TABELA 6.1 Dimensões de Diferenciação e Tipos de Usuários de Softwares

Dimensões	Usuários
Demora	Usuários pacientes/impacientes
Interface com o usuário	Usuários ocasionais/experientes
Conveniência	Usuários comerciais/domésticos
Resolução de imagem	Usos como boletim informativo/sofisticado
Velocidade de operação	Usuários estudantis/profissionais
Formato	Usos na tela/impressos
Capacidade	Usos gerais/específicos
Características	Usuários ocasionais/frequentes
Abrangência	Usuários leigos/profissionais
Aborrecimento	Usuários que atribuem muito/pouco valor ao tempo
Suporte	Usuários ocasionais/intensivos

Fonte: Shapiro e Varian (1999)

diferenciação é considerada vertical quando a utilidade de todos os consumidores aumenta quando o nível de uma característica do produto é aumentado. Normalmente, em mercados onde os produtos são verticalmente diferenciados, os diferenciais de preços são elevados. A diferenciação vertical de produto ocorre, por exemplo, entre dois carros do mesmo modelo, mas um com maior potência e conforto do que o outro. Ou seja, todos os consumidores se beneficiam quando o conforto de um carro é aumentado.

A diferenciação horizontal ocorre quando os produtos não podem ser considerados melhores ou piores, ou seja, não se pode ordenar a qualidade dos produtos. Isso ocorre porque, em situações de preços iguais, nem todos os consumidores escolhem a mesma variedade. A escolha irá depender do “gosto” do consumidor. Novamente, segundo uma definição estrita, a diferenciação é considerada horizontal quando a modificação em um atributo do produto causa aumento na utilidade de alguns consumidores e diminuição na de outros. Um exemplo de diferenciação horizontal é a cor de carros, alguns consumidores preferem carros pretos e outros, carros brancos. Normalmente, em situações de diferenciação horizontal, os preços das variedades são bastante próximos. Ainda que existam modelos que abordem a diferenciação vertical, os modelos mais conhecidos, que são apresentados neste capítulo, tratam da diferenciação horizontal.

6.3 Modelos de Competição Monopolística

Os modelos de concorrência monopolística, como o nome sugere, combinam elementos de situações de monopólio e de concorrência perfeita. Assim como no monopólio, as empresas não são consideradas tomadoras de preços. Ou seja, há espaço para a empresa decidir o preço cobrado, já que ela se defronta com uma curva de demanda residual negativamente inclinada, o que indica um poder de monopólio.

Por outro lado, as empresas estão sujeitas à concorrência de empresas que oferecem produtos substitutos próximos. Como a entrada é considerada livre nesses modelos, o resultado é um lucro econômico nulo, já que a ocorrência de lucros positivos atrai novas empresas até que este se iguale a zero novamente.

O modelo do consumidor representativo foi desenvolvido por Edward Chamberlin em 1933, o primeiro que incorpora a diferenciação de produto, e por isso é considerado uma das principais origens da Economia Industrial. Exetuando as hipóteses referentes à diferenciação de produto, as demais hipóteses utilizadas no modelo são as mesmas da concorrência perfeita apresentadas no Capítulo 1. A diferenciação do produto é considerada segundo duas hipóteses: os produtos são substitutos próximos e, apesar de produtos diferentes, a demanda e os custos são uniformes entre as empresas.

O objetivo final do modelo é verificar se o número de variedades é socialmente ótimo. A diferenciação de produtos gera um benefício decorrente do aumento das possibilidades de as características do produto se adequarem ao gosto do consumidor. Por outro lado, a criação de novos produtos gera maiores custos fixos associados aos esforços de venda (gastos em publicidade e outras estratégias de *marketing*). Assim, o modelo verifica se o número de variedades gerado por estímulos de mercado é o que maximiza a diferença entre benefícios e custos sociais.

Como vimos no Capítulo 1, se os produtos são diferenciados, a definição de mercado não é automática. Chamberlin utilizou o conceito de grupo de produtos, definido por agrupar produtos substitutos próximos (elasticidade preço cruzada elevada).

Os Gráficos 6.1a e 6.1b, que utilizam curvas de demanda lineares, descrevem o modelo de Chamberlin. As empresas atuam maximizando lucros, escolhendo o nível de produção que iguala a receita marginal ao custo marginal. Em uma situação inicial (Gráfico 6.1a), as empresas escolhem produzir q_i , com preços iguais a p_i , e realizam lucros extraordinários (área destacada no gráfico). Como a entrada é livre, essa situação não é sustentável.

A ocorrência de lucros econômicos atrai novas empresas para esse grupo de produtos. Com a entrada de novas empresas, que ocorre até igualar os lucros extraordinários a zero, há um deslocamento da curva de demanda residual de D_r para D'_r (Gráfico 6.1b). O novo ponto de equilíbrio (p'_i, q'_i) ocorre quando a curva de demanda residual é tangente à curva de custos médios, já que nesse ponto não há incentivos para as empresas entrarem ou saírem desse grupo de produtos, pois os lucros econômicos desaparecem.

O nível de produção socialmente ótimo ocorre quando as empresas são eficientes, ou seja, produzem ao menor custo médio possível. No Gráfico 6.1b, o nível de produção socialmente ótimo ocorre no ponto (p^*, q^*) . Assim, o resultado do modelo é que o nível de produção de equilíbrio é socialmente ineficiente, já que as empresas atuam em escalas em que o custo médio não é mínimo. Isto é interpretado da seguinte forma: o excesso de diversidade de produtos não permite às empresas realizarem a totalidade das economias de escala disponíveis. Ou seja, os custos fixos, que o autor frisa como aqueles resultantes dos esforços de vendas derivados da criação de novos produtos, não são diluídos de forma ótima, ocorrendo uma situação de excesso de capacidade.²

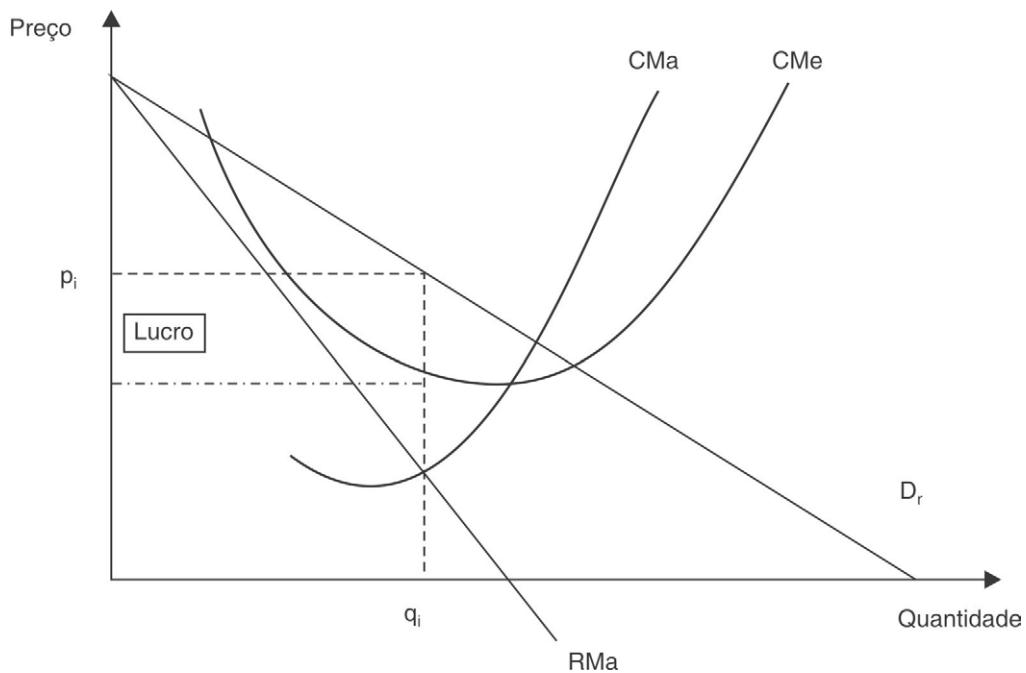


Gráfico 6.1a Modelo de Chamberlin antes da entrada de empresas.

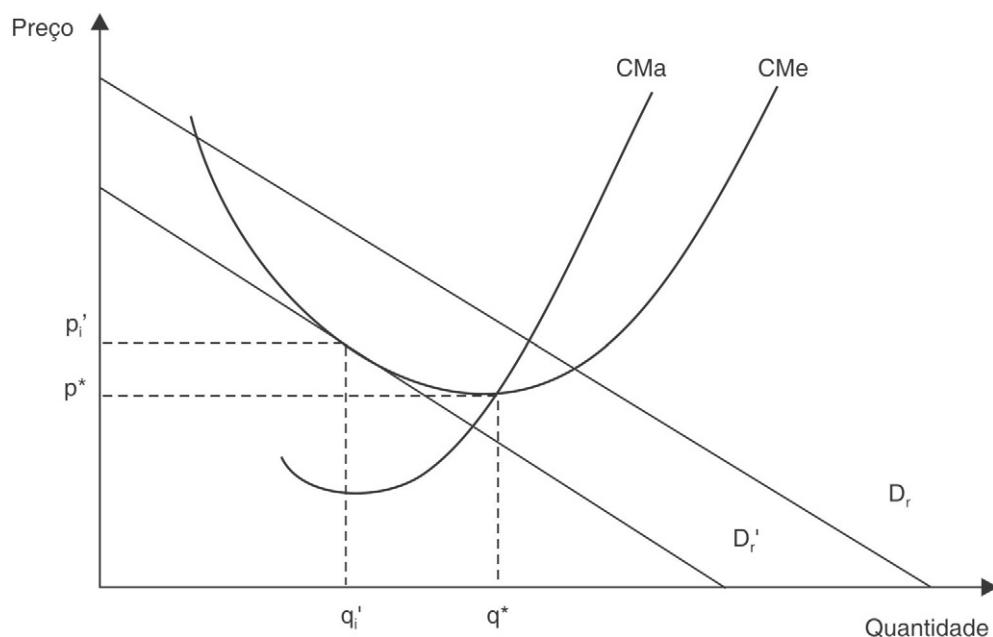


Gráfico 6.1b Modelo de Chamberlin após a entrada de empresas.

O modelo Chamberlin sofreu inúmeras críticas. Em função das hipóteses utilizadas, o tratamento da diferenciação de produtos não incorpora elementos fundamentais do processo. A hipótese de custos e demanda homogêneos entre as empresas é especialmente irrealista e restritiva. Se as empresas se deparam com demandas negativamente inclinadas (elasticidade finita), os produtos oferecidos têm de apresentar diferenças, mas, segundo o modelo, essas diferenças não têm impactos sobre os custos das empresas. Empiricamente, o que se observa são empresas que oferecem produtos diferenciados que têm preços e custos distintos entre si.

A hipótese de entrada livre na indústria também é incoerente com a possibilidade de diferenciação de produtos. A diferenciação é um fator que gera barreiras à entrada, já que as empresas entrantes têm de realizar gastos substanciais com esforços de venda para reverter a preferência do consumidor por produtos de empresas já estabelecidas, como será visto no Capítulo 7.

Apesar dessas limitações, a análise de Chamberlin teve grande importância para a evolução da ciência econômica por levar em conta de forma sistemática fatores desprezados até então, a diferenciação de produtos e os esforços de venda.

6.4 Modelos Locacionais

Já vimos que as empresas que oferecem produtos diferenciados competem entre si. No entanto, é de se esperar que quanto mais substituíveis os produtos mais intensa será a competição. Por exemplo, no mercado de refrigerantes, o preço da Coca-Cola tem maiores influências sobre a quantidade demandada de Pepsi do que sobre a demanda de Sprite *diet*. Os modelos locacionais têm o objetivo de enfatizar esse ponto na análise da diferenciação, avaliando os incentivos para as empresas produzirem mercadorias muito ou pouco diferenciadas.

Os modelos locacionais utilizam a analogia entre as características de produtos e a localização de lojas. Assim como os consumidores preferem adquirir produtos em lojas próximas às suas residências – o que só não ocorre se o custo de transporte para outra loja for inferior à diferença de preços, estes desejam também adquirir os produtos mais adequados às suas preferências, e somente não o fazem se o custo de adequação a um produto similar for inferior à diferença de preços.

Dois modelos são expostos: o modelo da cidade linear, conforme desenvolvido originalmente por H. Hotteling em 1929, e o modelo da cidade circular, elaborado por S. Salop em 1979.

6.4.1 MODELO DA CIDADE LINEAR

Nesse modelo, é considerada uma cidade com uma única avenida, onde os consumidores estão uniformemente distribuídos. Duas empresas que oferecem o mesmo produto decidem onde devem se localizar.

Apesar de, originalmente, o modelo de Hotteling ser voltado apenas para a distribuição espacial de empresas, desenvolvimentos teóricos posteriores demonstraram sua utilidade para analisar a diferenciação de um atributo dos produtos. Uma ilustração consiste em duas empresas que escolhem produzir balas mais ou menos doces. Nesse caso, cada produtor escolhe uma posição em um intervalo contínuo de possibilidades que corresponde à “doçura” da bala produzida, de forma que em um extremo do intervalo está a bala mais doce possível e no outro extremo, a bala menos doce.

O seu conteúdo analítico pode ser expresso de forma intuitiva. O Gráfico 6.2a ilustra, simplificadamente, o modelo de Hotteling. A empresa 1 se localiza a a quilômetros do início da cidade e a empresa 2 a d quilômetros do final. Considerando preços iguais entre as empresas, os consumidores devem adquirir produtos das empresas que estão mais próximas às suas residências. Assim, a empresa 1 captura todos os consumidores localizados nos intervalos a e b e a empresa 2 aqueles localizados nos intervalos c e d . O consumidor situado na posição i é indiferente entre adquirir na empresa 1 ou 2, pois a distância que tem de percorrer para chegar à empresa 1 (b) é mesma para a empresa 2 (c).

Na situação inicial em que as empresas se encontram equidistantes dos extremos, a empresa 1 tem incentivos para se encaminhar em direção ao centro da cidade, aumentando sua participação de mercado já que um número maior de consumidores terá menores custos de transporte para chegar à empresa 1 do que para chegar à empresa 2. Ou seja, o consumidor indiferente ficará mais à direita e o intervalo $a + b$ aumenta (Gráfico 6.2b).

No entanto, os mesmos incentivos também são válidos para a empresa 2, que também se encaminha para o meio procurando conquistar maior participação de mercado (este comportamento se assemelha ao das empresas que cortam preços no modelo

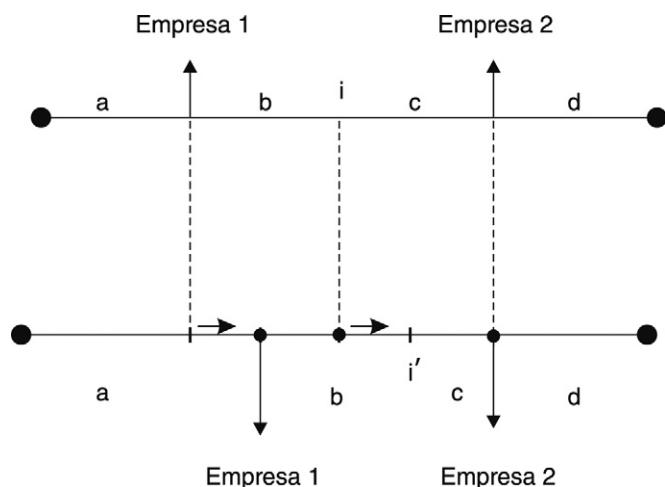


Gráfico 6.2a Modelo da Cidade Linear – Situação Inicial (Empresas Equidistantes).

Gráfico 6.2b Modelo da Cidade Linear – Movimento da Empresa 1.

de oligopólio de Bertrand). Após movimentos sucessivos, o equilíbrio ocorre quando ambas as empresas se encontram no meio da cidade.

Aplicando o raciocínio para a diferenciação de atributos dos produtos, o resultado é que as empresas têm incentivos para oferecer produtos semelhantes e não para diferenciar os produtos. No caso das balas, ambas as empresas ofereceriam balas com “doçura” moderada, nem muito, nem pouco doce. Esse resultado é conhecido como princípio da diferenciação mínima.

Podemos formalizar o modelo de Hotteling como se segue:

Sejam t o custo de transporte por quilômetro percorrido e u o comprimento da cidade em quilômetros ($u = a + b + c + d$), considerando custos de produção nulos e u consumidores uniformemente distribuídos, onde cada consumidor adquire uma unidade do produto.

Para o consumidor localizado no ponto i (indiferente entre adquirir da empresa 1 ou 2), vale a seguinte igualdade:

$$p_1 + tb = p_2 + tc \quad (1)$$

$$b - c = \frac{p_2 - p_1}{t} \quad (1b)$$

mas como $u = a + b + c + d$

$$b + c = u - a - d \quad (2)$$

somando (1b) e (2), temos:

$$\begin{aligned} 2b &= \frac{p_2 - p_1}{t} + u - a - d \\ b &= \frac{1}{2} \left(\frac{p_2 - p_1}{t} + u - a - d \right) \end{aligned} \quad (3)$$

subtraindo (1b) em (2), temos:

$$\begin{aligned} 2c &= \frac{p_1 - p_2}{t} + u - a - d \\ c &= \frac{1}{2} \left(\frac{p_1 - p_2}{t} + u - a - d \right). \end{aligned} \quad (4)$$

Em função das hipóteses, a quantidade vendida por cada empresa é:

$$q_1 = a + b \quad (5)$$

$$q_2 = c + d \quad (6)$$

Substituindo (3) em (5) e (4) em (6), temos:

$$\begin{aligned} q_1 &= a + \frac{1}{2} \left(\frac{p_2 - p_1}{t} + u - a - d \right) \\ q_1 &= \frac{1}{2} \left(\frac{p_2 - p_1}{t} + u + a - d \right) \end{aligned} \quad (7)$$

$$\begin{aligned} q_2 &= d + \frac{1}{2} \left(\frac{p_1 - p_2}{t} + u - a - d \right) \\ q_2 &= \frac{1}{2} \left(\frac{p_1 - p_2}{t} + u - a + d \right). \end{aligned} \quad (8)$$

Assim, pode-se definir os lucros de cada empresa:

$$\begin{aligned} \pi_1 &= p_1 q_1 = p_1 \frac{1}{2} \left(\frac{p_2 - p_1}{t} + u + a - d \right) \\ \pi_2 &= p_2 q_2 = p_2 \frac{1}{2} \left(\frac{p_1 - p_2}{t} + u - a + d \right). \end{aligned}$$

Verificando os preços nos quais as empresas maximizam lucros.

Condição de primeira ordem (CPO):

$$\frac{d\pi_1}{dp_1} = \frac{1}{2}(u + a - d) - \frac{p_1}{t} + \frac{p_2}{2t} = 0 \quad (9)$$

$$\frac{d\pi_2}{dp_2} = \frac{1}{2}(u - a + d) - \frac{p_2}{t} + \frac{p_1}{2t} = 0. \quad (10)$$

Condições de segunda ordem (CSO):

$$\frac{d^2\pi_1}{dp_1^2} = \frac{d^2\pi_2}{dp_2^2} = -\frac{1}{1} < 0.$$

A partir de (9) e (10) temos, respectivamente:

$$u + a - d + \frac{p_2 - 2p_1}{t} = 0 \quad (11)$$

$$u - a + d + \frac{p_1 - 2p_2}{t} = 0. \quad (12)$$

A partir de (11), temos:

$$p_2 = 2p_1 - t(u + a - d). \quad (13)$$

Substituindo (13) em (12) e após algum desenvolvimento algébrico, temos os preços de equilíbrio.

$$p_1^* = t \left(u + \frac{a-d}{3} \right) \quad (14)$$

por analogia:

$$p_2^* = t \left(u + \frac{d-a}{3} \right). \quad (15)$$

Substituindo (14) e (15) em (7) e (8), temos:

$$\begin{aligned} q_1^* &= \frac{1}{2} \left\{ u + a - d + \frac{1}{t} \left[t \left(u + \frac{d-a}{3} \right) - t \left(u + \frac{a-d}{3} \right) \right] \right\} \\ q_1^* &= \frac{1}{2} \left(u + \frac{a-d}{3} \right) \\ q_2^* &= \frac{1}{2} \left(u + \frac{d-a}{3} \right). \end{aligned}$$

Os lucros de cada empresa no ponto de equilíbrio serão:

$$\begin{aligned} \pi_1^* &= t \left(u + \frac{a-d}{3} \right) \cdot \frac{1}{2} \left(u + \frac{a-d}{3} \right) = \frac{t}{2} \left(u + \frac{a-d}{3} \right)^2 \\ \pi_2^* &= \frac{t}{2} \left(u + \frac{d-a}{3} \right)^2. \end{aligned}$$

Verificando a relação entre a localização das empresas e seus lucros:

$$\frac{d\pi_1^*}{da} = \frac{t}{3} \left(u + \frac{a-d}{3} \right) > 0, \text{ pois } u > \frac{d-a}{3}$$

$$\frac{d\pi_2^*}{dd} = \frac{t}{3} \left(u + \frac{d-a}{3} \right) > 0, \text{ pois } u > \frac{a-d}{3}.$$

Ou seja, o lucro da empresa 1 cresce conforme a aumenta, ocorrendo o mesmo para a empresa 2 em relação a d . Dessa forma, ambas as empresas têm incentivos para se afastarem dos pontos extremos, e o equilíbrio ocorre quando as duas empresas se localizam no ponto central da cidade. Vale, então, o princípio da diferenciação mínima.

Como esse resultado não é intuitivo nem comprovado empiricamente, já que a ocorrência de produtos diferenciados é característica fundamental de várias indústrias, outros autores desenvolveram modelos que contrariam os resultados de Hotteling. D'Aspremont, Gabszewicz e Thisse (1979) desenvolveram um modelo em que as empresas, além de definir a localização, também escolhem seus preços e apontam para um problema de descontinuidade da função de lucros, ocorrendo inexistência de equilíbrio se as empresas se encontram próximas ao ponto central da cidade. Para resolver esse problema de indefinição, os autores consideram custos de transporte quadráticos. Ou seja, os consumidores são mais sensíveis à distância que têm de percorrer para adquirir o produto, o que significa que valorizam mais o produto da loja mais próxima. Nesse caso, o equilíbrio ocorre com as empresas se localizando nos extremos da cidade. Dessa forma, segundo esses autores, o princípio da mínima diferenciação não é válido, e sim seu oposto, o da máxima diferenciação.

6.4.2 MODELO DA CIDADE CIRCULAR

O modelo de Hotteling analisa os incentivos existentes para que duas empresas estabelecidas diferenciem seus produtos, desprezando o efeito da entrada de novas empresas. O modelo de Salop, além de analisar a localização das empresas, também enfatiza os efeitos da entrada de empresas na indústria. O objetivo do modelo é verificar se o número de empresas que atuam no mercado, isto é, se o número de variedades geradas, é socialmente ótimo.

A consideração de um espaço circular, com consumidores distribuídos uniformemente por seu perímetro e empresas simétricas, é conveniente ao objetivo do modelo, já que não há vantagens iniciais de localização.³ Uma interpretação prática seria a decisão da localização de quiosques ao redor de uma lagoa.

O modelo exposto neste item⁴ considera consumidores que adquirem apenas uma quantidade do produto, n empresas estabelecidas em um círculo de perímetro 1, custo marginal c , custo de transporte t , custo fixo de implantação das empresas f . É adotado como hipótese que as empresas se localizam equidistantesumas das outras (ou seja, implicitamente considera-se o princípio da máxima diferenciação).

A Figura 6.1 descreve o modelo. No círculo, estão situadas n empresas a uma distância de $1/n$ umas das outras.

Suponha um consumidor indiferente entre comprar na empresa 1 ou 2, localizado a uma distância a da empresa 1. Para este consumidor vale a igualdade:

$$p_1 + ta = p_2 + t(1/n - a)$$

$$a = \frac{p_2 + t/n - p_1}{2t}$$

A empresa 1 tem a demanda definida por $2a$ (consumidores localizados mais próximos à empresa 1 do que os consumidores indiferentes entre a empresa 1 e a 2 e entre a empresa 1 e a n). Em função da hipótese de simetria, podemos considerar os preços das demais empresas iguais a p .

$$D_1(p_1, p) = 2a = \frac{p + t/n - p_1}{t}$$

$$\pi_1 = (p_1 - c) \frac{p + t/n - p_1}{t} - f$$

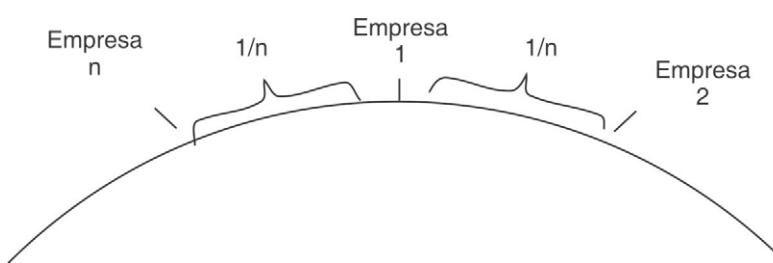


Figura 6.1 Cidade Circular.

Podemos calcular o preço que maximiza os lucros da empresa 1:

$$(CPO) \frac{d\pi_1}{dp_1} = -\frac{1}{t}(p_1^* - c) + \frac{p + t/n - p_1^*}{t} = 0$$

$$(CSO) \frac{d^2\pi_1}{dp_1^2} = -\frac{2}{t} < 0$$

No equilíbrio simétrico $p_1^* = p$, assim:

$$p = c + \frac{t}{n} \quad (1)$$

Como a entrada é livre, o número de empresas no mercado é determinado pela condição de lucro nulo.

$$(p - c) \frac{p + t/n - p}{t} - f = 0 \quad (2)$$

Substituindo (1) em (2):

$$\frac{t}{n^2} - f = 0$$

O número de empresas e o preço no ponto de equilíbrio serão:

$$n = \sqrt{t/f}$$

$$p = c + \frac{t}{\sqrt{t/f}} = c + \sqrt{tf}$$

Para verificar a eficiência do resultado deve-se compará-lo ao número de empresas que maximiza o bem-estar social. Nesse caso, o número de empresas é aquele que minimiza a soma do total de custos fixos (nf) com o total gasto com transporte.

O total gasto com transporte é igual ao custo de transporte (t) multiplicado pela distância percorrida por cada consumidor. A distância máxima percorrida por um consumidor é de $1/2n$ – pois, se a distância for maior, é preferível se dirigir para outra empresa – e a mínima é 0. Neste intervalo a distância média percorrida é: $\int_0^{1/2n} x dx$ (a densidade ao longo do círculo é igual a 1). Como existem $2n$ desses intervalos no círculo, o total gasto com transporte é igual a $t \left(2n \int_0^{1/2n} x dx \right)$.

Calculando n que minimiza os custos totais:

$$(CPO) \frac{d}{dn} \left[nf + t \left(2n \int_0^{1/2n} x dx \right) \right] = \frac{d}{dn} \left(nf + \frac{t}{4n} \right) = 0$$

$$f - \frac{t}{4n^2} = 0$$

$$(CSO) \frac{d}{dn} \left(f - \frac{t}{4n^2} \right) = \frac{t}{2n^3} > 0$$

Assim, o número de empresas que proporciona o maior nível de bem-estar:

$$n^* = \frac{1}{2} \sqrt{t/f}$$

Dessa forma, o número de empresas que corresponde ao ótimo social é metade do número que surge com a livre atuação do mercado. Portanto, como no modelo de Chamberlin, se o mercado opera livremente, é gerado um número de diversidades de produtos maior que o desejável.

6.5 Abordagens Alternativas

Os modelos locacionais, tais quais o modelo de Chamberlin, admitem hipóteses muito restritivas, constituindo-se, sobretudo, em um esforço de formalização. Nestes, não são consideradas as implicações do processo de diferenciação. Como não há uma interação entre a diferenciação e outros aspectos do desempenho da indústria, a diferenciação não é incorporada à formação de estratégias empresariais e à dinâmica da indústria.

Joe Bain, em 1956, considera a vantagem de diferenciação de produto uma das fontes de barreiras à entrada na indústria. Dessa forma, a entrada de novas empresas não é livre e as empresas podem obter lucros supranormais sem atrair competidores, como será visto com mais detalhes no Capítulo 7.

A abordagem evolucionista, apresentada na Parte III, incorpora o processo de diferenciação à dinâmica das indústrias. A diferenciação surge como resultado de uma inovação de produto que propicia poder de monopólio para as empresas inovadoras, dando lugar a lucros extraordinários. O surgimento de novos produtos ocorre segundo uma combinação de fatores estocásticos e determinísticos, que atuam no processo de geração de variedades (busca) e na sua seleção pelo mercado. Estes últimos são determinados pelas atitudes deliberadas das empresas e pela configuração do ambiente de seleção. P. Saviotti, em 1996, analisou o processo de geração de variedades na indústria, observando as regularidades presentes no processo.

As indústrias nas quais a diferenciação de produto se constitui em um importante instrumento de concorrência são classificadas como oligopólio diferenciado. Nessas indústrias, os gastos com comercialização e publicidade são elevados, assim como os esforços inovativos. O contínuo surgimento de novos produtos resulta em uma tendência à instabilidade estrutural, que só não é mais acentuada porque os gastos com vendas e com o processo inovativo são caracterizados pela cumulatividade. Dessa forma, frequentemente são as mesmas empresas que introduzem novos produtos.⁵

A competição pela diferenciação acaba atuando de três formas: aumenta as vendas de uma empresa em particular; atua sobre a demanda de mercado como um todo, incorporando novos consumidores; e garante a sobrevivência da empresa no mercado. Assim, a diferenciação é, ao mesmo tempo, uma arma agressiva, na medida em que a empresa está buscando garantir uma demanda maior para seus produtos, e defensiva, porque é uma forma de garantir sua posição no mercado em que atua.

A incorporação da estratégia da diferenciação do produto pelas empresas e de seus efeitos sobre a dinâmica da indústria ainda deve avançar do ponto de vista teórico. Esta representa uma das diretrizes da agenda heterodoxa.

6.6 Conclusão

Buscamos, neste capítulo, apresentar como a diferenciação de produto é analisada por diferentes correntes da Economia Industrial. A diferenciação de produtos envolve aspectos, como diferenciais de qualidade, que dificultam sua formalização. Os modelos apresentados recorrem a simplificações que não permitem aplicações imediatas. Os modelos se preocupam com questões específicas. Os modelos de Chamberlin focam o impacto da possibilidade de diferenciação sobre o bem-estar social e o de Hotteling os estímulos para a diferenciação.

Na corrente heterodoxa, as implicações do processo de diferenciação são enfatizadas, principalmente sobre a dinâmica da indústria. No entanto, esta proposta dificulta sensivelmente sua formalização. O instrumental existente se restringe, principalmente, a tipologias que associam características das estruturas de mercado e estratégias de empresas.

Notas

1. No modelo de concorrência perfeita, a curva de demanda residual é horizontal.
2. O modelo exposto é uma simplificação do modelo original. A análise de Chamberlin também incorporou outros elementos como o efeito da miopia decisória das empresas quando concorrem via preços. No modelo, as empresas cortam preços para aumentar a quantidade vendida considerando que as demais empresas manterão preços inalterados, mas como as outras empresas também acompanham o corte de preços, o efeito em aumento de vendas é inferior ao esperado. A consideração deste efeito de miopia torna o modelo mais complexo. Mas como o resultado não é alterado, optamos por centrar a análise na diferenciação de produtos.
3. No modelo linear, se três empresas se distribuem na reta, por exemplo, a que se situa na posição central tem desvantagens, pois tem de competir com as duas outras empresas, ao passo que a firma mais próxima à extremidade esquerda detém um intervalo de consumidores cativos (entre a firma e a extremidade), só tendo que competir pelos consumidores localizados a sua direita (o mesmo vale para a firma à direita).
4. A exposição segue, em grande medida, a de Tirole (1988).
5. Possas, M. L. (1985), p. 176.

Bibliografia

- Bain, J. (1956) *Barriers to New Competition*. Cambridge, MA: Harvard University Press.
Chamberlin, E. (1933) *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

- D'Aspremond, C., Gabszewicz, J. e Thisse, J. (1979) "On Hotteling's Stability in Competition". *Econometrica*, 17, pp. 1045-1151.
- Guimarães, E. A (1982) *Acumulação e Crescimento da Empresa: um Estudo de Organização Industrial*. Rio de Janeiro: Zahar.
- Hotelling, H. (1929) Stability in Competition. *Economic Journal*, 39, pp. 41-57.
- Possas, M. L. (1985) *Estruturas de Mercado em Oligopólio*. São Paulo: Hucitec.
- Possas, M. S. (1993) *Concorrência e Competitividade Notas sobre a Estratégia e Dinâmica Seletiva na Economia Capitalista*. Campinas, IE/Unicamp, Tese de Doutorado.
- Salop, S. (1979) Monopolistic Competition with Outside Goods. *Bell Journal of Economics*, 10, pp. 141-156.
- Saviotti, P. P. (1996) *Technological Evolution, Variety and the Economy*. Edward Elgar Publishing, Cheltenham, Reino Unido, 223 pp.
- Shapiro, C. e Varian, H. (1999) *A Economia da Informação – como os Princípios Econômicos se Aplicam à Era da Internet*. Tradução de Ricardo Inojosa do original "Information Rules". Rio de Janeiro: Campus.
- Tirole, J. (1988) *The Theory of Industrial Organization*. Cambridge, MA: The MIT Press.
- Waterson, M. (1994) Models of Product Differentiation. In: Cable, J. (Org.) (1994), *Current Issues in Industrial Economics*. Nova York: St. Martin's Press. Cap. 6. pp. 105-133.



Barreiras Estruturais à Entrada

David Kupfer

7.1 Introdução

Até aqui nossa discussão sobre o funcionamento dos mercados enfatizou a concorrência que ocorre no interior de uma determinada indústria. O número e o tamanho relativo das diversas empresas que formam a indústria e a natureza das funções de custos associadas às atividades produtivas realizadas foram as variáveis utilizadas para explicar a lucratividade de um setor industrial. Indústrias com grau elevado de concentração seriam as mais lucrativas. Inversamente, estruturas industriais mais atomizadas seriam as menos lucrativas. O monopólio – a estrutura industrial com apenas uma empresa e, portanto, com máximo grau de concentração – corresponderia ao limite superior de lucratividade. No outro extremo, estruturas em concorrência perfeita – as mais atomizadas – definiriam o limite inferior de lucratividade nula, isto é, a condição em que o preço equivale ao custo marginal de produção.

Estudos empíricos, porém, revelam que raramente grau de concentração e economias de escala são suficientes para explicar a estrutura industrial. O tamanho mínimo econômico (ou EME – escala mínima eficiente – ver Capítulo 3) quase nunca supera a casa de 10% do tamanho do mercado assim como o crescimento dos custos médios de produção de fábricas com escalas subótimas geralmente é de pequena monta. Por que, então, a maioria das indústrias é concentrada? Mais importante: diversas tentativas de comprovação empírica da hipótese de que a lucratividade de uma indústria é positivamente correlacionada com o grau de concentração também levaram a resultados pouco conclusivos.

A principal deficiência das análises estruturais de indústrias baseadas somente em curvas de custos e participação das empresas no mercado está no fato de que ela é desenvolvida sem levar em conta, ao menos diretamente, a existência de outras empresas que atuam em outros setores da economia. Em uma economia formada por múltiplas indústrias, não parece razoável imaginarmos que essas sejam independentes entre si.

A partir do trabalho pioneiro de Joe S. Bain nas décadas de 1940 e 1950, ganhou corpo nas teorias de Economia Industrial a ideia de que o principal fator na determinação dos preços e da lucratividade em uma indústria está relacionado à facilidade ou à dificuldade que as empresas estabelecidas encontram para impedir a entrada de novas empresas, isto é, a existência ou não de barreiras à entrada na indústria.

Neste capítulo iremos nos concentrar no enfoque estrutural das barreiras à entrada, também conhecidas como barreiras à entrada estáticas ou exógenas.¹ Como veremos, são barreiras à entrada que decorrem exclusivamente da relação preço-custo médio de longo prazo predominante na indústria. Outros enfoques, baseados em análises dos comportamentos ativos ou reativos das empresas no que se refere às escolhas de preços e quantidades, as chamadas barreiras à entrada estratégicas, endógenas ou também prevenção de entrada, serão abordados no Capítulo 12.

O capítulo prossegue com a discussão dos conceitos de concorrência real e potencial, da qual surge toda a construção teórica da noção de barreira à entrada. Na terceira seção apresentamos as principais definições de barreira à entrada

encontradas na literatura. O modelo conceitual de preço-limite, que explicita o mecanismo econômico básico da criação de barreiras à entrada, e o exame destas na prática, por meio da análise de suas diversas fontes empíricas, são os temas da quarta e da quinta seções. Na sexta seção, o foco da análise é transferido para a noção de barreiras à saída e a apresentação dos princípios centrais da teoria da contestabilidade. Na conclusão, discutimos as limitações do enfoque estrutural das barreiras à entrada e apontamos os principais tópicos que ainda permanecem abertos na agenda de pesquisa sobre esse tema.

7.2 Concorrência Real e Potencial

Os modelos tradicionais de análise microeconômica limitam-se a examinar o conceito “marshalliano” de concorrência limitada a cada mercado, isto é, a concorrência em função do número e do tamanho relativo das diversas empresas que formam cada indústria. A essa noção de concorrência – chamada concorrência real – opõe-se a noção de concorrência potencial. Concorrência potencial relaciona-se à competição por lucros entre empresas já estabelecidas em uma determinada indústria e novas empresas interessadas em iniciar operação nessa mesma indústria – as empresas entrantes (ou empresas potenciais).

A ênfase na concorrência potencial é, de fato, típica do pensamento econômico clássico. Para essa corrente, se uma indústria apresenta lucros elevados, é de se esperar que novas empresas venham a se estabelecer nessa indústria buscando compartilhar esses lucros extraordinários. Simetricamente, se uma indústria apresenta desempenho deficitário, algumas empresas desejarão encerrar as atividades e transferir-se para indústrias mais atraentes. Nos setores superavitários, aumento da oferta, consequente à adição de capacidade, reduz preços e contrai lucros, ao passo que, nos setores deficitários, a redução da oferta, consequente à saída de empresas, contrai a oferta e eleva os lucros. A migração intersetorial dos capitais cessaria apenas quando as taxas de lucros se igualassem em todas as indústrias.

Como resultado, a concorrência é vista como um processo dinâmico, caracterizado por livre entrada e saída de capitais da indústria e, considerando a economia como um todo, pela tendência à igualação da taxa de lucro. A mobilidade interindustrial dos capitais é a força econômica determinante da dinâmica dos mercados e, portanto, são as interações entre empresas no processo econômico como um todo que devem ser a base para a teoria geral de preços e da produção.

Nessa visão, uma indústria somente poderia apresentar lucratividade superior à média por um certo período de tempo – o tempo necessário para que uma decisão de construir capacidade se materialize. Consequentemente, se uma indústria apresenta lucros extraordinários permanentes, alguma restrição à mobilidade do capital existe. Dizemos então que existem barreiras à entrada nessa indústria.

A noção de concorrência potencial e a existência de barreiras à entrada trazem importantes implicações sobre as escolhas de preços e quantidades realizadas pelas empresas que, de certo modo, sempre constituíram um desafio para a modelagem do funcionamento dos mercados.

Antes de prosseguirmos, convém definirmos os elementos básicos envolvidos em um problema da entrada.

1. *Empresas estabelecidas.* São as empresas que já atuam na indústria considerada. Comumente, as análises de entrada supõem que todas as empresas estabelecidas se coordenam com o objetivo de impedir entradas. Uma interpretação possível para essa hipótese é imaginarmos que o receio da concorrência potencial anula a concorrência real.
2. *Empresas entrantes.* Também chamadas empresas potenciais, correspondem a qualquer capital interessado em atuar na indústria analisada. A rigor, o número de empresas entrantes em uma indústria é indefinido. Para resolver esse problema, é usual imaginarmos que elas formam uma fila organizada de acordo com a capacidade de competir na indústria. A primeira da fila, isto é, aquela que reúne os melhores requisitos competitivos, é a primeira empresa entrante.
3. *Incentivo à entrada.* Associado à possibilidade de uma nova empresa vir a se estabelecer na indústria e obter lucros extraordinários por um certo período de tempo. Nas análises estáticas de entrada é comum considerar-se que somente haja incentivo à entrada se esses lucros puderem ser auferidos imediatamente após a entrada.
4. *Entrada.* Uma entrada corresponde a uma adição líquida de capacidade produtiva na indústria por uma nova empresa. Por essa definição estão excluídas a expansão de capacidade de uma empresa já estabelecida, pois não significa um novo agente no processo competitivo, e a entrada via fusão ou aquisição de uma empresa já estabelecida, pois não significa adição de capacidade.
5. *Saída.* Uma saída significa que uma empresa encerrou suas atividades, isto é, que um certo montante de capacidade produtiva foi permanentemente eliminado da indústria. Se uma empresa abandonar uma indústria vendendo seus ativos produtivos a um terceiro, não haverá saída; terá ocorrido tão somente uma transferência do controle de um negócio.

7.3 Barreira à Entrada: Definições

São muitos os enfoques sobre barreiras à entrada na literatura de Economia Industrial. Todos têm em comum a ênfase conferida ao longo prazo e a concorrência potencial como bases teóricas para o conceito. Qualquer fator que impeça a livre mobilidade do capital para uma indústria no longo prazo e, consequentemente, torne possível a existência de lucros supranormais permanentes nessa indústria, constitui barreira à entrada.

Entretanto, quando se busca uma definição mais operacional, a convergência já não é tão nítida. É possível, no entanto, reunir as definições mais amplamente utilizadas sobre barreiras à entrada em quatro grupos.

No primeiro grupo está a definição atribuída a Joe S. Bain, pioneiro e principal formulador teórico dessa corrente. Barreira à entrada corresponde a qualquer condição estrutural que permita que empresas já estabelecidas em uma indústria possam praticar preços superiores ao competitivo sem atrair novos capitais. Em termos práticos, isso significa que é possível a existência de lucros extraordinários no longo prazo porque as empresas entrantes não conseguem auferir após a entrada os mesmos lucros que as empresas estabelecidas obtêm pré-entrada.

No segundo grupo está a definição de J. Stigler: Existe barreira à entrada em uma indústria se há custos incorridos pelas empresas entrantes que não foram desembolsados pelas empresas estabelecidas quando iniciaram a operação. Essa assimetria de custos entre empresas estabelecidas e empresas entrantes após a entrada impossibilita essas últimas de obterem a mesma lucratividade que as primeiras.

No terceiro grupo predominam as visões como a de R. Gilbert, na qual somente há barreiras à entrada se é possível configurar vantagens competitivas atribuíveis exclusivamente à existência da empresa. Nessa terceira definição, somente há barreira à entrada quando há um diferencial econômico entre empresas estabelecidas e entrantes simplesmente porque as primeiras já existem e as outras ainda não. Esse “prêmio pela existência” é, necessariamente, a tradução econômica de algum tipo de “vantagem da primeira empresa a se mover” (*first-mover advantages*). Nessa visão, uma teoria de barreiras à entrada não pode ser construída sem uma teoria do comportamento oligopolista e sem a análise das barreiras à saída existentes na indústria. Esse tipo de tratamento é a motivação teórica básica das teorias de prevenção estratégica de entrada que, como já mencionamos, serão discutidas no Capítulo 12.

Há ainda um último grupo de definições de barreiras à entrada que reúne os autores que enfatizam os aspectos normativos da questão da entrada, dentre os quais C. Von Weizsäcker é o principal representante. Nesse caso, a existência de diferenciais de custos entre empresas estabelecidas e entrantes não é condição suficiente para assegurar a presença de barreiras à entrada. É necessário, também, que impliquem distorções na alocação de recursos do ponto de vista social. Na Seção 7.5.2 faremos alguns comentários a esse respeito.

7.4 O Modelo Conceitual do Preço Limite

Vamos considerar uma indústria em equilíbrio temporário, na qual as empresas estabelecidas atuam em conjunto visando a prevenir entradas. As empresas podem produzir tanto bens homogêneos quanto diferenciados, conforme apresentado no Capítulo 6, mas sempre utilizam tecnologias que apresentam custos médios de longo prazo em forma de L, isto é, os custos médios são decrescentes com o aumento da escala até atingirem o nível equivalente à escala mínima eficiente, quando então se tornam constantes (ver Capítulo 3 para uma análise mais detalhada desse tipo de estrutura de custos).

Vamos considerar, ainda, o longo prazo como uma sequência de dois curtos prazos: o primeiro correspondente ao período pré-entrada e o segundo ao período pós-entrada. Finalmente, vamos supor que a empresa entrante (a mais capacitada a entrar dentre todas as empresas potenciais) avalia que haja incentivo à entrada somente se for possível a obtenção de lucro econômico positivo imediatamente após a entrada, isto é, o segundo dos dois períodos da análise é de pequena duração. Essa última hipótese é, de fato, a premissa fundamental das análises estáticas de entrada. Pode ser empiricamente interpretada como uma consequência direta da empresa entrante ser nova e, portanto, não ter suporte financeiro para operar com prejuízo mesmo que por um período curto de tempo. É importante registrar que essa premissa torna-se particularmente inadequada quando aplicada a empresas entrantes ligadas a capitais já constituídos, como no caso da entrada corresponder a uma diversificação de uma empresa já existente em outra indústria.

Para as empresas estabelecidas, uma possibilidade sempre à mão para prevenir entradas é fixar o preço no nível competitivo. Nesse caso, não haverá entrada simplesmente devido a ausências de incentivos. Essa escolha, porém, é também pouco atrativa para elas próprias, já que logicamente também não irão obter qualquer lucro, nos dois períodos. Melhor seria simplesmente fixar o preço no nível da maximização de lucros de curto prazo (do primeiro período). Nesse caso, será possível obter lucros positivos (e máximos) no primeiro período, após o que ocorrerão entradas que levarão o preço ao nível competitivo no segundo período.

Há, no entanto, uma opção intermediária. Se as empresas estabelecidas têm alguma vantagem competitiva em relação à empresa entrante, existe uma faixa de preços tal que é possível a elas obterem lucros positivos – mesmo que não os máximos possíveis no primeiro período – ao mesmo tempo em que nenhuma entrada seja incentivada. O valor superior dessa faixa é conhecido como preço limite. A adoção do preço limite torna possível às empresas estabelecidas auferirem um certo nível de lucros de forma permanente (isto é, no primeiro e no segundo períodos).

A questão é qual o preço limite que será escolhido pelas empresas existentes. Para responder a essa pergunta, Bain introduziu o conceito de condição de entrada, uma margem sobre os custos médios de longo prazo que as empresas estabelecidas podem incluir no preço sem atrair entradas.

Algebraicamente,

$$E = \frac{P_L - P_C}{P_C} \text{ ou } P_L = P_C(1 + E)$$

onde E é a condição de entrada, P_L o preço limite e P_C o preço competitivo no longo prazo (equivalente ao custo médio mínimo de longo prazo).

Em relação à condição de entrada E podem prevalecer quatro situações distintas:

1. *Entrada fácil.* As empresas estabelecidas não têm vantagens de custos em relação à empresa entrante e não podem sustentar lucros extraordinários. Não há barreira à entrada e prevalece o preço competitivo.
2. *Entrada ineficazmente impedida.* As empresas estabelecidas têm pouca vantagem competitiva e por isso preferem praticar o preço de maximização de curto prazo. Com isso irão obter os lucros mais altos possíveis embora apenas no primeiro período, porque ocorrerão entradas até que o preço retorne ao nível competitivo no segundo período. Outra razão para esse comportamento decorre da existência de fatores que impliquem um longo período de maturação dos investimentos de sorte que o tempo necessário para que a entrada se materialize seja longo.
3. *Entrada eficazmente impedida.* As empresas estabelecidas têm vantagem competitiva significativa em relação à empresa entrante, razão pela qual preferem praticar o preço limite e barrar entradas em vez de adotarem o preço de maximização dos lucros no primeiro período e permitir entradas que anulem os lucros no segundo período. A condição para que essa opção seja preferida é que o lucro acumulado nos dois períodos proporcionado pela adoção do preço limite seja superior ao lucro máximo possível no primeiro período (e nulos no segundo).
4. *Entrada bloqueada.* É a situação na qual as vantagens competitivas das empresas estabelecidas são tão grandes que mesmo o preço de maximização dos lucros no primeiro período é inferior ao preço limite. Nesse caso, o preço de maximização do primeiro período está dentro da faixa de preços que não incentiva entradas e, portanto, as empresas existentes irão manter esses lucros permanentemente.

7.5 Barreiras Estruturais à Entrada na Prática

Nesse ponto, para poder tornar operacional o mecanismo conceitual de definição do preço limite que abordamos na seção anterior é necessário que as características estruturais das indústrias sejam devidamente especificadas. Por essa razão, para prosseguirmos a análise é necessário detalhar os elementos presentes na estrutura da indústria que podem constituir fontes de barreira à entrada. São eles:

1. Existência de vantagens absolutas de custos a favor das empresas estabelecidas.
2. Existência de preferências dos consumidores pelos produtos das empresas estabelecidas.
3. Existência de estruturas de custos com significativas economias de escala.
4. Existência de elevados requerimentos de capital inicial.

7.5.1 VANTAGENS ABSOLUTAS DE CUSTOS

Quando o custo médio de longo prazo das empresas entrantes é superior ao das empresas estabelecidas em qualquer nível de produção de um bem homogêneo dizemos que essas últimas detêm vantagens absolutas de custos.

A Figura 7.1 mostra como a existência de vantagens absolutas de custos pode originar barreiras à entrada na indústria – as barreiras de custos (*cost barriers*). Na figura, A é a empresa estabelecida e B a empresa entrante. As curvas representam os custos médios de longo prazo (CMeLP) de ambas quando há uma vantagem absoluta em favor de A. DD' é a demanda de mercado. Se

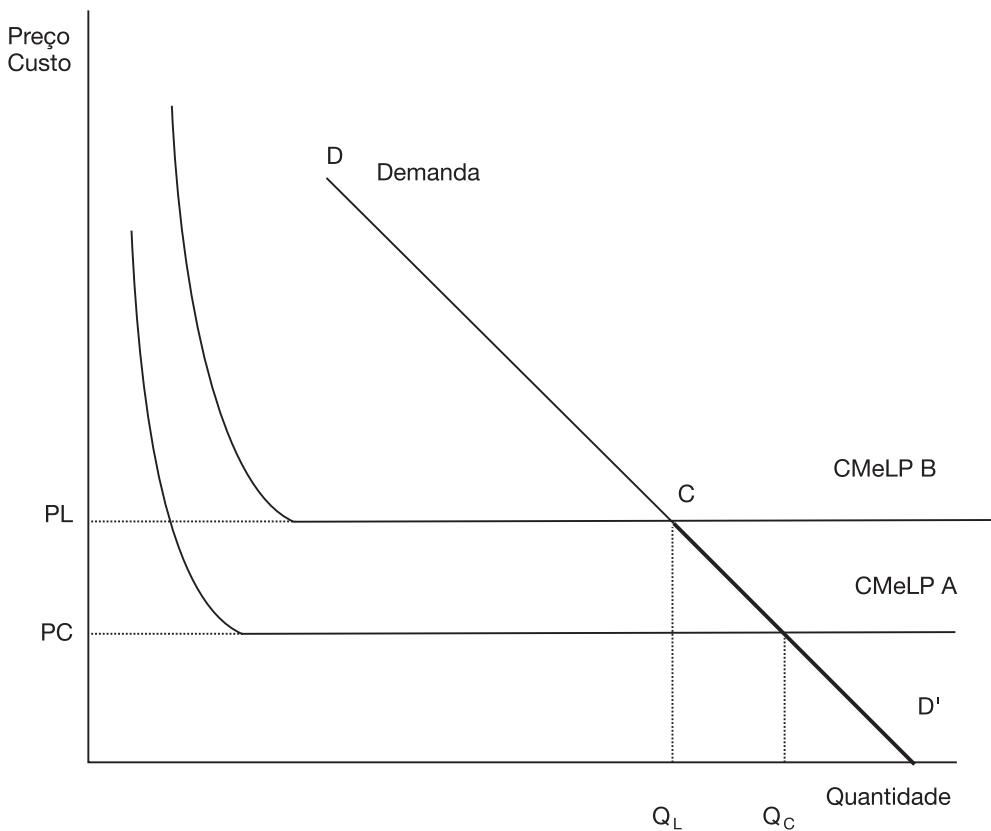


Figura 7.1 Barreira à Entrada Devido a Vantagens Absolutas de Custos. Fonte: Koutsoyannis (1975).

A fixa o preço limite PL no nível justamente abaixo do CMeLP mínimo de B , a demanda residual para B é CD . Nesse caso, não há plano de produção possível para B no qual o preço de mercado seja superior aos seus custos médios de longo prazo. Como resultado, B não irá entrar e A poderá produzir Q_L e auferir os lucros extraordinários correspondentes a $(PL - PC) \cdot Q_L$.

De modo geral, as vantagens de custos para as empresas estabelecidas surgem como reflexo de (1) melhores condições de acesso a fatores de produção, principalmente tecnologia e recursos humanos e naturais; (2) acumulação de economias dinâmicas de aprendizado; ou ainda (3) imperfeições nos mercados de fatores.

Uma das principais fontes de diferenciais absolutos de custos é a tecnologia. Empresas já em operação podem deter patentes que restrinjam ou impeçam o acesso das empresas entrantes às técnicas produtivas mais eficientes. Também é possível que economias decorrentes do aprendizado acumulado pelas empresas estabelecidas sejam relevantes na explicação dos diferenciais de custos.

O acesso a matérias-primas também pode ser mais favorável para empresas já estabelecidas em determinadas circunstâncias de funcionamento desses mercados. Isso tende a ser particularmente frequente no caso do abastecimento de recursos naturais no qual é provável que as empresas existentes já explorem as reservas de melhor relação custo-qualidade, deixando para as empresas entrantes fontes de matérias-primas com custos maiores de exploração ou transporte ou ainda de menor qualidade. Analogamente, é de se esperar que, se os recursos humanos de maior qualificação são escassos, esses já tenham sido contratados pelas empresas em operação, impondo às empresas entrantes despesas elevadas em treinamento de pessoal ou a necessidade de pagar salários mais altos para atrair a mão de obra já empregada.

Finalmente, o acesso ao capital também tende a se dar de forma mais favorável para empresas já existentes. Em vista de imperfeições dos mercados de capitais, que normalmente não conseguem avaliar riscos com exatidão, o financiamento a empresas já constituídas, que podem oferecer garantias reais, é concedido a taxas de juros inferiores ou prazos mais longos que para novos projetos. Em vista disso, os encargos financeiros tendem a ser maiores, pressionando os custos das empresas entrantes.

É importante salientar que embora sejam normalmente consideradas estruturais, as vantagens de custos podem ser modificadas por estratégias específicas de empresas, dentre as quais a busca de integração vertical, visando exercer posição de controle nos mercados de insumos, ou de proliferação de produtos, para ampliar os custos da empresa entrante, estão entre as mais relevantes.

Também é necessário enfatizar que há diversas situações de entrada que enfraquecem ou mesmo anulam as vantagens de custos para as empresas estabelecidas. Se, por exemplo, a empresa entrante é inovadora – detém melhor tecnologia que as empresas existentes – não é de se esperar que a entrada resulte eficazmente impedida.

Barreiras à entrada decorrentes de vantagens absolutas de custos são teoricamente compatíveis tanto com a definição de Bain quanto com a de Stigler, apresentadas na Seção 7.3. Empiricamente, no entanto, vantagens de custos são consideradas fontes pouco relevantes de barreiras à entrada na indústria em geral, apresentando importância restrita a um conjunto limitado de ramos industriais. Estão nesse grupo as indústrias extrativas de primeiro processamento de recursos naturais, tais como a metalurgia ou a indústria de minerais não metálicos ou ainda algumas agroindústrias.

7.5.2 EXISTÊNCIA DE ECONOMIAS DE ESCALA

Sugerida por Joe Bain como uma fonte igualmente fraca de barreira à entrada, a possibilidade de que a existência de significativas economias de escala em uma indústria² venha a constituir de fato impedimento à atuação de novas empresas é certamente a questão mais controvertida nesse tema. F. Modigliani e Sylos-Labini foram os autores pioneiros na discussão sobre a pertinência da adoção de um preço limite quando as estruturas de custos de empresas estabelecidas e empresas entrantes são rigorosamente iguais e o produto é homogêneo. Nessa linha de análise, as condições requeridas para a existência de barreiras de escala (*scale barriers*) são:

1. Escala mínima eficiente (EME) não negligenciável em comparação com o tamanho da demanda de mercado;
2. Custos médios de produção em escalas subótimas sensivelmente superiores aos custos médios mínimos de longo prazo, isto é, a elevada inclinação da curva de custos na região de escala subótima.

Podemos imaginar duas possíveis hipóteses de entrada. Na primeira, sabendo que a decisão de entrar em escala mínima eficiente provocará um incremento substantivo da oferta da indústria e, em consequência, uma sensível redução do preço de mercado, a empresa entrante opta por uma entrada em pequena escala (em escala subótima). Nesse caso, as empresas estabelecidas, que operam em escalas eficientes, poderão praticar um valor de preço limite correspondente ao diferencial de custos relativos a seu favor. Há consenso entre os economistas industriais que o resultado provável é a situação de entrada eficazmente impedida. Questiona-se, no entanto, o realismo dessa hipótese uma vez que somente em certas situações particulares, como por exemplo, quando a tecnologia confere grande flexibilidade de capacidade produtiva às empresas entrantes (modularização do investimento), pode-se esperar entradas em escalas subótimas.

Em uma segunda hipótese, mais realista, a empresa potencial decide entrar com a escala mínima eficiente. Nesse caso, poderá ou não haver excesso de capacidade produtiva na indústria, sendo necessário formular hipóteses adicionais a respeito da reação das empresas estabelecidas. É exatamente nesse ponto que a controvérsia ganha seus contornos mais nítidos.

As análises estruturais do problema da entrada geralmente aceitam que a reação mais provável das empresas estabelecidas é a chamada “não resposta em quantidade”. Denominada posteriormente como “postulado de Sylos”, a manutenção da quantidade produzida pelas empresas existentes implicará, no segundo período, excesso de oferta na indústria que, por sua vez, irá decretar uma queda do preço pós-entrada. Haverá portanto um nível de preço pré-entrada capaz de proporcionar às empresas estabelecidas uma lucratividade positiva, mas que irá reduzir-se para um nível pós-entrada incompatível com a obtenção de lucros pela empresa entrante.

A Figura 7.2 mostra como se dá a fixação do preço limite nas situações em que prevalece o “postulado de Sylos” de acordo com o modelo proposto por F. Modigliani.³ Na figura, DD é a demanda de mercado e CMe é o custo médio de produção das empresas existente e entrante, supostos idênticos. Q_{EME} é a escala mínima eficiente que, como podemos perceber, é significativa quando comparada ao tamanho da demanda ao preço competitivo Q_C . A empresa existente poderá escolher a quantidade Q_L , exatamente aquela para a qual a demanda residual disponível para a empresa entrante não permite remunerar os custos médios de produção e, portanto, impossibilita uma entrada lucrativa.

O grau de barreira à entrada de escala será tanto maior quanto maior for a redução esperada para o preço pós-entrada. Essa, por sua vez, será tanto maior quanto menor a elasticidade-preço da demanda de mercado e maior a relação entre escala mínima eficiente e o tamanho da demanda ao preço competitivo. Isso é facilmente estabelecido na formulação a seguir:

1. Sob o postulado de Sylos, se o incremento da quantidade total de mercado ΔQ é devido exclusivamente à entrada e esta se dá na escala mínima eficiente Q_{EME} , então, $\Delta Q = Q_C - Q_L = Q_{EME}$
2. O preço pré-entrada P_L é tal que após a entrada reduz-se para P_C . Então, $\Delta P = P_L - P_C$
3. A elasticidade-preço da demanda de mercado é:

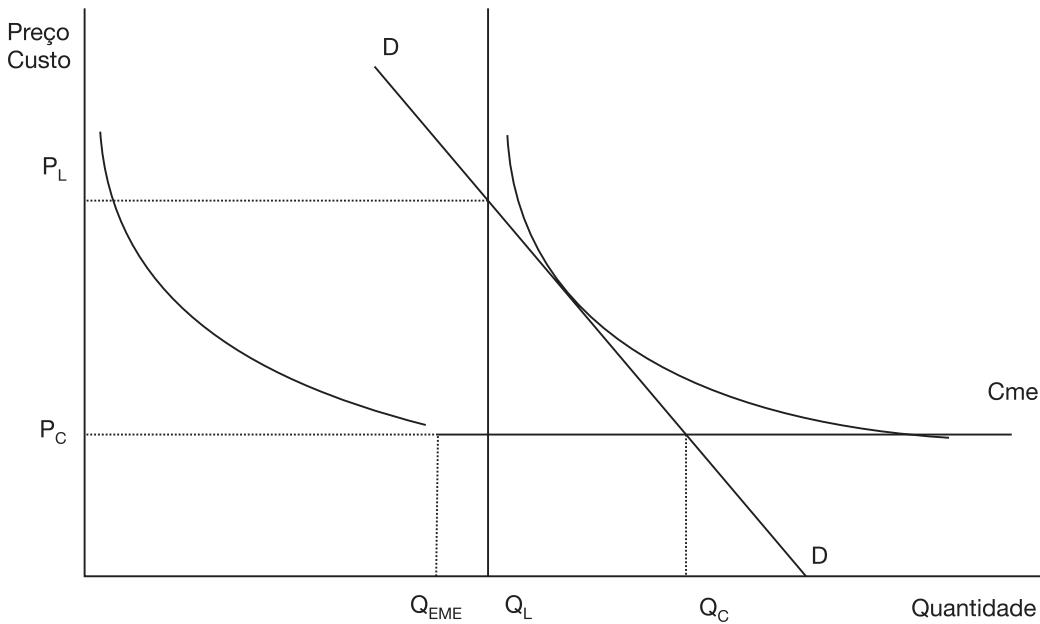


Figura 7.2 Economia de Escala e Barreira à Entrada: O modelo de Modigliani. Fonte: Hay e Morris (1994).

$$\varepsilon = \frac{\frac{\Delta Q}{Q_C}}{\frac{\Delta P}{P_C}} = \frac{Q_{EME}}{Q_C} * \frac{P_C}{P_L - P_C} \quad \text{ou} \quad P_L = P_C \left(1 + \frac{Q_{EME}}{Q_C * \varepsilon} \right)$$

4. Finalmente, a condição de entrada é $E = \frac{Q_{EME}}{Q_C * \varepsilon}$, diretamente proporcional à relação entre escala mínima eficiente e o tamanho do mercado ao preço competitivo e inversamente proporcional à elasticidade-preço da demanda.

Outra opção comportamental possível para as empresas existentes é a “não resposta em preço”. Nessa hipótese, as empresas estabelecidas optam por contrair o nível de produção para manter a oferta total nos níveis pré-entrada, promovendo, assim, a acomodação da capacidade adicional introduzida pela empresa entrante no mercado. Há os que defendam esse comportamento como o mais plausível, uma vez que a manutenção da quantidade significaria uma guerra de preços que imporia perdas também para as empresas existentes, ao passo que a acomodação poderia lhes propiciar ainda algum nível de lucratividade pós-entrada.

A hipótese de reação baseada na manutenção do preço, se corresponde a um cenário mais otimista para a empresa entrante, sugere um comportamento ingênuo das empresas estabelecidas. Na prática, é necessário que as empresas estabelecidas, primeiro, não zelem pela preservação dos níveis de participação no mercado alcançados, possivelmente à custa de grandes esforços competitivos anteriormente empreendidos, e, segundo, contentem-se com uma dupla pressão negativa sobre seus lucros, pois venderão quantidades menores e ainda enfrentarão um acréscimo de custos por passarem a operar mais próximo ou mesmo abaixo da escala mínima eficiente. Finalmente, o comportamento acomodativo poderá ainda estimular novas entradas na indústria, agravando as perdas iniciais sofridas pelas empresas estabelecidas.

É fácil perceber que o aprofundamento dessa questão vai além dos limites do enfoque estrutural do problema da entrada. A pura existência de economias de escala não impõe à empresa entrante qualquer custo que a empresa existente não tenha incorrido quando iniciou operação. Por essa razão, J. Stigler e seguidores rejeitam a possibilidade de que existam barreiras à entrada por esse motivo. Se não há qualquer assimetria de custos entre empresas recém-entrantes e empresas estabelecidas, não há razão para a empresa entrante acreditar que ocorra realmente uma guerra de preços após a entrada. Essa análise será retomada no Capítulo 12.

Quais as implicações da existência de barreiras de escala do ponto de vista da teoria do bem-estar? Basicamente, a existência de economias de escala de *per se* não significa perdas de bem-estar. Isso deixa de ser verdadeiro se derem margem à criação de barreiras à entrada. Se uma indústria envolve elevadas economias de escala, a eficiência alocativa requer uma estrutura

industrial com alto grau de concentração para que a minimização de custos possa ser alcançada. Esse é o caso do monopólio natural. Se, no entanto, a indústria é concentrada devido a barreiras à entrada, ocorre perda de bem-estar porque na ausência de concorrência perfeita, lucros extraordinários estão sendo obtidos pelas empresas existentes.

7.5.3 DIFERENCIAMENTO DE PRODUTOS

Na competição real, a existência de diferenciação de produtos influencia a estrutura de mercado nos moldes já discutidos no Capítulo 6. Implica a existência de algum grau de controle de seus preços pela empresa porque as curvas de demanda individual não são infinitamente elásticas, tornando possível elevar os preços acima do custo marginal sem comprometer completamente a receita.

Na competição potencial, a presença de diferenciação de produtos pode implicar a existência de barreiras à entrada. Se há lealdade dos consumidores para com os produtos vendidos pelas empresas existentes, as empresas entrantes têm forçosamente que vender a preços mais baixos para deslocar preferências estabelecidas ou incorrer em gastos superiores de publicidade para divulgar a nova marca. No primeiro caso, a curva de demanda para a empresa entrante ficará abaixo e à esquerda da curva de demanda das empresas estabelecidas. No segundo caso, haverá vantagens de custos para as empresas estabelecidas que farão com que os custos médios de longo prazo da empresa entrante sejam superiores em todos os níveis de quantidade. De fato, podemos notar que o custo médio total da empresa entrante equivalerá ao custo médio de produção da empresa existente mais uma parcela referente ao custo médio de penetração no mercado. Esse diferencial de custos médios poderá ser apropriado pelas empresas existentes como um sobrepreço permanente, tal qual quando existem vantagens absolutas de custos.

A análise dos custos médios de penetração é bastante complexa, de vez que forças contraditórias atuam sobre o resultado final. De um lado, há a possibilidade de que existam economias de escala importantes no que se refere ao esforço de venda, em particular, quando baseado em propaganda. De outro, é também esperado que o processo de conquista de novos consumidores enfrente dificuldades crescentes. A escala da entrada irá afetar o grau de barreira à entrada em uma ou outra direção, dependendo de qual efeito será predominante.

A mensuração do grau de barreira à entrada devido à diferenciação de produtos envolve um conhecimento aprofundado dos custos de servir ao mercado, mas difíceis de contabilizar que os custos de produção. De modo geral, dependerá das características do produto, da dimensão do esforço de venda imposto pelas firmas existentes e da natureza dos métodos de distribuição – exclusividades comerciais, montagem de redes de distribuição próprias ou de terceiros, entre outros. Em vista disso, é de difícil generalização.

Mesmo assim, há virtual unanimidade entre os economistas industriais de que a diferenciação de produto é a mais forte dentre as fontes de barreira à entrada na indústria. O motivo para isso é o fato de que são amplas as possibilidades de criação de vantagens da primeira empresa a se mover baseadas nas preferências dos consumidores. De modo geral, podemos esperar que a empresa entrante depare-se com a necessidade de deslocar preferências consolidadas pelas marcas das empresas estabelecidas. Em especial diante da diferenciação vertical, isso poderá ser muito custoso, pois a qualidade do produto oferecido pela empresa entrante é incerta para o consumidor. Essas vantagens da primeira empresa a se mover são particularmente intensas nas indústrias de bens duráveis – porque a pouca repetição do ato de compra dificulta o aprendizado do consumidor – e de maior valor unitário – porque os custos do “arrependimento”, que são proporcionais ao valor do bem serão altos. A indústria automobilística é um dos melhores exemplos dessa situação.

Mais uma vez, a eficácia do impedimento à entrada baseada em diferenciação de produtos é reduzida ou anulada quando a empresa entrante é uma subsidiária de uma empresa que detém produtos ou marcas conceituadas em mercados de outros produtos (empresas em diversificação) ou regiões (empresas transnacionais). Nessas situações ocorre um fenômeno conhecido como “transbordamento” (*spill-over*), pelo qual há a transferência da credibilidade da empresa do mercado original para o novo mercado.

Finalmente, se não houver necessidade de deslocar preferências dos consumidores, isto é, se o esforço de venda da empresa entrante for equivalente ao realizado pelas empresas estabelecidas quando estas iniciaram operação, não haverá assimetria de custos e, portanto, não é de esperar que haja impedimento à entrada de acordo com a definição de Stigler.

7.5.4 REQUERIMENTOS INICIAIS DE CAPITAL

Bain considerava que poderia existir uma quarta fonte de barreiras se a entrada de uma nova empresa em uma indústria exigisse a mobilização de elevada soma de capital para fazer face ao investimento inicial – as barreiras de capital (*capital barriers*). Assim como as barreiras de escala, as barreiras de capital surgem como consequência da existência de elevadas escalas mínimas eficientes. Porém, o surgimento de barreiras de capital exige somente que a escala mínima eficiente seja elevada em termos

absolutos, mesmo que pouco significativa em comparação com o tamanho da demanda. (Como vimos anteriormente, a relação entre escala mínima eficiente e tamanho da demanda explica a existência ou não de barreiras de escala.) Barreiras de capital seriam um reflexo direto da dificuldade em financiar os grandes volumes de capital requeridos quando o investimento inicial é muito elevado, não tendo qualquer relação com os impactos sobre os preços ou lucratividade provocados pelo aumento da oferta total da indústria em consequência da entrada.

Tema controverso na literatura de Economia Industrial, a rigor barreiras de capital não se encaixam em nenhuma das definições de barreira à entrada expostas na Seção 7.3. De fato, salvo em situações em que há imperfeições no mercado de capitais, não é de se esperar que um empreendimento lucrativo deixe de ser realizado em função de escassez de fundos para financiar o investimento inicial, ainda mais se isso não ocorreu quando do investimento inicial das empresas estabelecidas. E, se essas imperfeições existem, elas irão implicar diferenciais absolutos de custos: do ângulo estrutural, barreiras de capital seriam tão somente um caso particular das barreiras de custos (financeiros, nesse caso) já discutidos anteriormente, e não uma fonte própria de impedimento à entrada na indústria.

Por que, então, é intuitivo que quanto maior o investimento inicial a ser realizado, mais difícil tende a ser a entrada de uma nova empresa na indústria? A resposta está no fato de que elevados requerimentos iniciais de capital são geradores de barreiras à saída, pois esses investimentos tendem a ser formados, em grande parte, por custos irrecuperáveis que irão ter importantes implicações sobre o comportamento estratégico das empresas. A discussão desse tema, no entanto, exige elementos analíticos que somente serão apresentados no Capítulo 12.

7.6 Barreiras à Saída: os Modelos de Contestabilidade

Ao final da década de 1970 passou-se a atribuir grande importância às condições que regem a saída de empresas da indústria como elementos relevantes da análise da decisão de entrada.

Barreiras à saída decorrem da existência de custos que as empresas necessitam arcar para encerrar a produção. Esses custos podem ser desembolsos efetivos como, por exemplo, os custos de rescisão dos contratos em vigor, ou custos de oportunidade referentes a investimentos realizados e ainda não totalmente amortizados, e que não tenham valor de revenda; os chamados custos irrecuperáveis (ver Capítulo 12).

A Teoria da Contestabilidade ganhou corpo no início da década de 1980. Representa uma hipervalorização da competição potencial diante da competição real. Para essa teoria, estrutura e mesmo condutas pouco importam porque o desempenho é consequência das condições básicas (entenda-se, funções de custos, em particular os custos irrecuperáveis) dos mercados. A estrutura da indústria é o resultado da determinação conjunta dos “planos” de produção (preços e quantidades) das firmas que a constituem. O mecanismo de equilíbrio na contestabilidade é devido à entrada e à saída de empresas nas indústrias em que a configuração endógena resultante é “não sustentável”. O equilíbrio é assim assegurado pela existência de livre mobilidade do capital no sentido clássico e não propriamente por ações e reações das empresas rivais em uma dada indústria. Em outras palavras, o importante nas teorias de mercados contestáveis é a concorrência definida pela existência ou não de custos irrecuperáveis significativos para a empresa entrante.

Segundo essa teoria, uma configuração industrial é factível quando todas as empresas da indústria atendem a totalidade da demanda sem incorrer em prejuízo, o que depende somente da extensão do mercado, da técnica e dos preços dos fatores. Uma configuração industrial é sustentável quando, além de factível, não há plano possível para uma empresa entrante que lhe forneça lucros com os preços e quantidades que vigoram no mercado. Nesse caso, a indústria pratica preços tais que $P^e - C(y^e) < E(y^e)$, onde P^e são os preços, y^e as quantidades, $C(y^e)$ os custos totais de produção das empresas estabelecidas e $E(y^e)$ os custos de entrada.

Preços não sustentáveis seriam a princípio temporários, isto é, toda configuração industrial não sustentável estaria em desequilíbrio e tenderia para uma configuração sustentável. A exceção ocorreria se as empresas entrantes recebessem que as empresas estabelecidas reduzissem os preços para o nível sustentável após a entrada.

Há, no entanto, uma condição estrutural suficiente para assegurar que o equilíbrio da indústria ocorra sempre com configurações industriais sustentáveis: o mercado perfeitamente contestável, isto é, quando a entrada na indústria é livre e reversível sem custos. São os seguintes os requisitos para a existência de um mercado perfeitamente contestável:

1. Ausência de barreiras à entrada (na definição de Stigler) – não há diferenciais de custos entre empresas entrantes e estabelecidas porque ambas têm acesso aos mesmos fatores de produção e a mesma habilidade de servir mercados (qualidade, marcas etc.).
2. Ausência de barreiras à saída – custos irrecuperáveis são nulos, isto é, o capital empregado pode ser inteiramente recuperado seja porque pode ser revendido sem perdas, transferido para outra indústria ou ainda porque foi alugado ou subcontratado.

Fundamentalmente, essas duas condições permitem a existência de um tipo de competição de guerrilha (*hit and run competition*) capaz de disciplinar as decisões de preços das empresas mesmo no caso extremo do monopólio natural. Qualquer sobrepreço imposto pelas empresas estabelecidas daria lugar a uma entrada lucrativa até quando ocorresse a retaliação em preços. Nesse momento, a empresa entrante abandonaria a indústria, sem perdas devido à ausência de custos irrecuperáveis, retendo, assim, os lucros obtidos. O principal exemplo desse tipo de competição estaria no mercado de linhas aéreas porque o principal investimento envolvido – o avião – é um custo fixo, mas não é um custo irrecuperável.

A teoria da contestabilidade, como amplamente reconhecido, apresenta algumas limitações conceituais no que diz respeito à definição do monopólio natural e principalmente quanto à hipótese implícita, pouco geral, de que os preços se ajustam mais lentamente que as quantidades.

7.7 Conclusão

Nossa discussão sobre barreiras estruturais à entrada ao mesmo tempo em que trouxe uma grande riqueza de elementos para a análise do funcionamento dos mercados também deixou claro uma série de limitações. Embora a teoria seja útil para descrever características estruturais das indústrias, necessita ser enriquecida.

Análises estruturais da concorrência por meio de barreiras à entrada têm como principal limitação o seu caráter estático. A decisão de entrada é enfocada como dependente de uma expectativa de obtenção de lucros imediatamente após a entrada. Além disso, supõe que a demanda é dada, ao passo que a taxa de crescimento da demanda é forte determinante do grau de barreira à entrada na indústria. Quanto mais dinâmica a demanda, mais fracas tendem a ser as possibilidades de impedimento à entrada. Finalmente, falta uma reflexão mais aprofundada sobre as barreiras à saída.

Empresas estabelecidas e entrantes realizam investimentos e tomam decisões sob incerteza, baseadas em conjecturas sobre as suas quase-rendas futuras. As estratégias das empresas estabelecidas afetam tanto as conjecturas dos entrantes como as barreiras estruturais à entrada. Em consequência, o grau de barreira à entrada em uma indústria é somente em parte estrutural; em parte é também endógeno, isto é, decorrente das estratégias competitivas das empresas.

São muitos os elementos que influem na decisão de investimento da empresa entrante. As rendas presentes auferidas pelos ocupantes do mercado, o grau de barreira à entrada estática ou estrutural, as reações à entrada das empresas existentes, o comportamento provável dos outros membros da fila de empresas entrantes, os recursos relevantes já detidos pelos entrantes e a magnitude dos custos irreversíveis de reunir informações e decidir. Esses elementos variam de indústria para indústria e de tempo para tempo, implicando grande dificuldade de estabelecer princípios gerais sobre o problema da entrada.

É inegável que existem fragilidades no conceito de preço-limite. Preço e demais formas de competição que repercutem instantaneamente sobre a relação preço-custo não são os únicos mecanismos de concorrência. A luta competitiva envolve também mecanismos que exigem investimentos prévios, muitas vezes de grande monta, dos quais a inovação é o principal, mas não único exemplo. Mesmo considerando somente o preço, o grau de barreiras à entrada não é o único elemento a explicar o comportamento das empresas.

A despeito dessas limitações, o enfoque estrutural das barreiras à entrada continua sendo uma ferramenta útil para a análise do funcionamento da indústria.

A novidade introduzida por Bain e seguidores não foi a constatação da existência de restrições à mobilidade do capital, mas a eleição da barreira à entrada como principal elemento da determinação do preço: significa deslocar o eixo da teoria de formação de preços da concorrência real para a concorrência potencial e, consequentemente, do curto para o longo prazo.

Se há barreira à entrada em uma indústria, então se aceita a existência do diferencial econômico entre empresas estabelecidas e novas empresas que, no entanto, pode assumir várias formas. A mais comum é associá-lo à relação preço-custo das empresas existentes e entrantes e utilizá-las em teorias de preço limite, como explicação estrutural para o *mark-up* praticado pelas empresas.

Não há dúvida de que a decisão de entrada de uma empresa também depende das avaliações *ex-ante* que ela faça das reações da empresa existente – por exemplo, se responde em preço ou em quantidade a uma entrada em larga escala. Por isso, é inegável que barreira à entrada seja apenas, em parte, um elemento estrutural e não deva ser reduzida a um aspecto formal da estrutura do mercado como o número de concorrentes. Barreiras à entrada são um elemento constitutivo do próprio oligopólio, elemento-chave da formulação de uma teoria da concorrência em oligopólio, tanto potencial quanto real. Afinal, são as mesmas forças que orientam as decisões competitivas nos dois casos.

Notas

1. A literatura crítica a esse enfoque consagrou denominá-las também como barreiras à entrada inocentes.
2. Os conceitos e a análise empírica do fenômeno da economia de escala são discutidos em profundidade no Capítulo 3.
3. O modelo proposto por Sylos-Labini considera a existência de descontinuidades tecnológicas na indústria, de sorte que há três tamanhos de empresas (pequenas, médias e grandes) na estrutura industrial. Com isso, define-se uma faixa de variação para o preço limite, desde um valor superior correspondente ao preço máximo capaz de prevenir entradas até um valor inferior correspondente ao preço de exclusão. Independentemente dessa contribuição, as conclusões principais da modelagem de Sylos-Labini são as mesmas obtidas por F. Modigliani. Os leitores interessados em conhecer o modelo podem encontrar uma excelente resenha em Koutsoyannis.(1975).

Bibliografia

- Bain, J. (1956) *Barriers to New Competition*. Cambridge: Harvard University Press.
- Baumol, W.; Panzar, J. C.; Willig, R. D. (1982) *Contestable Markets and the Theory of Industry Structure*. Harcourt. San Diego.
- Caves R.E.; Porter M.E. (1977) From Entry Barriers to Mobility Barriers. *Quarterly Journal of Economics*. vol XCI. no 2. may. pp. 241-62.
- Davies, S.; Lyons, B. (1988) *Economics of Industrial Organization. Survey in Economics*. London. Longman.
- Ferguson, P.R.; Ferguson, G.L. (1994) *Industrial Economics: Issues and Perspectives*. Macmillan. UK.
- Gerosky, P. (1988) Competition Policy and the Structure-Performance Paradigm In Davies, S.; Lyons (1988). *op. cit.* 166-191.
- Gilbert, R. (1989). Mobility Barriers and the Value of Incumbency. Schmalensee, R. and Willig, R. (eds). *Handbook of Industrial Organization*. vol I. Elsevier. B.V. capítulo 8.
- Guimarães, E.A. (1982) *Acumulação e Crescimento da Firma*. RJ. Zahar.
- Hay, D. e Morris, D. *Industrial Economics: Theory and Evidence*. Harper Collins, 1994.
- Koutsoyannis, A. (1975) *Modern Microeconomics*. Macmillan. Houndsills.
- Modigliani F. (1958) New Developments on the Oligopoly Front. *Journal of Political Economy* 66(3): 215-232.
- Possas, M. L. (1985) *Estruturas de Mercado em Oligopólio*. São Paulo, Hucitec.
- Reid, G.C. (1987) *Theories of Industrial Organization*. Basil Blackwell. U.K.
- Scherer, F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*. Boston, Houghton Mifflin.
- Schmalensee R. (1988) Industrial Economics: An Overview. *Economic Journal*. 98. sept. pp. 643-68.
- Shepperd (1984) Contestability vs. Competition. *American Economic Review*. (74). pp. 572-87.
- Stigler, G. J. (1968) *The Organization of Industry*. Homewood: Richard D. Irwin.
- Sylos-Labini, P. (1980) *Oligopólio e Progresso Técnico*. RJ. Forense Universitária.
- Tirole, J. (1988) *The Theory of Industrial Organization*. Cambridge (Mass.): MIT Press.

Estrutura de Mercado e Inovação

Lia Hasencllever e Patrícia Moura Ferreira

8.1 Introdução

O estudo da inovação tecnológica foi durante muito tempo esquecido pela análise econômica que priorizava análises de equilíbrio de curto prazo ou, quando se tratava de analisar o longo prazo, dedicava-se à análise da acumulação de capital e da distribuição de renda.

Foi somente após a Segunda Guerra Mundial que as ideias apresentadas por Joseph Schumpeter começaram a florescer fundando o que hoje se denomina Economia da Inovação. Em seu livro *Teoria do Desenvolvimento Econômico*, publicado em 1912, o autor observa que a inovação cria uma ruptura no sistema econômico, no interior das indústrias, revolucionando as estruturas produtivas e criando fontes de diferenciação para as empresas.

A Economia da Inovação é o ramo da Economia Industrial que tem como principal objeto de estudo as inovações tecnológicas e organizacionais introduzidas pelas empresas para fazerem frente à concorrência e acumularem riquezas.

Em termos metodológicos coexistem dois enfoques sobre o problema da inovação: relação entre inovação e estrutura industrial e entre inovação e estratégias tecnológicas. Este capítulo trata do primeiro destes enfoques e resume os principais resultados dos estudos empíricos realizados sobre o tema entre as décadas de 1960 e 1980. No Capítulo 19, sobre estratégias de inovação, serão analisadas as estratégias de entrada e permanência das empresas baseadas em inovações tecnológicas e organizacionais e as consequências dinâmicas desse comportamento sobre a estrutura industrial.

Este capítulo está organizado em quatro seções. A primeira apresenta as principais definições sobre a inovação industrial e os elementos constitutivos desse processo. A segunda seção apresenta os principais modelos econômicos de compreensão da inovação industrial. A terceira seção discute a agenda de pesquisa derivada dos trabalhos de Schumpeter para explicar a intensidade das inovações a partir de sua relação com o tamanho das empresas ou os padrões de concorrência existentes no mercado. Finalmente, na conclusão, tecemos algumas considerações sobre o Capítulo 19 de estratégias tecnológicas, que complementa este com os avanços teóricos permitidos a partir da crítica aos trabalhos empíricos realizados sobre a relação entre inovação e estrutura industrial.

8.2 Inovação Industrial

Antes de abordar as questões sugeridas na introdução é importante definir o processo de mudança tecnológica. O processo de mudança tecnológica é resultado do esforço das empresas em investir em atividades de pesquisa e desenvolvimento (P&D) e na incorporação posterior de seus resultados em novos produtos, processos e formas organizacionais. Quando uma empresa

produz um bem ou um serviço ou usa um método ou insumo que é novo para ela, está realizando uma mudança tecnológica. Sua ação é denominada inovação.

Além da empresa e de suas atividades de P&D, o conjunto de instituições que contribui para a inovação e a ligação entre elas compreende o que o autor B.A. Lundvall chamou sistema de inovação nacional. Entre essas instituições estão as universidades, institutos públicos de pesquisa, agências públicas e privadas de fomento ao investimento em inovação e sistema educacional.

As atividades de P&D referem-se à pesquisa básica, pesquisa aplicada e desenvolvimento experimental. As primeiras regras e normas estabelecidas para a classificação dessas atividades, ainda hoje vigentes, são aquelas elaboradas pela Organização de Cooperação e Desenvolvimento Econômico (OCDE) há 30 anos e publicadas no, hoje popularizado, Manual Frascati. A partir da década de 1990, esse Manual foi revisto e ampliado para incluir uma visão mais abrangente do processo de inovação, quando foi criado o Manual Oslo¹, conforme discutido na subseção 8.4.1.

Como pesquisa básica entende-se trabalho teórico e experimental, empreendido primordialmente para compreender fenômenos e fatos da natureza, sem ter em vista qualquer aplicação específica. Como pesquisa aplicada, classificam-se as investigações originais concebidas pelo interesse em adquirir novos conhecimentos com finalidades práticas. E, como desenvolvimento experimental, entende-se a comprovação da viabilidade técnica/funcional de novos produtos, processos, sistemas e serviços, ou ainda o substancial aperfeiçoamento dos já existentes, buscada por meio de esforços sistemáticos a partir de conhecimentos técnico-científicos e/ou empíricos já dominados pela empresa ou obtidos externamente.

O ciclo de inovação pode ser dividido em três estágios: invenção, inovação e imitação ou difusão. O processo de invenção está relacionado com a criação de coisas não existentes anteriormente e utiliza como principais fontes conhecimentos novos ou conhecimentos já existentes em novas combinações. Os resultados desse processo podem ser patenteados, isto é, o inventor é investido de direitos de propriedade sobre o uso comercial de sua invenção. No entanto, nem todas as invenções ou mesmo patentes chegam a se transformar em inovações, isto é, serem lançadas no mercado com sucesso comercial. A introdução de inovações, por sua vez, permite a introdução de outras variações denominadas imitação (difusão das inovações). Essas variações são melhorias introduzidas nos bens e serviços inovadores para aproximar-los das necessidades dos usuários. Entretanto, o processo de imitação também pode ocorrer sem introdução de melhorias.

A introdução de uma inovação associada a um processo de invenção dá origem ao que se denomina inovação radical e o processo de imitação com introdução de melhorias é denominado introdução de inovação incremental.

8.3 Modelos de Análise Econômica da Inovação

Existem dois modelos alternativos e concorrentes utilizados para abordar a questão econômica da inovação. Esses modelos diferem tanto em função da concepção acerca da natureza do processo inovativo, quanto em termos das principais previsões decorrentes de suas conclusões. O primeiro deles deriva-se da tradição neoclássica e foi formulado, em 1962, por Kenneth Arrow. É comumente conhecido como *modelo de incitação* e foi concebido para duas formas opostas de concorrência: concorrência pura e monopólio, tendo sido posteriormente modificado para o estudo de várias formas de concorrência por P. Dasgupta e J. Stiglitz, em 1980. O segundo modelo denomina-se *modelo de seleção* e segue a tradição evolucionista iniciada por E. Penrose e A. A. Alchian, nos anos 1950, retomada por Sidney Winter por volta dos anos 1960 e, finalmente, consolidada por este autor em conjunto com Richard Nelson nos anos 1980. As diferentes concepções sobre a natureza do processo de inovação presentes nestes modelos e os seus conteúdos e principais previsões serão detalhados a seguir.

8.3.1 MODELO DE INCITAÇÃO

O modelo de Arrow considera as seguintes hipóteses derivadas da teoria neoclássica como quadro geral para formulação de sua teoria: (1) o conhecimento é um bem de informação e todos os agentes são capazes de obtê-lo de forma igual e sem custos derivados de investimentos passados ou decorrentes da experimentação, realizados pelas empresas, ao contrário, é fruto da genialidade dos inventores individuais; (2) as únicas situações de mercado existentes são a concorrência e o monopólio.

Decorrente desse quadro de análise a atividade de invenção possui características próprias, também indicadas pelo autor. Esta atividade é arriscada e/ou sujeita a incertezas. Em segundo lugar, o produto dessa atividade pode ser apropriado por outros agentes. E, enfim, existem rendimentos crescentes na utilização da nova informação.

Entretanto, sua problemática central é saber se existem vantagens decorrentes das características das duas estruturas de concorrência consideradas – concorrência pura e monopólio – no que diz respeito à motivação para investir em P&D. Para que

uma empresa invista em P&D, é preciso que o ganho resultante da inovação lhe seja atrativo, ou seja, é necessário que ele seja suficientemente elevado para financiar a P&D (atividade altamente arriscada) e assegurar um rendimento.

Arrow mostra que o monopólio já garante à empresa um sobre lucro derivado da própria situação de poder de mercado da empresa e, dessa forma, será necessário que a inovação permita uma redução de custos substancial para que o empresário esteja motivado a investir em P&D. Por outro lado, a empresa atuando em um mercado concorrencial e sem poder de fixar preços só tem uma forma de ampliar a sua margem de lucros, que é introduzir inovações permanentemente, ainda que seja recompensada com margens de lucros que irão se erodir rapidamente pela entrada de novas empresas imitadoras.

Para se entender melhor o modelo de Arrow, observemos a Figura 8.1 (a) que mostra uma indústria com curva de demanda DD e curva de custo médio e marginal constantes CC antes da inovação. Na situação pré-inovação a quantidade ofertada em monopólio é Q_m , o preço P_m e o lucro π_A ; por outro lado, se for um caso de concorrência perfeita, tem-se quantidade Q_c , preço P_c e lucro zero. O modelo de Arrow supõe que a inovação é feita por uma empresa inovadora de fora da indústria que irá adotá-la, como, por exemplo, uma indústria fornecedora de equipamentos, que tentará apropriar os benefícios da inovação cobrando um *royalty* r do uso da inovação.

Arrow trabalha com duas possibilidades de inovação: drástica e não drástica. Em termos das definições adotadas por Schumpeter, inovação radical e incremental. No caso de uma inovação drástica ocorre uma redução drástica dos custos. Na Figura 8.1 (a) essa redução é representada por $C - C'$, onde C' é o novo custo médio constante. Em concorrência perfeita o inovador cobra r por unidade produzida usando a nova tecnologia. O novo custo marginal é então $C' + r$ e, consequentemente, o preço que maximiza o lucro das empresas, $P = C' + r$. Entretanto, a decisão ótima do inovador envolve maximizar rQ com respeito a r . O seu retorno será a diferença entre a receita da indústria e o custo de produção, ou seja, $(P - C')Q$. Como nesse caso não existe restrição para o inovador, ele escolhe r tal que a receita marginal da indústria se iguale a C' , estabelecendo r^* que gera o retorno π_B .

No monopólio, o inovador cobra uma taxa fixa (*lump sum*) r como *royalty*, pois se fosse uma taxa por unidade o monopolista poderia ficar tentado a restringir a produção. Assim, o lucro do monopolista após adotar a inovação será π_B , onde a receita marginal é igual a C' . Entretanto, mesmo com a arrecadação de r , o máximo que o inovador pode ganhar com a inovação é o aumento do lucro resultante no monopólio, ou seja, $\pi_B - \pi_A$. Note ainda que mesmo que o inovador seja capaz de se apropriar de todo o aumento do lucro, o seu retorno ainda será menor do que no caso competitivo.

A Figura 8.1 (b) mostra uma inovação não drástica, onde a redução do custo é mais moderada. Nesse caso, existe uma restrição, $r \leq C - C'$. Em outras palavras, o máximo que o inventor pode ganhar no caso de concorrência perfeita é $(C - C')Q$, como se pode perceber na figura. Similarmente ao caso da inovação drástica, pode-se mostrar que esse montante é maior que o máximo que ele poderia ganhar no monopólio.

Nos dois casos, inovação drástica e não drástica, a motivação para inovar no monopólio é inferior à da concorrência perfeita. O modelo de Arrow conclui, então, que os mercados competitivos atraem mais atividade de pesquisa e desenvolvimento que as indústrias monopolísticas.

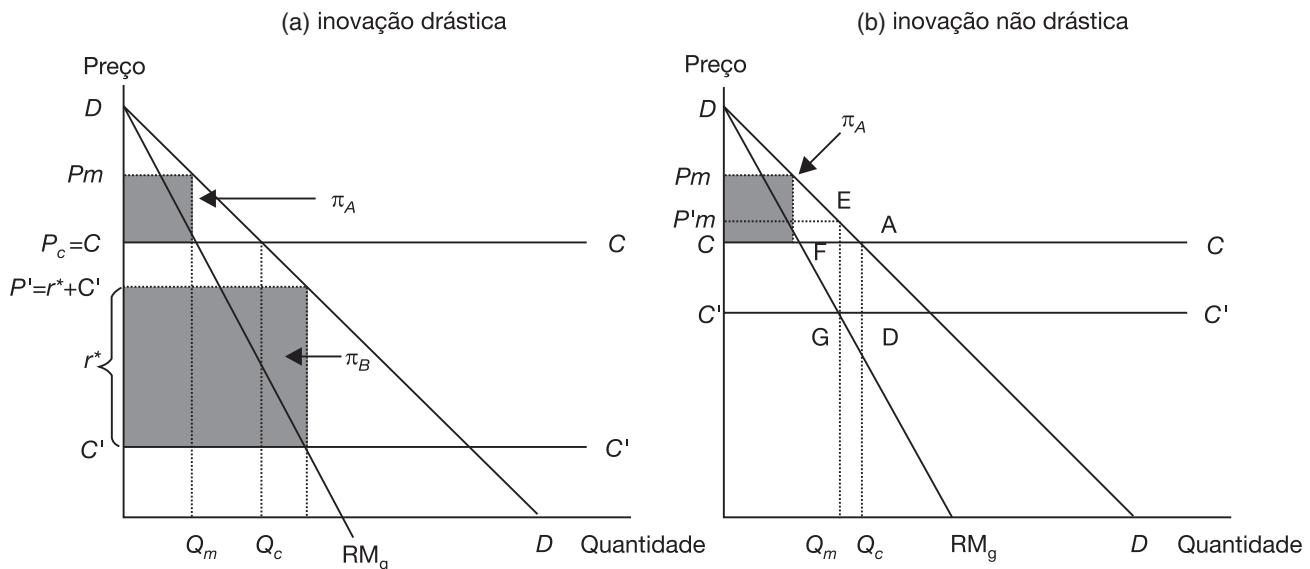


Figura 8.1 Incentivo para Inovar no Modelo de Arrow.

A partir dessas conclusões o autor afirma que uma economia de mercado iria subinvestir em P&D. Isso porque, se aceitamos a tese simples de que uma empresa irá investir uma parte de seus recursos em P&D para produzir uma melhor informação (uma melhor técnica) capaz de lhe agregar uma vantagem, ela estaria prejudicada pelo fato de que outras empresas poderiam igualmente se beneficiar de seus resultados, causando uma impossibilidade privada de apropriação dos resultados do investimento.

O modelo de Arrow, por mais rico que seja, não esgota o estudo das formas de concorrência entre as empresas geradas por meio da inovação. Em outras palavras, a natureza do processo de descoberta próprio das atividades de P&D e a interação entre as empresas dão lugar ao nascimento de interesses estratégicos que alinham as empresas na busca pelas patentes, gerando distintas estruturas de mercado que podem ser mais ou menos concorrentiais.

O modelo Dasgupta-Stiglitz se propõe a estudar o comportamento inovador das empresas e esclarecer como a taxa de inovação interage com a estrutura de mercado predominante. Além disso, o modelo visa avaliar o impacto de variáveis centrais dessa interação: elasticidade-preço da demanda, as barreiras à entrada e a ligação entre investimentos em P&D e redução dos custos unitários. Assim, a P&D é vista como uma variável estratégica que permite à empresa inovar e logo produzir mais eficientemente. Essa vantagem estratégica tem um papel na contenção dos entrantes potenciais, constituindo-se em uma barreira à entrada.

As hipóteses mais importantes do modelo são a ausência de inovação de produto e de competição por preços. Em outras palavras, as empresas concorrem entre si por meio da introdução de inovações de processo que reduzem os seus custos, permitindo-lhes ganhar partes do mercado. A relação entre os investimentos em P&D e a redução do custo médio é explicitada da seguinte maneira: a P&D permite pagar uma redução do custo unitário, ou seja, o custo médio é uma função do montante investido pela empresa em P&D. Seja $c(\cdot)$ o custo médio e R o investimento em P&D, logo:

$$c(\cdot) = \beta(R)^{-\alpha},$$

onde α é a elasticidade pesquisa-inovação.

A curva de demanda não se desloca no decorrer do processo (mas a elasticidade-preço não é infinita, é igual a e). Os preços dos fatores não variam, as empresas que conhecem as suas tecnologias de P&D maximizam seu lucro (as empresas ajustam instantaneamente as quantidades de insumos comprados ao seu nível ótimo). Não há incerteza, ou seja, as empresas têm perfeita informação sobre os preços e as quantidades. Os autores se situam em um cenário de Cournot, conforme visto no Capítulo 10; as empresas irão, então, escolher o seu nível de produto (todas as empresas definem a mesma estratégia).

Suponhamos agora que a indústria é composta de N empresas independentes, as decisões de uma empresa não têm influência sobre as decisões das demais. Consideremos a empresa i que deverá escolher o nível de seu volume de produção (Q_i) e seu investimento em P&D (R_i) que maximiza seu lucro de um período:

$$\text{Max}\{[p(\hat{Q} + Q_i) - c(R_i)]Q_i - R_i\},$$

sendo \hat{Q} o volume de produção da indústria (excetuando-se aquele da empresa i), $p(\hat{Q} + Q_i)$, o preço ou a função de demanda (prova-se que $Q^* = \hat{Q} + Q_i$).

Mostra-se que a condição de primeira ordem implica:

$$\begin{aligned} \frac{p - c}{c} &= \frac{1}{eN - 1} \\ \frac{R^*}{pQ^* \cdot N^{-1}} &= \frac{\alpha}{P/c} = \alpha \left(1 - \frac{1}{eN}\right) \end{aligned}$$

onde R^* é o investimento ótimo da empresa representativa, α é a elasticidade P&D do custo médio (uma alta de 1% da P&D reduz de $\alpha\%$ o custo médio), Q^* o produto (*output*) da indústria ($pQ^* \cdot N^{-1}$ é, consequentemente, o produto da empresa representativa). Em uma situação caracterizada por barreiras à entrada, N é dado e sua escolha ótima do investimento em P&D depende de α , e e N .

O indicador $\frac{p - c}{c}$ é uma medida do grau de monopólio, conhecida em Economia Industrial como Índice de Lerner, que aumenta quando o número de empresas diminui. Entretanto, a intensidade de P&D diminui com o número de empresas (R^*/pQ^*). As soluções descobertas por Dasgupta-Stiglitz implicam que assim que o número de empresas diminua, o produto total da indústria diminui da mesma forma, mas o investimento em P&D de cada empresa aumentará. Se fizermos a hipótese de que não há efeito de transbordamento (*spill over*)¹ possível, é o investimento em P&D da empresa que determina a taxa de inovação (quer dizer, a redução do custo unitário). Nesta configuração, essa taxa aumenta assim que o número de empresas diminui. Tal proposição admite finalmente que quando nos afastamos de uma configuração próxima da concorrência perfeita, a taxa de inovação cresce. Isso remete às duas chamadas hipóteses schumpeterianas, negando a hipótese de Arrow vista anteriormente.

Em uma configuração industrial onde há livre entrada, os novos entrantes erodirão as margens de lucro, reduzindo-o a zero. O grau de monopólio e a intensidade da P&D são também determinados por α . As duas variáveis diminuem com α . No limite para $\alpha=0$ (situação sem progresso técnico), reencontramos o modelo referente ao padrão da concorrência perfeita.

Se estivermos diante da presença de rendimentos de escala crescentes, uma só empresa monopoliza o mercado. A solução do problema de maximização de lucros indica que a intensidade de pesquisa ótima será igual a α (mesmo que ela seja igual a $\alpha/1\alpha$ no caso de uma estrutura oligopolista com entrada livre). Os investimentos em P&D são, portanto, menores que sob o regime de oligopólio com entrada livre; por outro lado, não pode haver desperdício devido à duplicação de esforços de P&D que caracterizam o modelo de concorrência. Em uma situação de monopólio, prevemos uma forte taxa de inovação (em oposição ao caso concorrencial em que existe muito investimento em P&D para pouca inovação).

A hipótese de ausência de transbordamento pode ser relaxada. Nesse caso, as conclusões acima expostas se modificarão ligeiramente. Mas, mesmo assim, o modelo não é capaz de incorporar a interatividade entre as empresas concorrentes (entre líderes e seguidoras), aspecto analisado pelo “modelo de seleção” exposto a seguir.

8.3.2 MODELO DE SELEÇÃO

Neste momento nos limitaremos a enunciar as estratégias de P&D, suas interações com a estrutura de mercado, progresso técnico e desempenhos industriais.² As empresas, tomadas como a unidade de análise fundamental, tendem a buscar técnicas alternativas às atualmente em uso, procurando aumentar sua rentabilidade. Essa é a melhor forma delas se desenvolverem. Uma atitude de se contentarem com as formas antigas de realizar suas atividades as leva em processo de falência.

Suponhamos que estamos dentro de um ambiente evolucionista em que as empresas não respondem da mesma forma aos sinais do mercado. As empresas que inovam e que não são rapidamente imitadas ou as empresas que imitam rapidamente podem em geral dominar a indústria. O modelo leva em conta dois tipos de comportamentos típicos das empresas: as políticas voltadas para a inovação e para a imitação.

Como afirmam Richard Nelson e Sidney Winter, as políticas das empresas não são determinadas por procedimentos de maximização. As empresas ao se engajarem em uma estratégia de P&D não sabem *ex-ante* se vão ou não ser bem-sucedidas; elas não conhecem o nível apropriado de P&D inovativo ou imitativo. Existe um processo de aprendizagem que exclui uma estratégia clássica de busca de um equilíbrio. Somente o curso dos acontecimentos é que revelará se a estratégia foi ou não bem-sucedida.

Suponhamos que existam N empresas na indústria e que a produção da empresa i no tempo t é proporcional a sua capacidade de produção.

$Q_i(t) = A_i(t) \times K(t)$; $A_i(t)$, o inverso do coeficiente de capital, resume a tecnologia.

A produção total da indústria se obtém por somatório e pode-se deduzir o preço de venda do bem.

$$Q(t) = \sum Q_i(t); \\ p(t) = d[Q(t)]$$

A margem unitária de lucro (ou o lucro por unidade de capacidade de produção) é:

$$\Pi_i(t) = p(t)A_i(t) - c_i - r_i m(t) - r_i n(t),$$

onde c é o custo de produção por unidade de capacidade de produção; $r m$ e $r n$ o montante das despesas em P&D por unidade de capital consagrado à imitação e à inovação, respectivamente.

As atividades de P&D geram progresso técnico (elevação da produtividade) por meio de um processo aleatório em duas etapas. Denominam-se por $d_i m(t)$ e $d_i n(t)$ variáveis aleatórias que podem assumir valores 0 ou 1 de acordo com a decisão da empresa i que a conduz a imitar ou a inovar ao longo do período t . As probabilidades de sucesso são proporcionais aos investimentos de P&D.

$$\Pr[d_i m(t) = 1] = a_i m r_i m K(t)$$

Os parâmetros são escolhidos de forma que cada probabilidade não possa ser superior a 1.

$$\Pr[d_i n(t) = 1] = a_i n r_i n K(t)$$

Esse conjunto de pressupostos leva às seguintes possibilidades:

- para $d_i n = d_i m = 0$, tanto os esforços da empresa i em imitar as melhores tecnologias existentes entre os concorrentes como aqueles que visam descobrir novas tecnologias fracassam e, nesse caso, $A_i(t+1) = A_i(t)$;
- para $d_i m(t) = 1$, a empresa tem acesso garantido à melhor tecnologia de produtividade: $\hat{A}_i(t)$;
- para $d_i n(t) = 1$, a empresa tem acesso a uma nova tecnologia que conduz à adoção de uma técnica de produtividade: $\tilde{A}_i(t)$.

Para uma empresa que realiza ao mesmo tempo um processo de imitação e de inovação, seu nível de produtividade no período seguinte é dado por:

$$A_i(t+1) = \text{Max}[A_i(t), \hat{A}_i(t), \tilde{A}_i(t)]$$

$\hat{A}_i(t)$ é a produtividade mais alta possível na indústria, ao longo do período.

$\tilde{A}_i(t)$ é a variável aleatória, tal que:

$\text{Log}[\tilde{A}_i(t)]$ tem uma distribuição normal, $N[\lambda(t)\delta^2]$

com $(\lambda) = a + bt$ ou, dito de outra forma, com o tempo a produtividade média resultante da inovação aumenta (efeito da aprendizagem).

A capacidade de produção da empresa i no período $(t+1)$ é definida pela relação contábil:

$$K_i(t+1) = I_i + (1 - \delta) K_i(t);$$

onde I_i é a função de investimento da empresa i e δ a taxa de depreciação.

A evolução da estrutura de concorrência pode seguir diversas regras. A partir desse modelo podem ser feitas algumas previsões. Se as empresas não fazem qualquer esforço de P&D e têm a mesma produtividade do capital, a dinâmica do modelo torna-se determinista; as N empresas sobreviventes dividem igualmente o mercado. A simulação torna-se necessária assim que a situação é mais complexa, ou seja, quando as empresas têm produtividades diferentes e níveis de custo diferentes. As empresas que têm os níveis de custo mais baixos poderão eliminar do mercado as empresas que têm custos mais altos e dividir entre elas as partes órfãs do mercado.

Suponhamos um processo generalizado de imitação, durante o qual algumas empresas não podem inovar (as despesas de P&D destinadas a inovar são nulas: $r_i n(t) = 0$, para todo i e para todo t). Supõe-se também que todas as empresas investem o mesmo volume de P&D, por unidade de capital, na estratégia de imitação ($r_i m(t) = r$). As empresas que sobrevivem terão a mesma produtividade, igual à produtividade máxima, associada à melhor técnica conhecida inicialmente. Entretanto, o processo de difusão expulsará algumas empresas do mercado, ou seja, algumas empresas sairão da indústria ao longo do processo de difusão. Segundo as especificações retidas quanto à saída do mercado, reinicia-se o jogo com o número de empresas sobreviventes e teremos processos de seleção/difusão completamente distintos. Assim, na ausência de inovações (isto é, de progresso técnico que venha romper o “equilíbrio” da indústria), a estrutura do mercado, que se estabiliza ao final do processo, é determinada pelo comportamento das empresas no que diz respeito à pesquisa de informações tecnológicas (comportamento imitativo) e às modalidades de difusão dos conhecimentos na indústria.

Os fenômenos mais interessantes a serem estudados são as formas de evolução de uma indústria, tomando-se a estrutura do mercado inicial, os comportamentos interativos das empresas e as possibilidades de progresso técnico. Em seus primeiros trabalhos, Nelson e Winter (1982) estabeleceram várias condições iniciais sobre o número de empresas (cada qual definindo uma estrutura de mercado) e sobre o regime de financiamento do investimento das empresas. Em cada uma das simulações a metade das empresas opta por uma estratégia imitativa e a outra metade por uma estratégia mista (inovação e imitação). Um dos resultados previsíveis dessa simulação é a tendência à baixa generalizada dos preços dos produtos.

Outra previsão estabelece que uma estrutura do mercado mais competitiva (caracterizada por um número maior de empresas) produz um desempenho produtivo médio inferior. Em oposição, uma estrutura mais concentrada é capaz de gerar uma produtividade mais elevada. Entretanto, é importante notar que a explicação, nesse caso, não está relacionada com uma maior propensão a inovar do monopólio ou do oligopólio, como supõe Schumpeter e será visto na próxima seção, mas sim a um efeito de inércia: a técnica com melhor desempenho se difunde mais lentamente em uma estrutura de mercado onde várias empresas interagem.

8.4 Intensidade da Inovação: Mudança Tecnológica e Estrutura Industrial

A análise das relações entre estrutura industrial (descrita em termos de repartição do tamanho das empresas, grau de concentração ou outros indicadores apontados nos Capítulos 5 e 7) e a mudança tecnológica durante um longo tempo foram encaminhadas por meio da seguinte questão: seriam algumas estruturas industriais mais favoráveis que outras ao investimento nas atividades de P&D e à sua eficácia?

Essa questão, que dominou a matéria Economia Industrial durante a década de 1960, é originária da denominada hipótese schumpeteriana que associa uma maior intensidade de inovação às grandes empresas. Em oposição, o quadro de análise neoclássica, como visto anteriormente, considera que o regime de concorrência, ou seja, em indústrias nas quais predominam as pequenas e médias empresas, promove mais adequadamente a inovação tecnológica, como pudemos inferir das previsões do modelo de incitação anterior. Além disso, algumas vantagens estruturais são apontadas em favor da inovação nas pequenas e médias empresas: organização menos burocratizada, maior motivação para a criatividade e maior disposição para dividir o mercado.

Os diversos estudos empíricos realizados a partir de 1960 traduziram essa observação de Schumpeter em duas proposições independentes: (1) a inovação cresce mais que proporcionalmente com o tamanho da empresa; (2) a inovação cresce com a concentração do mercado.

Várias hipóteses foram utilizadas nesses estudos empíricos para justificar essa proposição do efeito positivo do tamanho da empresa ou do grau de concentração do mercado na atividade de inovação. Entre as principais destacam-se: as imperfeições do mercado de capitais, que conferem vantagens para as grandes empresas por permitirem acesso mais fácil a financiamentos para os projetos de P&D, ou o fato de que as grandes empresas dispõem de recursos próprios; a existência de economias de escala na tecnologia, decorrentes da indivisibilidade de alguns equipamentos de P&D; os elevados custos fixos da inovação que podem ser compensados quando o inovador pode dividir por um maior volume de vendas; a complementaridade com outros ativos nas grandes empresas que permite aumentar a produtividade das atividades de P&D; e, finalmente, as empresas maiores e mais diversificadas estão mais bem posicionadas para explorar os resultados incertos das atividades de P&D pelo fato de atuarem em um escopo mais amplo de mercados.

Os contra-argumentos mais utilizados para negar a existência da mesma hipótese, por sua vez, referem-se às deseconomias de escala que a empresa estaria sujeita, provocadas entre outros fatores por perda de controle gerencial e decorrente redução da eficiência das atividades de P&D, bem como pela perda de incentivo ao espírito empreendedor, responsável pela habilidade em capturar benefícios de seus esforços, em função do aumento da burocratização dessas atividades.

Apesar da literatura empírica disponível sobre o teste dessas hipóteses ser bastante ampla, os resultados obtidos são relativamente inconclusivos e divergentes, como veremos na próxima subseção. Antes, porém, vejamos os tipos de medidas utilizadas para o teste dessas hipóteses e os problemas que podem derivar do uso dessas medidas.

8.4.1 MEDIDAS DE ENTRADA, SAÍDA E RESULTADOS

Tomando-se as atividades de P&D como um sistema organizado e rotineiro que transforma recursos (entradas ou meios da pesquisa) em novos conhecimentos (saídas ou resultados da pesquisa) que, por sua vez, serão incorporados em novos produtos, processos e formas de organização, impactando a atividade econômica (resultados ou impactos da pesquisa sobre o sistema produtivo), podemos visualizar, na Figura 8.2, o modelo conceitual e os diferentes tipos de indicadores utilizados para inferir os esforços de P&D e sua eficácia.

Esse conjunto de indicadores resultou de um esforço de padronização realizado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) que está reunida no *Manual Frascati*. Entre os indicadores mais utilizados na literatura empírica destacam-se como indicadores de entrada o número de pessoas empregadas em P&D, ou o seu equivalente em horas trabalhadas nessas atividades, e as despesas de P&D realizadas, que incluem as despesas com as pessoas empregadas, as despesas com material de consumo e as despesas com depreciação dos equipamentos laboratoriais, entre outras.

Entre os indicadores de saída, destaca-se o número de patentes e entre os indicadores de resultado empresarial, o percentual de receitas geradas atribuíveis aos novos produtos nos últimos cinco anos.

O modelo conceitual e os indicadores anteriormente enunciados, entretanto, apresentam algumas limitações, principalmente decorrentes do fato de que o *Manual Frascati* apenas registra as atividades de P&D que são realizadas de forma organizada e contínua. Dessa forma, por exemplo, podemos adiantar os seguintes problemas nos estudos empíricos realizados.

1. Incluir entre as despesas somente as despesas realizadas com as atividades de P&D subestima a intensidade dessas despesas, que podem estar ocorrendo nos processos de aprendizagem decorrentes do uso de equipamentos, gerados em outra indústria, por exemplo.
2. A mudança tecnológica pode ser decorrente de outras fontes que não as atividades de P&D da empresa, como, por exemplo, atividades de pesquisa desenvolvidas nas universidades.
3. Caracterizar os resultados da pesquisa pelo número de patentes indica melhorar a propensão a inovar do que a propensão a investir; além disso existe uma série de maneiras pelas quais as tecnologias podem se transformar em ativos rentáveis além do patenteamento; em alguns casos, por exemplo, o segredo industrial é a maneira mais eficiente de apropriação; outro importante aspecto é que nem toda patente se transforma em um novo produto ou processo.

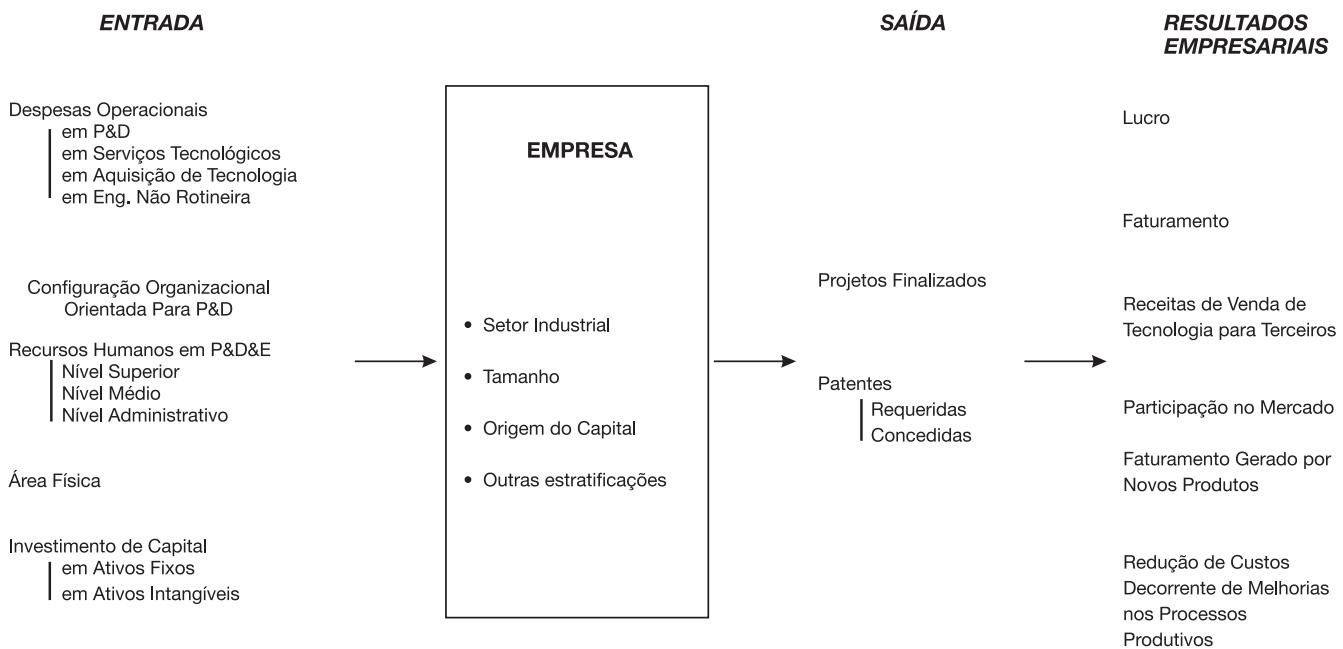


Figura 8.2 Modelo Conceitual sobre Indicadores Empresariais de Inovação Tecnológica. *Fonte:* Adaptado de Andreassi (1999) e ANPEI (1998).

Essa insatisfação com os indicadores de P&D está no centro da discussão de uma nova série de indicadores de ciência, tecnologia e inovação decorrente da adoção de uma visão menos linear e mais sistêmica do processo de inovação. Os novos indicadores incluídos são relativos às relações que se estabelecem entre as atividades de P&D, produção e *marketing* no interior da empresa e indicam que o processo de inovação não se esgota no departamento de P&D, mas transborda dentro da empresa e para além dela na relação com fornecedores e usuários. Estão relacionados com o processo de difusão da inovação e representam melhorias introduzidas ao longo do processo de produção, inclusive os desenvolvimentos conjuntos com fornecedores (novas máquinas e equipamentos, novos materiais, novos métodos de organização da produção, novos *designs* relacionados com aspectos funcionais e técnicos, entre outros) e melhorias introduzidas na relação com os usuários (*marketing*, logística de distribuição, *design* estético, entre outros), conforme o modelo de Kline-Rosenberg, que será visto no Capítulo 19.

Por essas razões que foi concebido um novo Manual, o Manual de Oslo (1992) para orientar as pesquisas nacionais sobre inovação. Essas pesquisas têm sido realizadas na OCDE, mas também em países em desenvolvimento, como o Brasil, por intermédio da Pesquisa de Inovação Tecnológica (Pintec), apresentada no Capítulo 28.

8.4.2 EVIDÊNCIAS EMPÍRICAS

A maioria dos estudos empíricos realizados utiliza, como visto anteriormente, as medidas de entrada. Isso se deve à maior facilidade de obtenção desses dados, uma vez que, em geral, fazem parte dos registros contábeis das empresas. Um resumo desses estudos e de seus resultados pode ser apreciado no Quadro 8.1.

Uma característica marcante do conjunto da pesquisa empírica sobre a relação entre tamanho da empresa e inovação é que existem várias categorias de resultados que podem ser identificadas no Quadro 8.1. Em uma primeira categoria observou-se que a intensidade de P&D cresce fracamente com o tamanho da empresa. Na segunda observou-se que a relação entre as variáveis é mais sutil, a atividade inovativa cresce mais do que proporcionalmente com o tamanho da empresa até certo limite, a partir do qual a relação pode ser fracamente negativa ou não existir. Encontram-se também resultados que indicam alguma evidência positiva em favor da hipótese de uma relação positiva entre as variáveis, mas, em alguns casos, ficou restrito a apenas para alguns setores, como é o caso da indústria química nos estudos de Mansfield (1964) e Grabowski (1968). Em contraposição, a quinta categoria mostra resultados que indicam que a intensidade de P&D diminui com o tamanho da empresa. E, por fim, de acordo com uma última categoria de resultados apresentados, as empresas muito pequenas ou muito grandes são as mais intensivas em P&D.

Na conhecida e divulgada resenha realizada por Cohen e Levin (1989) sobre inovação e estrutura de mercado é possível encontrar também estudos que utilizaram uma medida de resultado para analisar a relação entre tamanho de empresa e inovação e, até mesmo nesse caso, não se têm respostas conclusivas. Os autores citam os estudos empíricos realizados por Gellman

QUADRO 8.1 INOVAÇÃO E TAMANHO DA EMPRESA

1. Intensidade de P&D cresce fracamente com o tamanho da firma
 - Horowitz (1962)
 - Comanor (1967)
 - Hamberg (1964)
 - Worley (1961)
2. Atividade inovativa cresce mais do que proporcionalmente com o tamanho da firma até um certo limite, a partir do qual a relação pode ser fracamente negativa ou não existir
 - Scherer (1965)
 - Malecki (1980)
 - Loeb e Lin (1977)
 - Phillip (1971)
 - Link (1981)
 - Cohen (1987)*
3. Evidência positiva de uma relação monotônica entre tamanho e P&D
 - Soete (1979)
 - Meisel e Lin (1983)
 - Link (1980)
 - Loeb (1983)
4. Relação positiva apenas para alguns setores industriais, particularmente o caso da indústria química
 - Mansfield (1964)
 - Grabowsky (1968)
5. Relação negativa entre tamanho e intensidade de P&D
 - Smith e Creamer (1968)
 - Mueller (1967)
 - Shrieves (1976)
6. Intensidade de P&D primeiro diminui e depois aumenta com o tamanho da firma
 - Bound et al. (1984)
 - Cremer e Sirbu (1978)

* Esse autor analisa o tamanho da linha do negócio e não o tamanho da empresa como um todo.

Fonte: Cohen e Levin (1989) e Kamien e Schwartz (1982)

Research Associates (1976) e Bound *et al.* (1984), onde os resultados foram contrários à hipótese schumpeteriana: as empresas menores tiveram maior participação na introdução das principais inovações, ao passo que em Pavitt (1983) as maiores empresas responderam pelas maiores proporções de inovações.

Os estudos empíricos que testam a segunda hipótese atribuída a Schumpeter sobre a relação entre inovação e estrutura de mercado estão resumidos no Quadro 8.2. Da mesma forma que no Quadro 8.1, observamos pouco consenso quanto aos

QUADRO 8.2 INOVAÇÃO E ESTRUTURA DE MERCADO

1. Intensidade P&D é positivamente relacionada à concentração de mercado
 - Horowitz (1962)
 - Kelly (1970)
 - Hamberg (1964)
 - Shrieves (1978)
 - Mansfield (1968)
 - Rosemberg (1976)
2. Concentração de mercado tem efeito negativo sobre P&D
 - Bozeman e Link (1983)
 - Williamson (1965)
 - Mukhopadhyay (1985)
3. Evidência de uma relação não linear “U-invertido” entre intensidade de P&D e concentração de mercado
 - Scherer (1967)
 - Scott (1984)
 - Levin et al. (1985)

Fonte: Cohen e Levin (1989) e Kamien e Schwartz (1982)

resultados. Apesar de muitos estudos encontrarem uma relação positiva entre concentração de mercado e P&D, encontrou-se também evidência de que a concentração tem efeito negativo sobre a P&D. Em uma terceira categoria de resultados identificou-se ainda uma relação não linear (U-invertido) entre as variáveis, ou seja, a intensidade de P&D aumenta com a concentração de mercado até certo limite e a partir daí começa a declinar.

Apesar dos resultados divergentes, esses estudos concordam que a relação entre inovação e concentração depende de características específicas de cada indústria. Os estudos da última categoria do Quadro 8.2, por exemplo, forneceram evidência forte de que os resultados sobre o efeito da concentração na inovação são sensíveis às condições específicas de cada indústria. Isso porque cada indústria possui características específicas, como a cumulatividade da mudança tecnológica, as oportunidades tecnológicas e a apropiabilidade dos efeitos da mudança tecnológica, características estas que compõem o regime tecnológico de cada indústria e geram incentivos diferentes para inovar.

8.5 Conclusão

Esses resultados inconclusivos da literatura talvez reflitam as dificuldades a que estão sujeitos os testes empíricos das chamadas hipóteses schumpeterianas. A dificuldade mais fundamental talvez seja identificar uma inovação, que, em princípio, pode ser considerada como um deslocamento para cima da função produção ou um novo produto, mas é extremamente difícil identificar tais fatos na prática. Encontrar a definição apropriada para a medida de entrada e de saída do processo inovativo, assim como para o tamanho da empresa, é outro problema.

Outro aspecto seria a consideração de outras variáveis do esforço tecnológico da empresa, tais como atividades de aprendizado que se dão no processo produtivo e não nas atividades de P&D, e também o entorno institucional em que a empresa está situada e que também contribui para a inovação. É importante, ainda, considerar dados setoriais devido às peculiaridades sobre as mudanças tecnológicas específicas a cada setor. Ao examinar a relação entre tamanho da empresa e inovação, ou estrutura de mercado e inovação, é necessário considerar ainda as oportunidades tecnológicas de cada indústria, assim como as oportunidades de negócio. Assim, percebe-se que uma avaliação mais apropriada das hipóteses schumpeterianas deve considerar um modelo mais completo de determinação de progresso tecnológico.

É importante destacar ainda que a grande dificuldade de sustentar inequivocamente uma ou outra tese decorre da inexistência de indicadores confiáveis de inovação e de um consenso sobre a definição de estrutura industrial, como discutido na subseção anterior.

Mais recentemente, essa análise empírica foi renovada para incluir a natureza das influências recíprocas das estruturas sobre a taxa e a direção da mudança tecnológica, como veremos na Parte V sobre estratégias empresariais. Ou seja, se certas estruturas determinam as transformações tecnológicas, as estruturas também se modificam, por sua vez, em função das estratégias de inovação adotadas pelas empresas e as diversas formas de concorrência existentes.

Notas

1. Ver a respeito das causas e especificidades dessa mudança no texto de Freeman e Ivete (2005).
2. O termo significa que a P&D de uma empresa *i* qualquer serve a outra empresa *j* qualquer para diminuir o seu custo unitário. Em outras palavras, o investimento em conhecimento de uma empresa transborda para a outra sem que esta tenha custos.
3. Este modelo será mais bem compreendido após a leitura dos Capítulos 18 e 19.

Bibliografia

- Andreassi, T. (1999) Estudo das relações entre indicadores de P&D e indicadores de resultado empresarial em empresas brasileiras. Tese de Doutorado. Departamento de Administração/USP.
- ANPEI. Indicadores empresariais de inovação tecnológica: resultados da base de dados da ANPEI. Ano 7, dez., 1998.
- Bound, J.; Cummins, C.; Griliches, Z. B. H.; Jaffle, A. (1984) "Who does R&D and who patens?". In: Griliches, Z. (org.) *R&D, Patents and Productivity*. Chicago: University of Chicago Press for the National Bureau of Economic Research.
- Cohen, W. M.; Levin, R. C. (1989) "Empirical studies of innovation and market structure". In: Schmalensee, R.; Willig, R. D. *Handbook of Industrial Organization*, pp.1060-1107.
- Davies, S.; Lyons, B. (1988) *Economics of Industrial Organizations. Survey in Economics*. Londres: Longman.
- Freeman, C.; Soete, L. (2009) "Developing science, technology and innovation indicators: what we can learn from the past". *Research Policy*, v. 38, pp. 583-589.
- Gellman Research Associates. *Indicators of International Trends in Technological Innovation*. Final report to the National Science Foundation, NTIS document PB-263-738, Jenkintown, Penn: Gellman Research Associates, 1976.

- Grabowski, H. G. (1968) "The determinants of industrial research and development: a study of chemical, drug and petroleum industries". *Journal of Political Economy.*, 76, pp. 292-306.
- Hall, P. (1994) *Innovation, Economics & Evolution. Theoretical Perspectives on Changing Technology in Economic Systems*. Nova York: Harvester Wheats-heaf.
- Kamien, M.; Schwartz, N. (1982) *Empirical Studies of the Schumpeterian Hypotheses. Market Structure and Innovation*. Cambridge University Press.
- Le Bas, C. (1995) *Economie de l'Innovation*. Paris: Ed. Economica.
- Mansfield, E. "Industrial research and development expenditures: determinants, prospects, and relation of size of firm and inventive output". *Journal of Political Economy*, v. 72, p. 319-340, 1964.
- OECD (1997) *Manual de Oslo 1992*. Edição brasileira. Rio de Janeiro: FINEP.
- OECD (2002) *Manual de Frascati 1963*. Edição espanhola. Madri: FECYT.
- Resende, M.; Hasenclever, L. (1998) "Intensidade em P&D e Tamanho da Firma: uma Análise Exploratória do Caso Brasileiro". *Estudos Econômicos*, SP, v. 28, n.4, out.-dez., p. 1-18.
- Schumpeter, J. A. (1942) *Capitalismo, Socialismo e Democracia*. Zahar Editores, edição brasileira 1984.

PARTE

|||

INTERAÇÃO ESTRATÉGICA

Uma Introdução à Teoria dos Jogos

Larry C. Cardoso e Luís Otávio Façanha

9.1 Introdução

A teoria dos jogos pode ser definida, em princípio, como conjunto de técnicas de análise de situações de interdependência estratégica, já possuindo história, tradição e prestígio consolidados no âmbito da matemática. Parte importante daquele prestígio, entretanto, se deve à mobilização dos recursos oferecidos pela teoria dos jogos para utilização nas disciplinas das ciências sociais e da biologia, nas quais interações e conflitos entre atores e agentes constituem problemas tão reconhecidamente centrais quanto propícios para serem explorados e analisados com o emprego da teoria dos jogos. Nos dois sentidos, há, pelo menos, cinco autores principais a referenciar e prestigiar. Em primeiro lugar, J. von Neumann e O. Morgenstern (vN&M), autores do famoso *Theory of Games and Economic Behavior*, publicado em 1944, e, adicionalmente, J. Harsanyi, H. Selten, e J. Nash, laureados em 1994 com o Prêmio Nobel de Economia pelo conjunto de seus trabalhos.

De fato, desde a publicação do livro de vN&M, a penetração da teoria dos jogos em vários ramos das ciências aplicadas tem sido muito expressiva e, de um relativo aprisionamento inicial na esfera de trabalho de matemáticos e no estudo de jogos de soma-zero, a teoria dos jogos difundiu-se. Nos dias de hoje, a terminologia e as noções de jogos de soma-zero, de jogos cooperativos, de jogos não cooperativos e mesmo dos jogos evolucionários passaram a fazer parte da linguagem corrente de economistas, cientistas políticos, administradores, biólogos e jornalistas e a ser objeto de interesses comprometidos com o domínio daquelas técnicas.

Este capítulo tem objetivo pedagógico, tira partido do aprendizado adquirido com excelentes livros-texto já disponíveis e pretende ser de utilidade para os que se interessam por quaisquer das ramificações apontadas da teoria dos jogos. Em particular, espera-se que a teoria dos jogos venha a ser valorizada pelo leitor como instrumento de identificação, descrição e análise de regras de jogos e de conflitos, propício, portanto, para a organização e estruturação de diagnósticos e teorias, para o qual, como se verá detalhes – de processos de competição e de questões organizacionais, em especial – talvez não devam ser tratados como meros detalhes. Abordaremos as situações de interdependência estratégica por meio dos jogos não cooperativos, sem contemplar jogos com informação incompleta, que são situações nas quais características e regras do jogo não são conhecimento comum dos jogadores.

Além desta Introdução, o capítulo contém quatro outras seções e um Apêndice. Na Seção 9.2, apresentam-se elementos e conceitos que exploram virtudes descritivas da teoria dos jogos e a noção de regras de jogos. A análise de jogos e o estudo de conceitos de equilíbrio têm início na Seção 9.3, que se dedica aos jogos estáticos e de decisões simultâneas. A Seção 9.4 estuda os jogos dinâmicos com informação completa, e a Seção 9.5 oferece comentários finais, inclusive a respeito da bibliografia. Em apêndice, o leitor encontrará breve discussão do conceito de estratégias mistas e do conceito de equilíbrio de Nash em estratégias mistas.

9.2 Jogos de Estratégias: Representações e Explicitação de Regras de Jogos

O objetivo desta seção inicial é familiarizar o leitor com as representações de jogos e de regras de jogos. Como já foi assinalado na Introdução, pode ser útil recorrer à dimensão descritiva da teoria dos jogos, antes de passar à análise dos jogos e aos conceitos de solução, uma vez que esses conceitos estão intimamente associados às regras dos jogos e às situações de conflito que os jogos representam.

Neste texto, discutiremos jogos não cooperativos com informação completa. Nesses jogos, a interação estratégica não contempla manipulações de informação (por exemplo, a ocultação intencional dos custos por parte de uma empresa). Os jogos são jogos de informação completa e perfeita quando os jogadores podem observar as ações escolhidas pelo oponente. Quando isso não ocorre, dizemos que os jogos são jogos de informação imperfeita. Em todos os casos, podemos caracterizar os jogos por meio da forma extensiva (também denominada de forma dinâmica) ou da forma estratégica (às vezes denominada de forma normal). A forma estratégica é a representação mais popularizada de jogos, mas a subseção que se segue apresenta os jogos com informação completa na forma extensiva, uma vez que essa representação fornece descrição mais rica e completa dos jogos.

9.2.1 REPRESENTAÇÃO EXTENSIVA DE JOGOS COM INFORMAÇÃO COMPLETA

Consideremos as decisões interdependentes de duas empresas quanto a investimentos a realizar na capacitação tecnológica e nas atividades de Pesquisa e Desenvolvimento (P&D). De forma realista, os montantes a serem investidos não podem ser avaliados com precisão, distinguindo-se apenas como investimentos elevados (E), ou investimentos menores (B), por comparação com alguma referência histórica. Sabe-se que a empresa I sempre toma suas decisões antes da empresa II, após o que a empresa II observa a escolha da empresa I, decidindo a seguir entre E ou B ou entre e ou b. Após a decisão da empresa II, a empresa I e a empresa II auferem ganhos finais, indicados no gráfico que se segue pelos números em parênteses (o primeiro número indica o ganho da empresa I e o segundo o ganho da empresa II).

Impossibilidades de avaliação quantitativa do potencial de mercado, taxas internas de retorno e outros critérios de seleção de oportunidades de investimento e decisão permitem apenas às firmas que ordenem as situações tidas como possíveis, de acordo com suas preferências. Para a empresa I, o melhor dos desenlaces possíveis seria aquele em que seus investimentos pudessem ser baixos, pois isto implica menos esforço e menores custos organizacionais, com investimentos também baixos da empresa II (daí o ganho final de 4 para a empresa I). O segundo melhor desenlace para a empresa I caracteriza-se pela escolha de E pela empresa I e pela escolha de b pela empresa II (daí o pagamento de 3 para a empresa I). Acontece que, diante de investimentos menores da empresa I, a empresa II também pode realizar investimentos elevados, e esta é a situação que a empresa II prefere a qualquer outra (daí o ganho final de 4 para a empresa II). Na ordem das preferências da empresa I, este seria o terceiro melhor desenlace do jogo (daí o ganho final de 2 para a empresa I). A empresa II também prefere investir pouco quando a empresa I investe pouco, a investir pouco quando a empresa I investe muito. Quando as duas empresas investem muito, configura-se a pior das situações para ambas, o que está indicado por (1,1). Com isto em mente, pode-se recorrer à Figura 9.1, que traduz o problema em um “diagrama de árvore” simples.

A árvore é constituída de ramificações, cada uma das quais conecta dois pontos, chamados “nós” ou “nodos”. O nó que se encontra mais à esquerda é denominado de “raiz” da árvore, e representa o início do jogo. Existem nodos sem ramificações

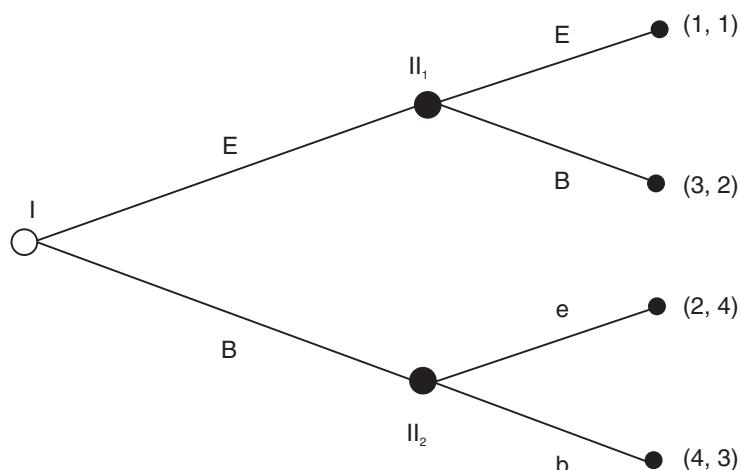


Figura 9.1 A Forma Extensiva de um Jogo com Informação Completa e Perfeita.

(quatro no caso), que recebem, por isso, a denominação de “nós terminais”. As sequências de eventos e as ramificações configuram “caminho de jogadas” e maneiras possíveis do jogo ser jogado, tendo como finalizações os nós terminais. Após cada nó terminal, a Figura 9.1 mostra a dupla de números que denota os ganhos finais (*payoffs*).

Como representação do jogo, a árvore explicita incertezas – associadas às interdependências das ações possíveis de serem adotadas pelos jogadores – que antecedem os ganhos finais. Na Figura 9.1 está representado um jogo com *informação completa e perfeita*, pois as ações do jogador I são sempre observadas pelo jogador II antes do jogador II tomar suas decisões. O jogador sabe que está no nó e conjunto de informação II.1, ou no nó e conjunto de informação II.2 quando é a sua vez de jogar. Uma incerteza adicional a ser contemplada a seguir é a que decorre de impossibilidades dos jogadores observarem as decisões tomadas pelo outro jogador, diferentemente do que ocorre no jogo da Figura 9.1.

Tal como acontece em leilões e licitações, em que os competidores (sem comunicação prévia entre si) apresentam suas propostas em envelopes fechados, nos *jogos com informação completa e imperfeita*¹ representa-se dificuldades de comunicação entre os jogadores. No caso, quando a empresa II não pode observar as decisões tomadas pela empresa I, o jogo passa a ser descrito como a seguir, na Figura 9.2, situação em que a empresa I também terá dificuldades de prever as decisões tomadas pela empresa II (ponto que será mais elaborado quando estudarmos os conceitos de solução).

Agora, os dois nós de decisão pertencentes à empresa II não podem mais ser distinguidos (como indica a linha tracejada) e passam a pertencer ao mesmo conjunto de informações. Vale dizer, na sua vez de jogar a empresa II (e a empresa I, pois a situação é simétrica) não sabe se está no nó de decisão superior (que foi anteriormente indicado por II.1), ou inferior (antes indicado por II.2). Omitiremos os ganhos (que suporemos serem os mesmos do caso anterior) da Figura 9.2, para tornar mais nítida a apreensão visual das novas regras do jogo.

Com esses elementos, podemos agora elaborar o conceito de estratégia. Para um jogador determinado, uma *estratégia* é a regra de conduta que especifica um movimento e uma escolha de ramificação para cada um dos *conjuntos de informação* do jogador. No jogo com informação perfeita, o jogador e a empresa I possuem apenas um conjunto de informação, a partir do qual pode selecionar E ou B. Para o jogador e empresa II, uma possível estratégia deve especificar o que fazer quando estiver na posição II.1 (escolher E ou B), e o que fazer se for chamado a jogar na posição e nó II.2 (escolher e ou b). No jogo com informação imperfeita, especificamos que os dois jogadores decidem entre E ou B apenas, pois dos nós que pertencem a um único conjunto de informação devem se suceder as mesmas alternativas de escolha (a descrição do jogo está levando em conta o fato de o jogador II não saber em qual dos nós do mesmo conjunto de informação ele se encontra). Em um sentido mais amplo, é importante reconhecer que os jogadores jogam com objetivos (condicionais) múltiplos, e a partir de *conjuntos de estratégias* que, nas situações hipotéticas comentadas, podem ser especificadas como se segue:

Conjuntos de Estratégias

	Quadro I	Quadro II
Empresa I	{ E,B }	{ E,B }
Empresa II	{ Ee, Eb, Be, Bb }	{ E,B }

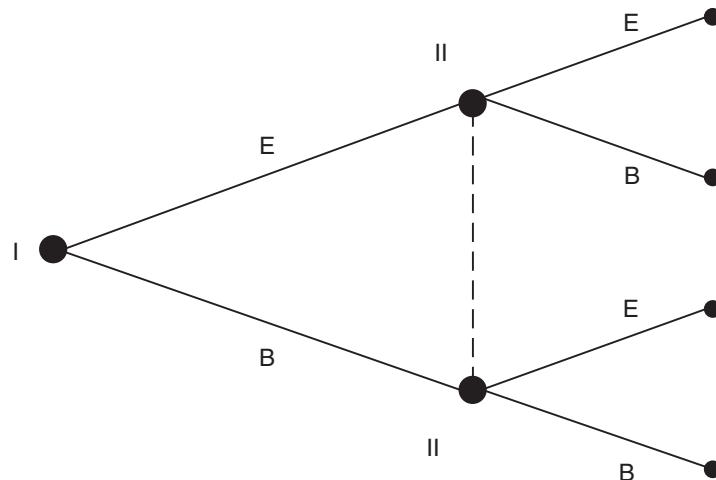


Figura 9.2 A Forma Extensiva de um Jogo com Informação Completa e Imperfeita.

	C_{II}	Ee	Eb	Be	Bb
C_I					
E		1,1	1,1	3,2	3,2
B		2,4	4,3	2,4	4,3

	C_{II}	E	B
C_I			
E		1,1	3,2
B		2,4	4,3

Figura 9.3 Representações Normais de Formas Estratégicas.

9.2.2 A REPRESENTAÇÃO NORMAL E FORMA ESTRATÉGICA DE JOGOS

Para definirmos um jogo em forma estratégica é necessário apenas especificar o conjunto de jogadores, o conjunto de opções e estratégias disponíveis para cada jogador, e os ganhos possíveis associados a cada jogada e escolha possível de estratégias. As formas estratégicas mostradas a seguir correspondem aos jogos representados por suas formas extensivas nas Figuras 9.1 e 9.2, respectivamente, à esquerda e à direita da Figura 9.3.

Deve-se desde já adiantar que o recurso à forma estratégica pode se tornar tão trabalhoso quanto pouco promissor, observação que não justifica abdicar de seu uso. Entretanto, mais importante do que isso é reconhecer que a *forma estratégica* é um modelo estático de jogo, pois ignora questões temporais e trata o jogo como se os jogadores estivessem escolhendo suas estratégias simultaneamente (mesmo que haja hiato de tempo entre a tomada de decisão de um e do outro). Diferentemente, a *forma extensiva* é um modelo dinâmico de jogo, pois leva em conta e descreve de forma completa as sequências de movimentos e eventos possíveis de ocorrer ao longo do tempo, quando o jogo for jogado.

A generalização do que foi exposto anteriormente se faz necessária e as notações que se seguem podem ser oportunas, justificadas e necessárias, uma vez que serão úteis em desenvolvimentos posteriores. Em primeiro lugar, os jogadores constituirão um conjunto J de J elementos, com $j = 1, \dots, J$. Cada jogador j tem conjunto Cs_j de m estratégias, e $Cs_j = \{s_{j1}, \dots, s_{ji}, \dots, s_{jm}\}$. Por sua vez, um perfil de estratégias será representado pelo conjunto $S = \{s_1, \dots, s_J\}$, que dirá respeito à combinação de estratégias que define uma célula na forma normal e estratégica dos jogos. Ou seja, S é extraído do conjunto Cs , com cada jogador genérico extraíndo uma estratégia específica do seu conjunto de estratégias. Além disso, como a busca de equilíbrios obrigará a que se foque nas decisões de cada jogador, é conveniente representar a estratégia do jogador em questão por s_j e denotar o perfil de estratégias dos demais jogadores por s_{-j} . Finalmente, se dirá (por ora) que os jogos estão também descritos por coleção de funções de pagamento $U = \{u_1(\cdot), \dots, u_J(\cdot)\}$ que atribuem utilidades aos jogadores em cada nó terminal que pode ser alcançado, e em cada célula da forma estratégica ou normal do jogo.

9.3 Conceitos de Solução: Análise de Jogos Estáticos com Decisões Simultâneas

A partir das observações da seção anterior, cabe agora indagar acerca de como elaborar, intuir e prever os desenlaces possíveis de jogos. O leitor terá acesso a conceitos de soluções para jogos e ao conceito de equilíbrio de Nash – EN. Este conceito é inseparável da teoria dos jogos e, muito frequentemente, torna-se fonte de mal-entendidos.

O equilíbrio, quando encontrado, não deve ser interpretado como imposição normativa, do tipo “o jogo deve ter este ou aquele desenlace” e/ou do tipo “o EN configura acordo autoimposto”, como se verá logo adiante. Acordos que se pretendem autoimpositivos devem possuir características adicionais de incentivo e/ou constituir referências virtuosas de convenções ou pontos focais, para que sejam tácitos, e não compulsoriamente adotados pelos jogadores. Se pretendermos supor a existência de acordos (que tenham a força de um equilíbrio), é recomendável que antes descrevamos e analisemos o jogo que permitiu a emergência do acordo.

Convém assinalar que a noção de equilíbrio não tem diretamente a ver com ganhos finais, e Kreps, persuasivamente numa das mais conhecidas maneiras de se referir ao desenlace de um jogo, referiu-se ao equilíbrio como “maneira razoável do jogo vir a

ser jogado”, no sentido de que, ao compartilharem a crença de que existe maneira óbvia de o jogo ser jogado, então os jogadores alcançariam o EN. Mas essa sugestão não deve levar o leitor a inferir que os desenlaces dos jogos venham a ser inequívocos. Em particular, os jogos podem não apresentar equilíbrio (quando jogados por meio de estratégias puras – veja o Apêndice), podem ter equilíbrio único e podem ter equilíbrios múltiplos. Ou seja, o EN é um conjunto que pode ser vazio, pode ser composto de elemento único e pode ser composto de mais de um equilíbrio. No caso de equilíbrios múltiplos, a “maneira razoável” de jogar contemplaria a identificação dos equilíbrios e a seleção do equilíbrio final (problema que será ilustrado melhor logo a seguir).

Isso sugere que, para alcançar o EN, a teoria dos jogos supõe que os jogadores estariam selecionando e depurando o conjunto de estratégias disponíveis. Duas *maneiras razoáveis* dos jogos serem jogados serão brevemente expostas a seguir. No primeiro método, os jogadores eliminam e descartam estratégias que não resistem a critérios de dominância, constituindo-se de método muito criticado por sobrecarregar, como se verá, a suposta racionalidade dos jogadores (e a confiança depositada pelos jogadores na racionalidade dos oponentes, o que pode ser grave, especialmente quando as estratégias forem numerosas e as eliminações supostas forem muitas). No segundo método, as exigências de racionalidade atuam menos diretamente. Os jogadores descartam estratégias disponíveis que não constituem melhores respostas a escolhas de estratégias que esses jogadores conjecturam que venham a ser adotadas pelos oponentes. As estratégias que sobrevivem a esse processo são denominadas de *estratégias racionais*, sendo reconhecido que se trata de método mais poderoso de refinamento dos conjuntos de estratégias dos jogadores.

9.3.1 EQUILÍBRIO DE NASH – EN

Nesta subseção, o EN será formalmente apresentado, mas, antes disso, se discutirão dois métodos de solução que ajudam a explicitar os pressupostos de racionalidade a serem adotados, assim como o alcance deles.

9.3.1.1 A Eliminação de Estratégias Dominadas

Considere-se a representação normal de um jogo em forma estratégica, tal como na Figura 9.4, em que estão explicitados apenas os ganhos do jogador I, associados às estratégias s_1 e s_2 do jogador I, e t_1 e t_2 do jogador II. Como segue:

	II	t_1	t_2
I		x	y
s_1		z	w
s_2			

Figura 9.4 Representação Normal de um Jogo em Forma Estratégica.

Diz-se que a estratégia s_1 é fortemente (e/ou estritamente) dominante (e s_2 é fortemente e/ou estritamente dominada) quando for constatado que $x > z$ e $y > w$. Caso se constate que $x = z$ (com $y > w$) ou que $y = w$ (com $x > z$) diz-se então que s_1 é fracamente dominante e s_2 é fracamente dominada.²

O importante a reter, no caso, é que os ganhos associados a dada estratégia do jogador j deve ser cotejado com o conjunto de ganhos associados às suas demais estratégias. Reconsideremos, agora, a representação normal e estratégica do jogo da Figura 9.3 (b), reproduzido na Figura 9.5, sem levar em conta, no momento, as barras colocadas sobre e sob alguns dos números e pagamentos.

Por comparação entre linhas, pode-se observar que a *estratégia (pura)* E é forte e estritamente dominada para a empresa e jogador I: seja qual for a escolha da empresa/jogador II, a empresa I obtém pagamentos maiores escolhendo B (com probabilidade igual a 1) do que escolhendo E (com probabilidade igual a 1). A noção de dominância tem as implicações seguintes.

Podemos dizer que, dificilmente, a estratégia dominada E viria a ser escolhida pelo jogador I, fato que é também do conhecimento do jogador II, e que permite em princípio suprimir a estratégia E da representação do jogo, sem alterar as regras do jogo. Com isso feito, a representação estratégica do jogo passa a ser a da Figura 9.5(b). Diante dessa versão simplificada do jogo, é esperado que o jogador II preferirá a estratégia E à estratégia B, caso o jogador II fosse também “racional”, como o jogador I, e no sentido elementar em que a noção de racionalidade foi utilizada. Assim sendo, outra representação estratégica poderia ser apresentada, incluindo apenas as estratégias B e E, para os jogadores I e II, respectivamente, e o *equilíbrio do jogo* estaria definido por (B,E).

		(a)	
		C _I	C _{II}
C _I	E	1,1	3,2
	B	2,4	4,3

		(b)	
		C _I	C _{II}
C _I	E	1,1	3,2
	B	2,4	4,3

Figura 9.5 Representações Estratégicas de Jogos, Completa (a) e com Estratégias Dominadas Suprimidas (b).

A eliminação de estratégias pelo critério de dominância é método potencialmente eficaz de simplificação, mas a teoria dos jogos recomenda cautela no seu uso, especialmente quando as eliminações ignoram a forma extensiva do jogo (ponto a ser mais esclarecido a seguir), ou quando muitas eliminações são executadas, como já foi assinalado. Neste último caso, o analista estaria impondo aos jogadores que refizessem por muitos turnos suas confianças na racionalidade do outro jogador, o que deve ser evitado, mas não a noção de dominância. Mais importante, talvez, é chamar a atenção para o uso indevido do método quando há identificação de estratégias fracamente dominadas.

Melhores Respostas (best-responses)

Tendo avaliado o alcance possível do método de eliminação de estratégias pelo critério da dominância, pode ser do interesse do leitor dispor de outro método, considerado mais poderoso e confiável, que possa vir a ser também atribuído aos jogadores. Agora, suporemos que os jogadores eliminarão estratégias que jamais serão as melhores respostas às suas conjecturas acerca das estratégias a serem adotadas pelos demais jogadores.

Exploratoriamente, e com o auxílio do jogo explicitado na Figura 9.5(a), pode-se dizer que o jogador I (com estratégias representadas nas linhas) prefere adotar a estratégia pura B quando conjectura que o jogador II (com estratégias puras representadas nas colunas) adotará a estratégia E. A estratégia B propicia ao jogador I ganho maior do que E, o que recomenda *denominar a estratégia B de melhor resposta do jogador I à conjectura de que o jogador II jogará a estratégia E*; e assim por diante, devendo-se fazer o mesmo para a conjectura do jogador I de que o jogador II jogará B e permitir ao jogador II que adote o mesmo procedimento.

As estratégias que resistem e sobrevivem a esse procedimento são denominadas de *estratégias racionalizáveis*. Pode-se argumentar que esse método de eliminação está fortemente apoiado nos pressupostos de racionalidade individual e de conhecimento comum das regras do jogo e das racionalidades mútuas; as estratégias remanescentes foram selecionadas a partir de conjectura justificável (e/ou racionalizável) que os jogadores sustentam a respeito das estratégias que serão adotadas pelos demais jogadores.

O recurso ao formalismo permite generalizar os procedimentos adotados. Em um jogo em forma estratégica, diz-se que a estratégia s_{ji} é uma melhor resposta para o jogador j para o perfil de estratégias s_{-j} adotado pelos rivais se $u_j(s_{ji}, s_{-j}) \geq u_j(s'_j, s_{-j})$ para todas as outras estratégias, que não i , s'_j pertencentes ao conjunto de estratégias Cs_j do jogador j , ($s'_j \in Cs_j$).

9.3.1.2 Equilíbrios de Nash – ENs

No caso anterior Figura 9.5(a), se o jogador I escolhe E, a melhor resposta do jogador II é B. Mas a melhor resposta do jogador I à escolha de B pelo jogador II não é E, mas B. Portanto, (E, B) não é equilíbrio de Nash. Quando o jogador I escolhe B, a melhor resposta do jogador II é E e não B. Caso o jogador II escolha E, a melhor resposta do jogador I é B. Portanto, E é melhor resposta do jogador II à escolha de B pelo jogador I e B é melhor resposta do jogador I à escolha de E pelo jogador II, ou seja, compatibilidade mútua que aponta (B, E), como equilíbrio; e (B, E) configura conjecturas e previsões corretas dos jogadores, o que, podendo-se mais uma vez argumentar, não é imposição obrigatória do pressuposto de racionalidade, que apenas leva às estratégias racionalizáveis.

O equilíbrio de Nash virá a ser definido a partir do requisito de que os jogadores estejam mutuamente corretos acerca de suas conjecturas, caso de (B, E) citado anteriormente. Mais formalmente, o equilíbrio de Nash refere-se ao perfil $s = (s_1, \dots, s_j)$, associado aos j jogadores, de modo que para todo j , $j=1, \dots, J$, $u_j(s_{-j}, s_j) \geq u(s'_j, s_j)$. Devem todas as estratégias $s'_j \in C_s$.

Uma forma talvez menos trabalhosa de se chegar às estratégias racionalizáveis e ao(s) equilíbrio(s) de Nash – ENs, é a de assinalar os ganhos com barras, como se fez na Figura 9.5(a). Os ganhos que identificam melhores respostas do jogador I às conjecturas do jogador I acerca das escolhas de estratégias por parte do jogador II estão marcados com barras sob os números, e os pagamentos que identificam melhores respostas do jogador II às conjecturas do jogador II acerca das escolhas de estratégias do jogador I estão marcados com barras sobre os números. Células que vierem a conter barras sob e sobre os números identificam as estratégias que constituem ENs.

9.3.2 O DILEMA DO PRISIONEIRO: ESTRATÉGIAS DISCRETAS E EQUILÍBRIO DE NASH ÚNICO

É uma das mais famosas situações de interdependência estratégica analisadas pela teoria dos jogos e descreve, entre outras histórias concebíveis, a detenção em flagrante de dois ladrões por parte de um detetive zeloso. O detetive chega à delegacia com os ladrões, mas sem testemunhas, sabendo que o delegado exigirá provas do delito. Ocorre ao detetive colocar os ladrões em celas separadas, tornando-os incomunicáveis. Após alguma reflexão, o detetive resolve apresentar proposta a cada um dos detentos, em separado. A um deles sugere a traição, delatar o outro detento como responsável pelo crime, em troca de recompensa e confirmações perante delegado e juízes do testemunho contra o colega, que resultariam em menos anos de detenção. Ao outro detento o detetive apresenta a mesma proposta, o que motiva dissecar a armadilha proposta. Os detentos somente têm duas alternativas: cooperar (entre si) e não ceder à proposta do detetive e negar a cooperação que os levou juntos ao crime, cedendo à proposta e indiciando o colega. Os detentos não sabem se o outro confessou ou não, pois o jogo é jogo de informação imperfeita. Se ambos confessarem eles pegam cinco anos de prisão, se ambos não confessarem eles pegam um ano de prisão e se um confessar e o outro não, o que confessou é solto e o outro é penalizado com dez anos de prisão.

Sendo I e II os prisioneiros, o jogo na forma estratégica seria:

		II	n	c
		I	(-5, -5)	(0, -10)
		n	(-10, 0)	(-1, -1)
		n: não cooperar		
		c: cooperar		

Figura 9.6 O Dilema do Prisioneiro.

Convidamos o leitor a exercitar o uso das barras (que identificam as melhores respostas), sendo, assim, fácil concluir que o equilíbrio de Nash desse jogo é atingido quando os detentos decidem indicar o colega e se negarem a cooperar. Surpreende no resultado o fato de os jogadores obterem pena de cinco anos cada, no lugar da pena de apenas um ano de prisão, que ocorreria no caso de ambos cooperarem.

O resultado do jogo dramatiza as possibilidades frequentes da interdependência estratégica resultar em perdas coletivas e em ineficiências sociais, o que já levou muitos filósofos a interpretarem o dilema como representação bastante fiel da “guerra de todos contra todos” anunciada pelo Leviatã, de Hobbes. Note que o resultado foi obtido sem que supuséssemos, como se faz correntemente, que, em cada célula, o ganho de um jogador é a perda do oponente, e que a soma dos ganhos em cada célula é nula (o famoso *jogo de soma-zero*).

9.3.3 MODELO DE COURNOT: ESTRATÉGIAS CONTÍNUAS E EQUILÍBRIO ÚNICO

Uma interpretação corrente da literatura especializada de Microeconomia, Organização e Economia Industrial é a de que os jogos estáticos de oligopólio possuem a mesma estrutura do dilema do prisioneiro. Examinemos a proposta analisando o comportamento estratégico de duas empresas que disputam um mercado fazendo escolhas de quantidades a produzir e ofertar. Para simplificar a

descrição das regras do jogo, suponha que as condições de demanda são conhecidas e que o preço é dado por $p(q) = 1 - q$. Ignore-se, por enquanto, os custos de produção, ou suponha que os custos unitários de produção são constantes, iguais a c tal que $c=0$.³

Para os *duopolistas de Cournot*, as estratégias e escolhas possíveis de quantidades q_j situam-se, em princípio, no intervalo $[0,1]$, o que configura, diferentemente dos jogos até agora estudados, conjunto de estratégias com infinitos elementos e querendo isto também dizer que, no caso de uma das empresas selecionar 1, todo o mercado poderia ser abastecido pela empresa. Mas as empresas não tomam decisões como empresas completamente independentes (ainda que, nos jogos de decisões simultâneas, as decisões individuais dos jogadores não sejam observadas pelos oponentes) mas, reconhecendo antes interdependência estratégica, que poderia ser explicitada, para a empresa I inicialmente, por função de lucro individual da forma.

$$L_1 = (1 - q_1 - q_2) \cdot q_1$$

Com isso estabelecido convém retornar aos procedimentos anteriores, antes de identificar o EN para o jogo dos duopolistas. Admitamos, assim, que as firmas identifiquem “melhores respostas” às escolhas possíveis do competidor.

Quando a empresa I conjectura que a empresa II virá a produzir zero, a melhor resposta é selecionar a produção de monopólio, e $q_1 = 1/2$. Identifique-se esta estratégia na Figura 9.3, usando o eixo das abscissas para representar as escolhas da empresa I. Em outro extremo, caso a empresa I conjecture que a empresa II deseja abastecer todo o mercado, a melhor resposta da empresa I é selecionar o nível de produção igual a zero. Estes dois pontos definem assim a *Correspondência de Melhores Respostas (e/ou de Reação)* da empresa I, denotada por R_1 . A situação é simétrica, podendo o mesmo raciocínio ser aplicado à empresa II, para definir sua Curva de Melhor Resposta ou Curva de Reação, dada por R_2 .

C_I	C_{II}	C	N
C	C_I	$L^*/2, L^*/2$	L^C, L^N
N	N	L^N, L^C	L^C, L^C

Figura 9.7 As correspondências de melhores respostas no modelo de Cournot.

Analiticamente, as melhores respostas das empresas seriam dadas pela regra de maximização de lucros. Para a empresa I, dado que $\partial [(1 - q_1 - q_2) \cdot q_1] / \partial q_1 = 0$, as melhores respostas seriam fornecidas pela expressão $q_1 = R_1(q_2) = (1 - q_2)/2$.

O *Equilíbrio de Cournot-Nash, ECN*, ficará definido pela interseção das duas correspondências, ponto em que as conjecturas mútuas estão corretas e se compatibilizam, prevendo $1/3$ do mercado para cada uma das empresas e $2/3$ de quantidade total produzida, com $p=1/3$ e lucros individuais de $1/9$. Assim, a quantidade total a ser produzida é superior à de monopólio, com preços e lucros inferiores aos que este último arranjo propiciaria, mas inferior à quantidade total que seria produzida em concorrência perfeita.

A forma estratégica apresentada na Figura 9.8 ilustra a semelhança de estruturas, do jogo dos duopolistas de Cournot com o Dilema do Prisioneiro (L^* é o lucro de monopólio e L_c o lucro previsto pelo modelo de Cournot) e é ilustração do que se denomina de análise da estabilidade do equilíbrio. De fato, caso as regras do jogo e instituições fossem outras e as empresas

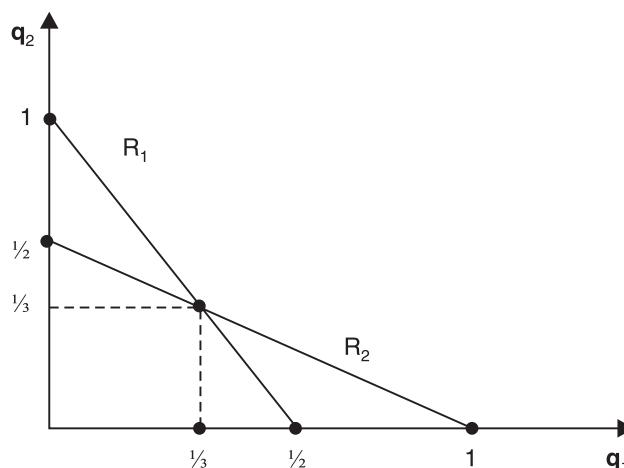


Figura 9.8 As correspondências de melhores respostas no Modelo de Cournot.

pudessem se coordenar como um monopólio (dividindo igualmente a quantidade a ser produzida e agindo como se ambas cooperassem), a produção individual seria de $1/4$. A célula correspondente à escolha de estratégias {C,C} e os valores de $L^*/2$ podem, então, ser identificados, recomendando-se fazer o mesmo com relação à célula {N,N} e aos lucros individuais, L_c , alcançados no equilíbrio de Nash-Cournot.

E por que o acordo de monopolização não seria efetivo? Segundo o modelo aqui proposto, porque, uma vez que uma das empresas venha a anunciar a produção de $1/4$, a outra empresa teria incentivos para, obedecendo sua Correspondência de Melhor Resposta, responder com a produção $(1 - 1/4)/2 = 3/8$. Identifique os lucros das empresas I e II e alcance os valores das células correspondentes a {N,C}, {C,N}, L^N e L^C , que são os lucros individuais obtidos quando a escolha de estratégias é divergente. Com a forma estratégica completada, você terá alcançado a previsão de que o equilíbrio de Nash-Cournot viria a ser jogado. Mas vale também continuar o exame da estabilidade. Identificada a produção individual quando uma das empresas coopera e a outra desvia, você pode supor que a outra empresa virá, também, a desviar, o que levaria a empresa que inicialmente escolheu C a responder, agora, com $(1 - 3/8) = 5/16$; o que levaria a empresa que decidiu por não cooperar no primeiro lance a adotar a melhor resposta de $11/32$, resultados que configurariam convergência das estratégias de melhores respostas para $(1/3, 1/3)$.

Essa comparação sugere que, para a coletividade de empresas envolvidas, o ECN prevê “ineficiências organizacionais”. Mas a situação também contrasta com os resultados da concorrência perfeita, em que o mercado estaria sendo totalmente abastecido a preços nulos (pois o custo unitário variável é nulo, por hipótese, e a quantidade total produzida seria de 1, valendo sugerir que se localize no gráfico o par de estratégias $(1/2, 1/2)$). Portanto, o ECN configuraria também “ineficiências sociais”, à luz da produção máxima e do preço mínimo que poderia ser oferecido ao mercado e aos consumidores, resultado que virá a ser questionado pelo Modelo de Bertrand.

9.3.4 O MODELO DE BERTRAND: PRODUTOS HOMOGENEOS

Este modelo tem estrutura semelhante à do Modelo de Cournot e à do Dilema do Prisioneiro, entretanto diferem destes porque agora as empresas fazem escolhas de preços, em vez de selecionarem estratégias de quantidades a ofertar. Suponham-se condições de demanda idênticas às anteriores, em que os consumidores sempre optam por comprar da empresa com menor preço, uma regra acordada de divisão de mercado e custos unitários constantes. A regra (explicitada para facilitar o entendimento) diz que, a preço comum, cada empresa fica com metade da demanda total. Assim sendo, as duas empresas anunciam seus preços simultaneamente e as vendas para a empresa I, por exemplo, podem ser especificadas por D_1 e por:

$$\begin{aligned} D_1(p_1, p_2) &= D(p_1) \quad \text{quando } p_1 < p_2; \\ D_1(p_1, p_2) &= 1/2 D(p_1) \quad \text{quando } p_1 = p_2; \\ D_1(p_1, p_2) &= 0 \quad \text{quando } p_1 > p_2; \end{aligned}$$

e a função objetivo como

$$L_1 = (p_1 - c) D_1(p_1, p_2)$$

A situação é simétrica e algo análogo pode ser estabelecido para a empresa II. O *equilíbrio de Bertrand-Nash – EBN* propõe que as duas empresas selecionarão $(p_1^*, p_2^*) = (c, c)$, o que, no caso de $c = 0$, equivale a preços nulos. Com preços nulos, as quantidades individuais produzidas seriam de $1/2$ e a quantidade total alcançaria 1, como se os duopolistas estivessem competindo em regime de concorrência perfeita.

Antes de passar a qualificações, é importante proceder aos mesmos testes de consistência a que se submeteu o ECN. Considere-se, por exemplo, que os custos unitários são positivos e iguais a c e que $p_1 > p_2^* > c$. Nesse caso, a empresa I fica sem demanda e com lucros nulos. Se a empresa II mantiver o preço em p_2^* , e se a empresa I cobrar $p_1 = p_2^*$, o lucro unitário da empresa I será de $(p_1 - c)/2$. Caso, entretanto, a empresa I decida cobrar preço um pouco inferior ao da empresa II, na medida de $r > 0$, a empresa I auferá a margem unitária de lucro de $p_2^* - r - c$. A estratégia de cobrar $p_2^* - r$ domina, portanto, aquelas duas possibilidades. Além disso, empresa alguma cobrará preços inferiores ao custo unitário, e se $p_1 > p_2 = c$, a empresa I ficará sem mercado, possibilitando inclusive à empresa II aumentar seu preço em medida pequena. A empresa II está diante de conjecturas semelhantes, o que confirma a estabilidade do EBN, no qual $p_1^* = p_2^* = c$.

9.3.5 ENs MÚLTIPLOS: SELECIONANDO EQUILÍBRIOS MÚLTIPLOS

Como já assinalamos, os jogos podem apresentar equilíbrios de Nash múltiplos e vale voltar ao jogo da Figura 9.1, agora na sua forma estratégica, para ilustrar os problemas a serem enfrentados. A Figura 9.9 reproduz o jogo da Figura 9.1, com os equilíbrios assinalados, como a seguir.

	C_{II}	Ee	Eb	Be	Bb
C_I					
E		1,1	<u>3,2</u>	<u>3,2</u>	<u>3,2</u>
B		<u>2,4</u>	<u>2,4</u>	<u>2,4</u>	<u>4,3</u>

Figura 9.9 Equilíbrios de Nash para o jogo do quadro I.

Como se pode conferir, os equilíbrios de Nash definem o conjunto $\{(B, Ee), (E, Eb), (E, Be)\}$ como maneiras prováveis de o jogo vir a ser jogado, situação que já foi caracterizada como embaracosa para efeitos de previsão e, talvez, crítica para os jogadores, pois estão diante de dificuldades de se coordenarem para selecionar o desenlace do jogo. Por esta razão, diz-se que os equilíbrios múltiplos caracterizam falhas de coordenação, ponto que voltará a ser exemplificado a seguir.

9.3.5.1 Exemplo de falha de coordenação

O exemplo que se segue, referenciado pela Figura 9.10, tem versão curiosa, denominada de *batalha dos sexos*. Neste jogo, dois parceiros decidem o que fazer no final de semana, sabendo ambos que um deles prefere ir ao cinema (uma das estratégias) e o outro prefere prestigiar o futebol (a segunda estratégia). Entretanto, ambos preferem estar juntos a ficar separados, o que na figura pode ser representado por ganhos de zero, no lugar de 1. Mas outro roteiro possível para o jogo é o que discute interdependências estratégicas entre dois gerentes de uma empresa. O gerente de *marketing* (C_I) e o gerente de produção (C_{II}) procuram tomar decisões acerca, respectivamente, do tamanho de lotes a produzir e do número de variedades de produtos a oferecer aos clientes. Há intenções da parte dos gerentes de maximizar o lucro da empresa para a qual trabalham, ou de minimizar custos, de estocagem de produtos, de paradas na linha de produção e outros conflitos a serem superados. Para um dos gerentes, oferecer mais produtos significa atender melhor aos desejos dos clientes, e preços mais elevados. Entretanto, para um tamanho definido de lote, aumentar o número de produtos significa também aumentar estoques. Para o outro gerente, um dado número de variedades está associado a um tamanho apropriado de lotes. Lotes menores estão associados a estoques menores, mas obrigam a empresa a incorrer em custos de paradas nas linhas de produção e de reorientação destas para outros produtos. Quanto maior o tamanho de lotes, mais onerosa a estratégia de aumentar a gama de variedades. Na Figura 9.10, N denota a estratégia de produzir um número grande de produtos e n a estratégia de produzir número menor de produtos; I denota a estratégia de operar com lotes menores e L denota a estratégia de operar com lotes maiores. Os pagamentos traduzem as preferências e os ganhos de 1 anunciam que as perdas associadas a descoordenações não seriam totais.

	C_{II}	I	L
C_I			
n		1,1	3,4
N		4,3	1,1

Figura 9.10 Exemplo de Falência de Coordenação.

O jogo tem dois equilíbrios de Nash e a teoria dos jogos sugere que, em casos como estes, a seleção do equilíbrio estará sendo feita por elementos não incluídos na descrição do jogo. Consta que o equilíbrio (n,L) teria sido o escolhido pela Ford na época de fabricação do modelo T e que o equilíbrio (N,I) foi o selecionado pela General Motors, por constituírem *Pontos Focais*⁴ de uma empresa e da outra, respectivamente.

9.4 Jogos Dinâmicos com Informação Completa: Indução Retroativa (*backward induction*), Subjogos e Equilíbrio Perfeito em Subjogo

Antes de se recorrer a situações mais complexas, suponhamos que o jogo entre as empresas I e II se dê como está representado na Figura 9.1. Uma das maneiras de analisar o jogo apresentado, já conhecida, é recorrer à representação estratégica.

Como se viu, o conjunto de estratégias da firma I é {E,B}, e o conjunto de estratégias da empresa II é {Ee, Eb, Be, Bb}. Procedimentos análogos aos adotados na seção anterior permitem que se identifique o *conjunto de equilíbrios de Nash*

$\{(B,Ee),(E,Eb),(E,Be)\}$, ou, renomeando convenientemente as estratégias, $\{(B, e \text{ dado II.2}), (E, e \text{ dado II.1}), (E, b \text{ dado II.1})\}$, resultado que merece ser por nós retido, para elaboração posterior.

Recorramos, agora, à representação extensiva e dinâmica do jogo e levemos em conta que, sendo o jogo de conhecimento comum das empresas/jogadores, a empresa I, antes de tomar sua decisão, tire partido da possibilidade de se colocar no lugar da empresa II, seja no nó II.1, seja no nó II.2. Tentando antever os resultados condicionais de suas possíveis decisões, por empatia, confiando na racionalidade da empresa II, a partir de II.1, e caminhando no sentido do nó final do jogo, a empresa I pode deduzir que a empresa II decidiria por b , que propicia à empresa II ganho de 2, superior ao ganho de 1 que seria auferido caso a empresa II decidisse por e (a duplicação das ramificações ajuda a explicitar o procedimento). Repita o procedimento, levando em conta a possibilidade de o jogo chegar em II.2. Tendo descartado as ações e , no caso de II.1 e b , no caso de II.2, considere então que os nós II.1 e II.2 são nós finais do jogo e transfira os ganhos $(3,2)$ e $(2,4)$, respectivamente, para os novos nós finais. Você chegou à *forma reduzida* do jogo inicial, com a qual irá se defrontar a empresa I ao ter de tomar sua decisão inicial e depois de proceder à indução retroativa.

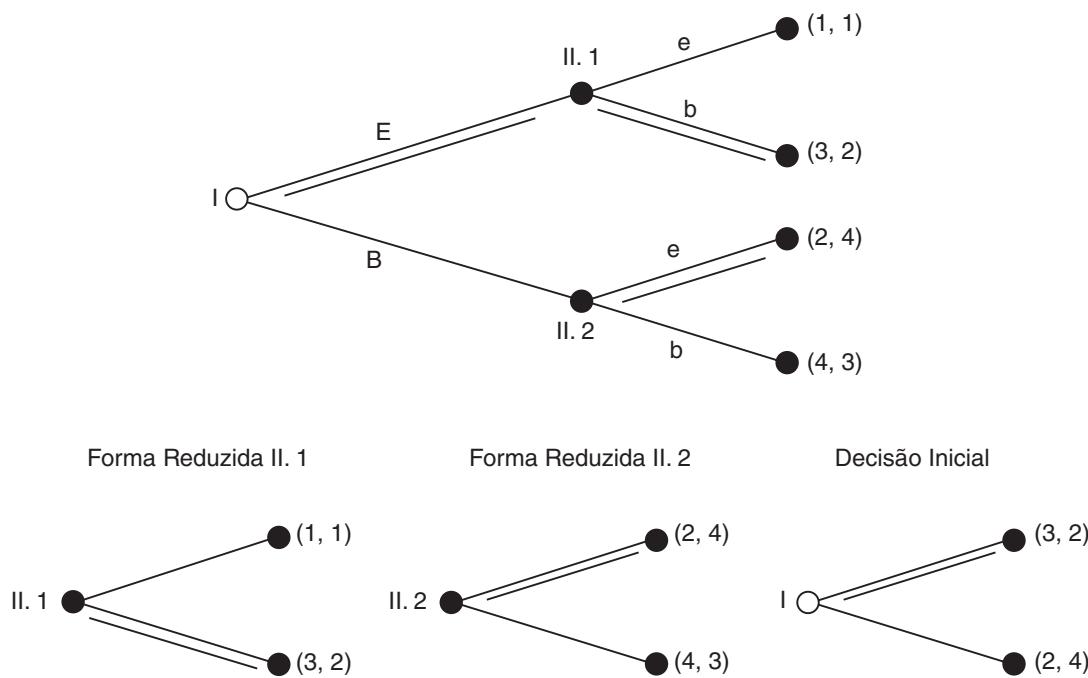


Figura 9.11 Exemplificação da Indução Retroativa (*backward induction*).

Como esperado, a empresa I decidirá por E, e o equilíbrio por indução retroativa, que se denominará de *EN sequencialmente racional*, será $(E, b \text{ dado II.1})$. Como se pode verificar, a indução retroativa refinou as previsões identificadas na análise da forma estratégica do jogo. Os equilíbrios que não passam pelo *teste da racionalidade sequencial*, segundo o procedimento adotado, mereceriam ser descartados.⁵

Convém, agora, a partir do resultado alcançado, indagar se é possível identificar ENs que satisfaçam o princípio de racionalidade sequencial em jogos mais gerais que envolvam informação imperfeita. Isso se adequaria às motivações iniciais do texto e se justifica recorrendo à forma extensiva que se reproduz na Figura 9.12.

Agora, podemos indagar ao leitor como proceder à obtenção das formas reduzidas do jogo. Notemos que, após alcançar os nós II.1 e II.2, as empresas I e II passam a jogar jogos de decisões simultâneas (afinal, a teoria dos jogos pode conviver com mudanças nas regras do jogo). Como se pode verificar, o jogo que se segue ao nó II.2 é um Dilema do Prisioneiro e o jogo que se segue ao nó II.1 é um jogo de coordenação. O equilíbrio de Nash do jogo à direita (a partir do nó II.2) é encontrado quando ambos renegam e perdem 5. O jogo que se inicia no nó II.1 tem dois equilíbrios de Nash (f,F) e (F,f) . Resolvendo o jogo por indução retroativa, já observando os equilíbrios de Nash, o jogador I prefere jogar B, pois assim ele poderá ganhar ou perder 1 enquanto que se jogasse E deveria perder 5.

Antes de adotar procedimento natural, convém justificar sua adoção com novas definições. Um *subjogo* de jogo em forma extensiva é subconjunto do jogo com as seguintes propriedades: (1) o subjogo tem início em um conjunto de informações que contenha um único nó de decisão, e contém todos os nós de decisão que se sucedem, e apenas estes nós; e (2) se um nó de decisão x pertence

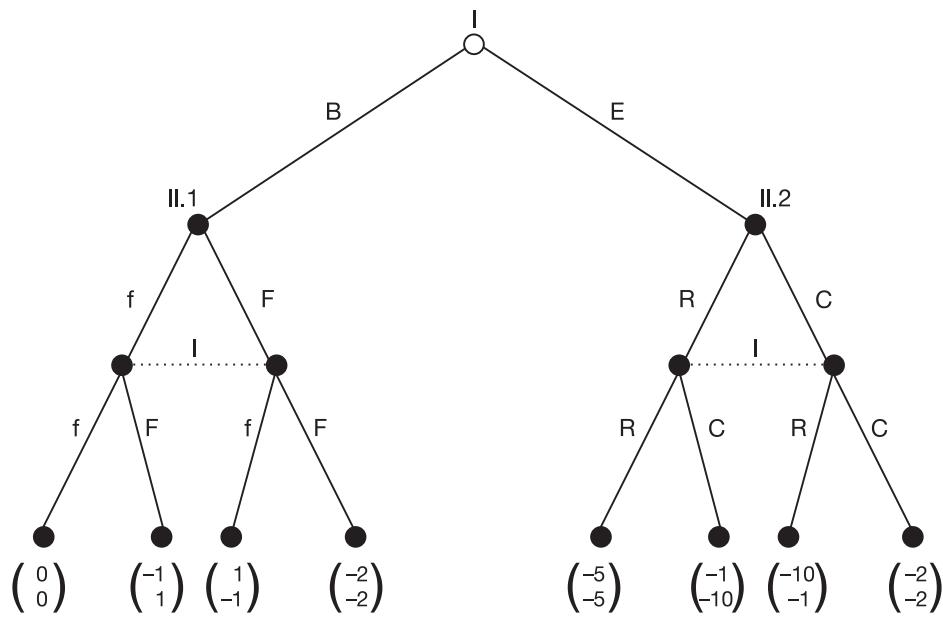


Figura 9.12 Representação Extensiva, Subjogos e Racionalidade Sequencial.

ao subjogo, então todo nó de decisão que pertence ao mesmo conjunto de informação ao qual pertence x deve pertencer ao subjogo. Obviamente, o jogo por inteiro é um dos subjogos a serem encontrados. O jogo do exemplo anterior possui três subjogos.

Sugerimos, então, que você adote *método de indução retroativa generalizada* (v. MWG, *op.cit.*, p. 277, para prova formal), ou seja, que inicie pelo final da árvore do jogo, e identifique ENs para cada um dos subjogos finais; depois, selecione um EN em cada um dos subjogos e derive a forma extensiva reduzida, repetindo os passos anteriores para o jogo reduzido e assim por diante, até alcançar perfil de equilíbrios, como acontece no exemplo citado, ou o equilíbrio único, quando for o caso. Mas examine também a definição que se segue.

Segundo Selten, um perfil de estratégias $s = (s_1, \dots, s_J)$ em um jogo em forma extensiva de J jogadores é *Equilíbrio Perfeito de Nash em Subjogo* (*Subgame Perfect Nash Equilibrium – SPNE*), se este equilíbrio induz um EN em cada subjogo do jogo em forma extensiva, devendo-se notar que, a partir de analogias com resultados já assinalados, todo SPNE é EN, mas nem todo EN é SPNE. Além disso, vale assinalar o resultado que o leitor já teria encontrado no exemplo anterior, o fato de que o conjunto de SPNEs não é necessariamente único.

9.4.1 O MODELO DE STACKELBERG

No Modelo de Stackelberg, supõe-se que a empresa I seja empresa já estabelecida (*incumbent*) no mercado, que pode fazer valer seu compromisso de permanecer nesta atividade instalando capacidade antes da empresa E (*entrant*). Consideremos que os custos irrecuperáveis (*sunk costs*) comprometam as decisões da empresa I por certo período de tempo. Com isso, a empresa I estaria sinalizando à empresa E sua disposição de permanecer no mercado em períodos futuros. Isso dado, no primeiro estágio do jogo a empresa I deve decidir pela quantidade a produzir. Com as simplificações adotadas, pode-se dizer que a empresa E é concorrente potencial que pode ingressar sem custos no mercado, ou é empresa que já participa deste, como empresa marginal. No segundo estágio do jogo, a empresa E decide então pela quantidade a ser produzida, tendo observado a decisão prévia da empresa I .

Como se viu, em jogos dinâmicos o jogador que tem a liderança – o responsável pelo primeiro movimento do jogo – pode se colocar no lugar dos demais jogadores e antecipar suas decisões. No caso, tudo se passaria como se a empresa I pudesse inspecionar as melhores respostas, dadas por RE, a serem selecionadas pela empresa E , dadas suas conjecturas a respeito das estratégias a serem selecionadas no primeiro estágio pela empresa I , vindo a empresa I a usar esta informação para decidir pela capacidade de produção que mais lhe convém, no primeiro estágio do jogo. Dada a decisão inicial da empresa I , e como a empresa E confere credibilidade à decisão da empresa I , a empresa E observa a decisão da primeira e decide sob o comando de sua correspondência de melhores respostas.

Considere, por exemplo, que a correspondência de melhores respostas da empresa E seja dada por $q_E = (A - c - q_I)/2$, como em exemplo anterior. Como se viu, nesse caso a função de demanda é dada por $p(q) = A - q$, e os custos unitários são constantes

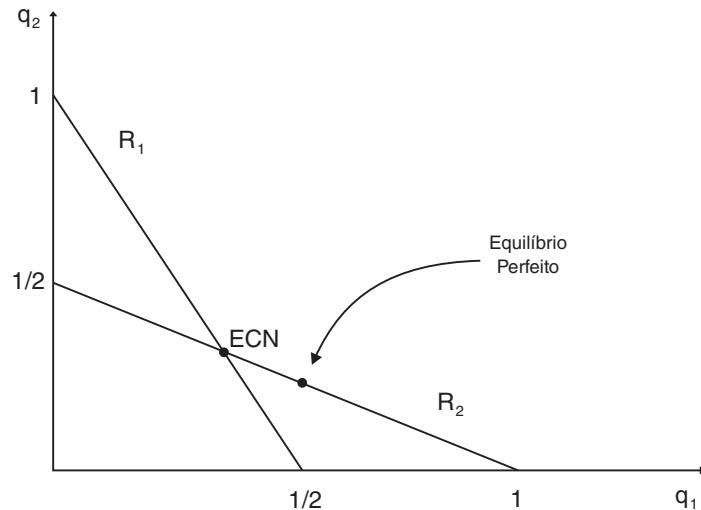


Figura 9.13 O equilíbrio perfeito de Nash em subjogo – SPNE no modelo de Stackelberg.

e iguais a c para as duas empresas. Tendo esse conhecimento, a empresa I pode, assim, especificar sua função objetivo, que passa a ser $L_I = (A - c - q_I - (A - c - q_I)/2) q_I$. A otimização dessa função indicará à empresa I a decisão a ser tomada, ou seja, $q_I = (A - c)/2$. Na sua vez de jogar, a empresa E usará essa informação em R_E , obtendo como resposta $q_E = (A - c)/4$.

O Equilíbrio Perfeito de Nash em Subjogo, SPNE, é assim $[(A - c)/2, (A - c)/4]$, e está representado na Figura 9.13, onde se supõe $A = 1$ e $c = 0$, e que q_I representa as quantidades produzidas da empresa I e q_E representa as quantidades produzidas da empresa E . Note, a propósito do modelo estudado, que as escolhas da empresa I induzem pontos na correspondência de melhor resposta da empresa E .

No equilíbrio encontrado, a empresa I produz a quantidade de monopólio, $1/2$ (no caso), seus lucros são máximos, e a empresa E aufera lucros positivos, mas inferiores aos lucros que obteria caso o jogo fosse um jogo de Cournot. Recomendamos que o leitor examine, como exercício útil, que os lucros da empresa E decrescem, ao longo de sua correspondência de melhores respostas, de $1/2$ que configura situação em que a produção da empresa I é nula, e a empresa E pode decidir como monopólio, auferindo lucros máximos, até o ponto em que a empresa I produz 1 , e os lucros da empresa E são nulos, valendo comentário equivalente para a empresa I .

Notemos que, no caso da disputa exigir da empresa E que realize investimentos iniciais para ingressar no mercado e no caso de esses investimentos serem superiores aos lucros correspondentes ao equilíbrio final, a empresa E poderá ter sua entrada no mercado bloqueada.

9.4.1.1 Estratégias Mistas e o Equilíbrio em Estratégias Mistas

Até aqui desconsideramos o uso de estratégias mistas por parte dos jogadores. Pela nossa experiência, o conceito de estratégia mista é de assimilação inicial difícil (razão porque, aliás, Dixit e Nalebuff e Dixit e Skeath dedicam longas e saborosas partes de seus livros ao tópico), mas se torna não descartável depois de apreendido. Correremos aqui o risco de oferecer uma abordagem sintética para as estratégias mistas, e para o equilíbrio em estratégias mistas, lembrando ao leitor (e recorrendo ao argumento de) que jogos com estratégias finitas que não possuem equilíbrio de Nash em estratégias puras possuem equilíbrio de Nash em estratégias mistas.

Há inúmeras justificativas para supor o uso de estratégias mistas, tal como se exemplificará. Uma das mais consagradas e utilizadas é a que atribui aos jogadores intenções de tornar suas escolhas de estratégias difíceis de serem previstas pelo outro jogador (como em jogos de tênis, basquete e futebol). Mas o objetivo de transferir imprevisibilidades ao oponente obriga ao jogador evitar que seu comportamento se torne previsível e tornar suas escolhas aleatórias.

Por razões didáticas, em vez de recorrer a jogo que não possua equilíbrio em estratégias puras, recorreremos a jogo que possua dois equilíbrios e procuraremos identificar o terceiro equilíbrio, em estratégias mistas. No jogo da Figura 9.14 não é do interesse do jogador I assegurar ao jogador II que Cooperar (C) será implementada com certeza e probabilidade unitária. Também não há interesse por parte do jogador II de dar garantias ao jogador I de que (N,C) pode ser implementado.

Mas isso obriga a que a escolha de C ou N por parte do jogador I ou do jogador II seja uma escolha probabilística (ou errática, se as probabilidades forem interpretadas como a frequência relativa de escolhas de C ou N em jogos similares). Assim sendo, supõe-se que o jogador I estará escolhendo C com probabilidade $(1 - p)$, N com probabilidade p , que o jogador

		II	C	N
		I		
		C	2,2	2,3
		N	3,2	0,0

Figura 9.14 Exemplo de equilíbrio de Nash em estratégias mistas.

II estará escolhendo C com probabilidade $(1 - q)$, N com probabilidade q e que as escolhas de probabilidades dos jogadores sejam independentes.

O Equilíbrio de Nash em estratégias puras será um par (p, q) , obtido da seguinte maneira. Quando o jogador II escolhe entre C ou N de acordo com $(1 - q, q)$, o jogador I pode distinguir entre C e N à luz de *ganhos esperados* (*e utilidades esperadas*), que denotaremos por $Eu_1(C)$ e $Eu_1(N)$, respectivamente, que suas estratégias propiciariam, dado o comportamento probabilístico suposto do jogador II, ou seja,

$$\begin{aligned} Eu_1(C) &= 2(1 - q) + 2q = 2, \quad \text{relativo à escolha de } C; \text{ e,} \\ Eu_1(N) &= 3(1 - q) = 3 - 3q, \quad \text{quando escolhe } N. \end{aligned}$$

Cooperar-C será melhor resposta para o jogador I, $Eu_1(C) \geq Eu_1(N)$, ou quando $q \geq 1/3$, situação em que C será escolhida com probabilidade 1. Quando $q < 1/3$, a melhor resposta para o jogador I será escolher N com probabilidade 1. Quando $q = 1/3$ o jogador I será indiferente entre a escolha de C ou N, vindo a selecionar p no intervalo $[0, 1]$. Podemos desenvolver o mesmo raciocínio para o jogador II e sintetizar com expressões e graficamente o que identificamos em termos de respostas ótimas e reações do jogador I e do jogador II na Figura 9.15. A expressão à esquerda, por exemplo, nos diz que o jogador I escolherá p segundo suas respostas ótimas $R(q)$, e p terá valor de 0 quando $q > 1/3$, estará no intervalo $[0, 1]$ quando $q = 1/3$, e terá valor de 1 quando $q < 1/3$. A figura integra as melhores respostas dos dois jogadores, sendo fácil verificar que o Equilíbrio de Nash em estratégias mistas será $(p^*, q^*) = (1/3, 1/3)$.

Com isso, os ganhos e utilidades totais dos jogadores podem ser obtidos, sendo, para o jogador I, de:

$$Eu_1(p^*, q^*) = 1/3 \cdot 1/3(3) + 1/3 \cdot 2/3(3) + 2/3 \cdot 1/3(4) + 2/3 \cdot 2/3(0) = 17/9$$

De forma mais genérica, podemos denotar a estratégia mista do jogador j por σ_j entendendo por isso que o jogador j atribui às estratégias puras s_{j1}, \dots, s_{jM} as probabilidades $\sigma_{j1}, \dots, \sigma_{jM}$, que somam a unidade, ou seja $\sigma_j = (\sigma_{j1}, \dots, \sigma_{jM})$, e $\sum_m \sigma_{jm} = 1$, $\sigma_{ji} \geq 0$. No caso citado, denotemos C de s_{11} e N de s_{12} (para o jogador N e C podem ser chamadas s_{21} e s_{22} , respectivamente). Como anteriormente explicado, um equilíbrio em estratégias mistas é um perfil de estratégias mistas, com $p^* = (1 - p^*, p^*)$, para o caso do jogador I, e $q^* = (1 - q^*, q^*)$, para o jogador II, e um perfil para o conjunto de jogadores pode ser denotado por $\sigma = (\sigma_1, \sigma_2)$, sendo os componentes de σ vetores (no equilíbrio, p^* e q^*).

Com a motivação já oferecida pelo cálculo da utilidade total, pode-se agora definir a *utilidade de von Neumann & Morgenstern – vN&M*. Como já assinalado, cada célula mostra o resultado da atuação da utilidade u (que representa as preferências dos jogadores) sobre os desenlaces possíveis dos jogos, tendo ali o *status* cardinal. Essa é a utilidade de *vN&M*, nada obrigando a que venha a ser linear (de maneira geral u pode ser uma função côncava se os jogadores forem avessos ao risco, e não neutros ao risco, como aqui supomos). Por sua vez, Eu é uma função linear que atua sobre os “prêmios”. O leitor está convidado a verificar que, no equilíbrio, e com jogadores neutros ao risco, teremos $uE = Eu$.

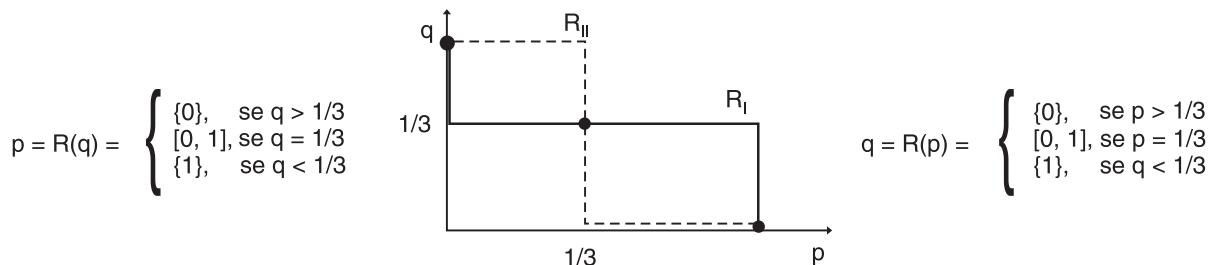


Figura 9.15 Respostas Ótimas e Equilíbrio de Nash em Estratégias Mistas.

9.5 Nota sobre Jogos Repetidos

Podemos, agora, passar à intuição básica elaborada por tipo especial de jogos dinâmicos, os jogos repetidos. Recorra-se ao exemplo simples que se segue. O jogo a ser considerado tem a estrutura do Dilema do Prisioneiro, em que a Cooperação (C) bilateral permite às empresas dividirem os mercados e obter conjuntamente lucros de monopólio, denotados por L^* . Quando uma das empresas se nega a cooperar – N, e quebra o acordo, a empresa pode (instantaneamente) abastecer o mercado como ofertante único e auferir os lucros de monopólio L^* . Mas a situação é simétrica, e quando ambas se negam a cooperar, restam-lhes excedentes e lucros menores, como os resultados associados ao equilíbrio de Bertrand-Nash, denotados por L_B . As estratégias e pagamentos estão indicados a seguir, na Figura 9.16.

(N,N) é o equilíbrio de Nash – EN, único do jogo estático. Mas concebemos agora a possibilidade desse jogo vir a ser jogado repetidamente, sabendo cada um dos jogadores que, a cada turno do jogo, as estratégias jogadas no turno anterior serão divulgadas e virão a se tornar conhecimento público. Procure não recorrer de imediato à forma extensiva desse jogo dinâmico e nem à possível representação estratégica da forma extensiva. Alerta-se que a mera descrição do jogo repetido é peculiarmente complexa, em sentidos que se motivam a seguir.

Em relação ao primeiro turno do jogo, cujo equilíbrio ainda não é conhecido, denotemos as estratégias do primeiro jogador por s_1 e s_2 e as do segundo jogador por t_1 e t_2 . Como as jogadas do primeiro turno são conhecidas no segundo turno do jogo, o conjunto H , que segue, denotaria as combinações possíveis de estratégias, e $H = \{(s_{1,t1}), (s_{1,t2}), (s_{2,t1}), (s_{2,t2})\}$. H configura a *história do jogo*, quando os jogadores passam do primeiro para o segundo turno. Procuremos, agora, identificar as estratégias puras de um dos jogadores no segundo turno do jogo. Essas estratégias seriam definidas por um s_i , que teria sido jogado no primeiro turno, e, adicionalmente, pela estratégia pura que viria a ser selecionada no segundo turno, em função da história do jogo. Na verdade, o segundo elemento do par que define a estratégia pura do segundo turno é uma função, que associa a cada história possível, H_h (e $h=1, \dots, 4$ no caso), a estratégia pura a ser jogada no segundo turno. A h_1 podem estar associadas às estratégias s_1 , ou s_2 , ou s_3 , ou s_4 , o mesmo ocorrendo com relação a h_2 , a h_3 e a h_4 . Com esse raciocínio, nós identificaremos 16 funções e 32 estratégias puras!

Jogos repetidos são, assim, peculiarmente difíceis de serem descritos, mas configuram questionamento útil e obrigatório das situações estáticas que analisamos anteriormente. Em particular, a previsão do Leviatã e do Dilema do Prisioneiro, da “guerra de todos contra todos” merece ser cotejada por desenlaces que tenham a força das reciprocidades (coopero hoje, confiando que, no futuro, você, ou outro jogador que esteja no seu papel, venha a cooperar). Mais concretamente, é difícil antever que duas empresas venham a estabilizar um mercado no equilíbrio de Bertrand-Nash, com lucros nulos.

Assim sendo, recorramos a uma possível história, enredo ou rotina de reciprocidade. Por exemplo, a *rotina* ou *estratégia intimidadora* (também denominada de *estratégia de reversão de Nash*) inicia cooperando, mas sucumbe ao primeiro sinal de desacordo, transitando para N, e aí permanecendo. Voltando ao exemplo, isto quer dizer que os dois jogadores jogarão a estratégia intimidadora e/ou de reversão e que todos os elementos de H até o turno em questão são integrados por C apenas, ou que se está no primeiro turno do jogo. Mais formalmente, pode-se assim descrever as estratégias puras nesses jogos simples como,

$$s_{jt}(H^{t-1}) = C \text{ se todos os elementos de } H^{t-1} \text{ forem } (C, C) \text{ e / ou se } t = 1; \text{ ou} \\ s_{jt}(H^{t-1}) = N, \text{ de outra forma.}$$

Nesse ponto, duas possibilidades devem ser contempladas e a primeira delas supõe que o jogo será jogado um número limitado de jogadas. Argumentos de Indução Retroativa e de Equilíbrio em Subjogo permitem prever que (N,N) seriam as estratégias escolhidas pelos jogadores no último turno de jogadas, conhecimento que permite prever que (N,N) estaria sendo obtido no penúltimo turno e, assim, sucessivamente “para trás”, levando a prever que (N,N) viria a ser implementado no estágio inicial. Ou seja, (N,N) é SPNE e induz o EN (N,N) em cada subjogo e não (C,C).

	II	C	N
I			
C	$L^*/2, L^*/2$	0, L^*	
N	$L^*, 0$	L_B, L_B	

Figura 9.16 Representação estratégica para o Dilema do Prisioneiro, com pagamentos dados pelo equilíbrio de Bertrand-Nash, e por acordo de monopolização.

Consideremos, entretanto, que o horizonte para o término do jogo é infinito. Equivalentemente, os jogadores podem ter incertezas quanto ao término do jogo, em vez de saberem que o jogo termina com probabilidade igual a 1, após número determinado de fases, como no parágrafo anterior. Pode-se, também, supor que os jogadores talvez não sejam totalmente imediatistas, como anteriormente, descontem pagamentos futuros e possuam taxa de desconto $\delta \in (0;1)$.

O mais importante a reconhecer, nesse ponto, é que, depois de qualquer história prévia do jogo, as estratégias especificadas pela rotina intimidadora constituem EN do subjogo que é o jogo (anteriormente caracterizado) jogado número infinito de vezes. O cálculo dos pagamentos associados às estratégias C e N é simples. Cooperando, cada um dos jogadores poderia obter

$$V(\text{Cooperar}) = L^*/2 + \delta \cdot L^*/2 + \delta^2 \cdot L^*/2 + \dots = (L^*/2) / (1 - \delta)$$

em que δ representa a probabilidade do jogo continuar, ou o desconto que os jogadores aplicam a pagamentos futuros. Caso um dos jogadores implemente N, no turno do jogo em que isto acontece o jogador aufera os lucros de monopólio L^* , mas nos turnos seguintes este jogador estará auferindo apenas L_B , ou seja, lucros nulos, então:

$$V(\text{Não cooperar}) = L^*$$

O acordo será compatível do ponto de vista do incentivo aos jogadores quando

$$V(\text{Cooperar}) - V(\text{Não cooperar}) \geq 0, \text{ ou quando } \delta \geq 1/2.$$

O equilíbrio obtido, (C,C), sugere então que a *Monopolização e Colusão Tácita* entre as empresas e jogadores é um resultado viável (há outros equilíbrios possíveis, inclusive o não cooperativo). O importante a reter é que se trata de resultado cooperativo (possível de qualificar outras situações de interdependência) obtido por meio de mecanismos puramente não cooperativos.⁶ Curiosamente, o resultado decorre de importante teorema da teoria dos jogos, denominado de Teorema Popular (*folk theorem*), cuja autoria ainda é desconhecida.

9.6 Conclusão

Esperamos que esta introdução tenha atendido às expectativas do estudante e do leitor e contenha boas motivações para recorrerem a textos de maior fôlego. Nesse sentido, fazemos a seguir um comentário da bibliografia indicada. No nível introdutório e com exigências técnicas menores do que as que o presente texto em alguns momentos apresentou, recomendamos entusiasticamente o livro de Dixit e Skeath (1999), que se trata de obra preparada para o ensino da teoria dos jogos em disciplinas básicas, que antecedam (segundo a proposta dos autores) disciplinas formais de teoria (em cursos de Administração, Biologia, de Economia, e de Ciência Política). O livro amplia e sistematiza a excelente introdução de Dixit e Nalebuff (1991), que se tornou *best-seller* nos Estados Unidos da América. A bibliografia contém outros textos de introdução que merecem ser examinados (os que não estarão sendo contemplados pelas citações que se seguem). Como textos de base para estudos mais avançados, Gibbons (1992) e Binmore (1992) (com ou sem as extensões monumentais de 1994 e de 1998), continuam atuais. Myerson (1991) e Osborne e Rubinstein (1994) já são considerados textos clássicos, mas se tratam de livros de leitura mais exigente. Uma abordagem mais compactada da teoria dos jogos pode ser encontrada nos capítulos 7, 8 e 9 dedicados à teoria dos jogos por Mas-Colell, Whinston e Green, MWG (1995), que contém excelente sistematização de conceitos básicos, e extensões naturais da teoria dos jogos, no capítulo dedicado ao estudo de Poder de Mercado (12), e nos capítulos 13 e 14, dedicados a problemas de contratos e a desenho de mecanismos contratuais de incentivo. A esse respeito, vale chamar a atenção do leitor para os textos de Macho nos quais a informação incompleta e a teoria de contratos são tratadas de forma rigorosa e acessível aos estudantes de graduação (há texto didático dos autores, disponível no Instituto de Economia da UFRJ, com o título “Teoria dos jogos, organizações e contratos: uma introdução” que contempla estes tópicos). Para os estudantes que vêm demonstrando interesse crescente pelas abordagens evolucionistas e pelos jogos evolucionários, recomendamos fortemente recorrer a Binmore (1998), a Samuelson (1997) e a Young (1998).

Notas

1. De maneira mais geral, diz-se que um jogo é um jogo de *informação imperfeita* sempre que for recomendável incluir mais de um nó de decisão num mesmo “conjunto de informação”, o que estará sendo aqui atendido pela linha tracejada. A linha indica que o jogador II decide sem saber se a sua posição é a que se segue a E, ou a que se segue a B. No jogo de informação perfeita, o conjunto de informação é definido por apenas um nó. Cada um dos nós de decisão é um conjunto de informação e encerra uma história do jogo.

2. Mais genérica e formalmente, uma estratégia s_{ji} do jogador j será considerada estritamente dominante se, consideradas todas as outras estratégias do conjunto S_j , que serão denotadas por s'_{ji} , e para todos os possíveis perfis de estratégias selecionados pelos demais jogadores, ou seja, para todos os possíveis s_{-j} (pertencentes ao conjunto S_{-j}) se tiver $u_j(s_{ji}, s_{-j}) > u_j(s'_{ji}, s_{-j})$, valendo modificações recomendáveis para o caso de estratégias fracamente dominantes (em que a desigualdade estrita seria substituída por \geq) e de estratégias estrita e fracamente dominadas.(em que o sinal da desigualdade se inverte).
3. Para referências futuras, vale lembrar que, naquelas condições, um monopólio estaria decidindo pela produção de $q = 1/2$, como se adotasse regra conhecida de maximização de lucros que recomenda igualar a receita marginal, $\partial(p(q) \cdot q) / \partial q = 1 - 2q$, ao custo marginal, no caso igual ao custo unitário suposto nulo. Com isso o monopólio cobraria $p(q) = 1 - 1/2 = 1/2$, com lucro total de L_j , em que $L_j = p \cdot q - c \cdot q = 1/4$, $j = 1, 2$. Se a demanda estivesse sendo atendida por muitas firmas competindo em regime de *concorrência perfeita*, regra idêntica de maximização de lucros individuais estaria recomendando que as firmas produzissem $q = 1$, uma vez que para cada firma individual a demanda é infinitamente elástica em p .
4. Uma esclarecedora discussão sobre o tema pode ser encontrada em Kreps(1990c).
5. O resultado encontrado é caso particular de famoso *teorema de Zermelo* (há provas em Binmore, 1992, e em MWG, 1995, entre outros), que afirma que todo jogo finito com informação perfeita tem equilíbrio de Nash em estratégias puras. Na verdade, segundo Walker, 1995, o teorema de E. Zermelo é o primeiro teorema da teoria dos jogos (publicado em alemão em 1913), e afirma que o jogo de xadrez é estritamente determinado, no sentido de possuir apenas um equilíbrio em estratégias puras.
6. Em Kreps (1990c), em Milgrom e Roberts (1992) e em Binmore (1992, 1994 e 1998), o leitor encontrará elaborações mais autorizadas do material aqui apresentado.

Bibliografia

- Alt. J.; K. Shapsle, (orgs.) (1990) *Perspectives on Positive Political Economy*, Cambridge University Press.
- Bierman, H. S.; Fernandez, L. (1998) *Game Theory with Economic Applications*, Addison-Wesley Publishing Co.
- Binmore, K. (1992) *Fun and Games: A Text on Game Theory*, D.C. Heath and Company.
- Binmore, K. (1994) *Game Theory and the Social Contract I: playing fair*, The MIT Press.
- Binmore, K. (1998) *Game Theory and the Social Contract II: just playing*; The MIT Press.
- Binmore, K.; Dasgupta, P. (1986) *Economic Organizations as Games*, Basil Blackwell.
- Dixit, A.; Nalebuff, B. (1991) *Thinking Strategically*, W.W. Norton e Comp. (há tradução disponível, com o título *Pensando Estrategicamente*, editada pela Editora Atlas, 1994).
- Dixit, A.; Skeath, S. (1999) *Games of Strategy*, W.W. Norton & Comp.
- Fink, E. C.; Gates, S.; Humes, D. (1998) *Game Theory Topics: Incomplete Information, Repeated Games and n-Player Games*, Sage Publications.
- Fudenberg, D.; Tirole, J. (1986) “Noncooperative game theory”, in Schmalensee e Willig, orgs. HIO:260-327.
- Gibbons, R. (1992) *Game Theory for Applied Economists*, Princeton University Press.
- Kreps, D. (1990a) *Game Theory and Economic Modelling*, Clarendon Press.
- Kreps, D. (1990b) *A Course in Microeconomic Theory*, Princeton University Press.
- Kreps, D. (1990c) “Corporate culture and economic theory”, in J. Alt e Shapsle, orgs.:90-143.
- Macmillan, J. (1992) *Games, Strategies & Managers*, Oxford University Press.
- Mas-Colell, A.; Whinston, M.D.; Green, J. R. (1995) *Microeconomic Theory*, MWG, Oxford University Press.
- Milgrom, P.; Roberts, J. (1992) *Economics, Organization, and Management*, Prentice Hall.
- Myerson, R. B. (1991) *Game Theory: Analysis of Conflict*, Harvard University Press.
- Osborne, M. J.; Rubinstein, A. (1994) *A Course in Game Theory*, The MIT Press.
- Rasmusen, E. (1990) *Games and Information: An Introduction to Game Theory*, Basil Blackwell.
- Samuelson, L. (1997) *Evolutionary Games and Equilibrium Selection*, MIT Press.
- Schmalensee, R.; Willig, R. (orgs.) (1989) *Handbook of Industrial Organization- HIO*, North-Holland.
- Shapiro, C. (1986) “Theories of oligopoly behavior”, in Schmalensee e Willig, org., HIO:329-414.
- Straffin, P. D. (1993) *Game Theory and Strategy*, The Mathematical Association of America.
- Tirole, J. (1989) *The Theory of Industrial Organization*, MIT Press.
- Varian, H. (1993) *Microeconomic Analysis*, W.W. Norton & Comp.
- Walker, P. (1995) *An Outline of the History of Game Theory*, p.walker@ econ.canterbury.ac.nz;
- Young, P. (1998) *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton University Press.

Modelos de Concorrência em Oligopólio

Hugo Pedro Boff

10.1 Introdução

Iniciaremos este capítulo definindo vários conceitos fundamentais para a descrição e a análise da concorrência industrial. Ao mesmo tempo, introduziremos notações que serão utilizadas no tratamento formal do tema, apresentado nas seções seguintes.

As variáveis sobre as quais os produtores podem exercer um controle (*variáveis de decisão*) são as quantidades (q) ou os preços (p): Como o resultado obtido pela ação de um produtor individual é afetado pelas ações dos outros produtores, é possível presumir que cada um adotará um comportamento estratégico, isto é, tomará suas decisões levando em conta as possíveis reações dos outros às suas próprias ações. Isto faz com que p e/ou q sejam vistos como variáveis *estratégicas* pelos produtores.

A escolha de uma variável estratégica e a forma de considerar as possíveis interações existentes entre as ações individuais estabelecem os diversos regimes ou modelos de concorrência na indústria.

10.1.1 INTERAÇÕES ESTRATÉGICAS

Seja

$$Q = \sum_{j=1}^n q_j$$

a oferta total de uma indústria formada por n firmas; a i éssima firma oferta a quantidade q_i ao preço unitário p_i . Notamos q_{-i} (ou p_{-i}) para o vetor das quantidades (ou preços) da indústria privado do seu i éssimo componente q_i (ou p_i), de modo que $Q_{-i} = \sum_{j(j \neq i)} q_j = Q - q_i$, é a oferta agregada das empresas concorrentes do produtor i : Estes, serão denominados doravante empresas “rivais”.

Chamamos variação conjectural, notada por v_i a variação na oferta das empresas rivais antecipada por cada produtor, quando este decide alterar o nível da sua produção (ou preço).¹ Assim por exemplo, se os produtores competem em quantidades em um mercado de produto homogêneo, a variação conjectural do produtor i será:

$$v_i = \frac{dQ_{-i}}{dq_i}.$$

Em um mercado de produto heterogêneo (ou diferenciado), se os produtores competem em quantidades, a reação do produtor j antecipada pelo produtor i é:

$$\nu_{ij} = \frac{dq_j}{dq_i};$$

se os produtores competem em preços a reação antecipada é: $\nu_{ij} = \frac{dp_j}{dp_i}$.

Dependendo do valor atribuído pelos produtores à variação conjectural ν , diferentes regimes concorrenceis são gerados na indústria. Entretanto, nada garante que a reação do rival antecipada por cada produtor coincidirá com a ação realmente empreendida por ele. Quando a coincidência acontece e os prognósticos de cada um se cumprem, dizemos que as conjecturas são consistentes.²

10.1.2 MODELOS DE CONCORRÊNCIA

Assumimos neste capítulo que o principal objetivo das empresas é a maximização do lucro. A maneira de se escrever a função lucro depende da variável estratégica (q ou p); da natureza do produto (homogêneo ou diferenciado) e do horizonte de programação dos produtores. Esse último aspecto diz respeito à análise dinâmica da concorrência, a qual não será empreendida aqui.

Sendo C_i a função custo do produtor i , o Quadro 10.1 apresenta as funções lucro individuais $\pi_i = p_i q_i - C_i$ que o produtor i maximizará de acordo com a natureza do produto e a variável estratégica escolhida.

Os modelos de concorrência se especificam a partir das hipóteses adotadas com relação ao tempo e à existência ou não de cooperação entre as empresas.

Em um contexto estático no qual a dimensão temporal é ignorada (somente os efeitos instantâneos das decisões são considerados), os produtores tomam suas decisões simultaneamente. Focalizaremos neste capítulo, quase exclusivamente, padrões competitivos em mercados estáticos. Uma exceção é feita ao modelo de Stackelberg, que pressupõe dois momentos. Quando as decisões são tomadas *sequencialmente*, em uma determinada ordem preestabelecida, temos um modelo *dinâmico*.

Com relação às formas de considerar as possíveis interações entre as empresas, os regimes competitivos podem ser classificados como concorrenceis, colusivos e *mistas*.

Na primeira classificação, supomos que os produtores atuem de maneira não cooperativa, cada qual maximizando sua própria função-lucro (antecipando ou não a possível reação dos rivais). Os exemplos mais conhecidos são o regime de *competição perfeita*, o regime de Cournot, o regime de Bertrand e o regime de *competição monopolística*.

Na segunda classificação, todos os produtores atuam de maneira *cooperativa*, maximizando o lucro agregado da indústria. Aqui, a cooperação pode ser *explícita* (existe um contrato entre os produtores, prevendo punições para desvios individuais) ou apenas *tácita*. Este segundo caso inclui situações no qual comportamentos não cooperativos levam a resultados colusivos (quando existe informação imperfeita ou incerteza entre os produtores). Ambos os casos caracterizam a formação de *cartel*.

Em regimes *mistas*, uma ou várias coalizões atuam no mercado conjuntamente com produtores independentes (*franja concorrencial*).

Trataremos neste texto unicamente da concorrência perfeita e de três modelos de competição imperfeita: Cournot (variável estratégica q e decisão simultânea); Bertrand (variável estratégica p e decisão simultânea) e Stackelberg (variável estratégica p ou q e decisão sequencial). A análise dos regimes de monopólio, de competição monopolística, do cartel e dos regimes mistos é feita em outros capítulos deste livro (Capítulos 1, 6 e 12).

Além da introdução, este capítulo inclui três outras seções. Na próxima seção, introduziremos as funções de reação, um conceito central para a análise do desempenho das empresas no equilíbrio industrial. Na seção, são apresentadas as condições de primeira ordem para o oligopólio homogêneo e para o oligopólio diferenciado. Na terceira seção introduziremos as medidas

QUADRO 10.1 FUNÇÕES LUCRO SEGUNDO A VARIÁVEL DE CONTROLE E A NATUREZA DO PRODUTO

Produto	Variáveis estratégicas	
	q	p
Homogêneo	$q_i P(Q) - C_i(q_i)$	$p q_i(p) - C_i(q_i(p))$
Diferenciado	$q_i P_i(q_i; q_{-i}) - C_i(q_i)$	$p q_i(p_i; p_{-i}) - C_i(q_i(p_i; p_{-i}))$

do excedente econômico utilizadas para uma análise normativa dos diferentes regimes concorrenceis apresentados. A decomposição do excedente total no excedente do consumidor e do produtor é particularizada para cada empresa individualmente, nos casos em que o produto é homogêneo ou heterogêneo. A quarta seção analisa mais detalhadamente a atuação das empresas nos diferentes regimes competitivos, a partir das condições do equilíbrio industrial. A leitura vem acompanhada de resultados explícitos obtidos no caso do duopólio. Uma análise comparativa dos excedentes gerados em cada regime é feita ao final das subseções. Na conclusão, realizamos uma síntese comparativa dos resultados proporcionados pelos diferentes regimes concorrenceis enfocados no capítulo.

10.2 Funções de Reação e Equilíbrio

Definimos o equilíbrio concorrencial da indústria pelos vetores de quantidades e preços ($q_e; p_e$) que maximizam o lucro individual de cada produtor, isto é, tais que:

$$\pi_i(q_e, p_e) = \max_{q_i \text{ ou } p_i} \pi_i(q_i, p_i; q_{-ie}, p_{-ie}) \quad \forall i$$

Mediante a introdução de algumas hipóteses adicionais podemos determinar o equilíbrio da indústria de acordo com o regime de concorrência vigente. De modo geral, a existência de equilíbrios para a indústria requer que as funções lucro de cada produtor (π_i) sejam quase côncavas com relação à variável estratégica escolhida (q_i ou p_i). A unicidade do equilíbrio requer o atendimento de condições de segunda ordem mais fortes para estas funções.

Sob a hipótese de diferenciabilidade (até a segunda ordem) das funções lucro com relação a todos os seus argumentos, apresentamos a seguir as condições de primeira ordem,

$$\frac{\partial \pi_i}{\partial q_i} = 0 \quad \text{ou} \quad \frac{\partial \pi_i}{\partial p_i} = 0, \quad i = 1, 2, \dots, n$$

que devem ser satisfeitas para a existência do equilíbrio nos diferentes regimes de concorrência. As condições de primeira ordem engendram as chamadas funções de reação de cada produtor (ou *funções de resposta*) assumindo valores no conjunto das suas escolhas ótimas, para cada nível das variáveis de controle dos produtores rivais.

No caso, por exemplo, em que todas as empresas produzem um bem homogêneo e escolhem as quantidades como variável estratégica, a condição de primeira ordem para o produtor i é:

$$\frac{\partial \pi_i}{\partial q_i} = P + q_i P' [1 + \nu_i] - C'_i = 0 \quad (1a)$$

onde $\nu_i = \frac{dQ_{-i}}{dq_i}$ é a variação da oferta agregada dos rivais conjecturada pelo produtor i ao alterar seu nível de produção de equilíbrio.

No caso de produto *diferenciado* e competição em quantidades, a condição de primeira ordem do produtor i é:

$$\frac{\partial \pi_i}{\partial q_i} = p_i + q_i \sum_{j=1}^n \frac{\partial p_j}{\partial q_i} \nu_{ij} - C'_i = 0 \quad (1b)$$

onde $\nu_{ij} = \frac{dq_j}{dp_i}$ é a variação da oferta do produtor j antecipada pelo produtor i quando este altera a sua oferta de equilíbrio.

Se a variável estratégica for o *preço*, a condição de primeira ordem é:

$$\frac{\partial \pi_i}{\partial p_i} = q_i + (p_i - C'_i) \sum_{j=1}^n \frac{\partial q_j}{\partial p_i} \nu_{ij} = 0 \quad (1c)$$

onde $\nu_{ij} = \frac{dp_j}{dp_i}$ é a variação no preço do produto j conjecturada pelo produtor i quando este altera seu preço de equilíbrio.

10.2.1 ESTABILIDADE DO EQUILÍBRIO

Seja x o vetor (com n componentes x_i) das variáveis de controle dos produtores ($x = p$ ou $x = q$) e notemos $\mu_i(x) = \frac{\partial \pi_i}{\partial x_i}$ para o lucro marginal auferido pelo produtor i com uma mudança na sua variável de controle.

Se \dot{x}_i é a derivada de x_i com relação ao tempo, a equação $\dot{x}_i = s_i \mu_i(x)$ fornece uma descrição (míope) da dinâmica da variável x_i a partir dos desvios do equilíbrio, sendo $s_i > 0$ a velocidade dos ajustes. Evidentemente, no equilíbrio temos: $s_i \mu_i(x_e) = 0$. Para escrever as n equações na forma vetorial, notemos $D_s = \text{Diag}[s_1; s_2; \dots; s_n]$ para a matriz diagonal dos coeficientes de ajustamento e $\mu(x) = [\mu_1(x), \mu_2(x), \dots, \mu_n(x)]$. Vale então a equação $\dot{x}_i = D_s \mu_i(x)$ para o sistema dinâmico completo.

Uma maneira usual de se efetuar um estudo local da estabilidade do sistema acima, consiste em se proceder a uma aproximação linear da equação dinâmica ao redor de $x = x_e$. Dado que neste ponto temos $\dot{x}_i(x_e) = 0$, vem:

$$\dot{x}_i \cong D_s H_\pi (x - x_e)$$

onde $H_\pi = [h_{ij}]$ é a matriz de ordem n das segundas derivadas das funções lucro individuais, avaliadas no equilíbrio industrial:

$$h_{ij} = \frac{\partial \mu_i(x_e)}{\partial x_j} = \frac{\partial^2 \pi_i(x_e)}{\partial x_i \partial x_j}, \quad i, j = 1, 2, \dots, n.$$

A estabilidade do mercado na vizinhança do equilíbrio exige que se $x \gg x_e$, então, $\dot{x} \ll 0$, o que requer que a matriz $D_s H_\pi$ seja definida negativa.

No caso $n = 2$ (duopólio), para que a definição negativa de $D_s H_\pi$ seja satisfeita é necessário (e suficiente) que o traço desta matriz seja estritamente negativo e seu determinante estritamente positivo, isto é: $s_1 \frac{\partial^2 \pi_1(x_e)}{\partial x_1^2} + s_2 \frac{\partial^2 \pi_2(x_e)}{\partial x_2^2} < 0$ e

$s_1 s_2 \left[\frac{\partial^2 \pi_1(x_e)}{\partial x_1^2} \cdot \frac{\partial^2 \pi_2(x_e)}{\partial x_2^2} - \frac{\partial^2 \pi_1(x_e)}{\partial x_1 \partial x_2} \cdot \frac{\partial^2 \pi_2(x_e)}{\partial x_2 \partial x_1} \right] > 0$. Como estas desigualdades devem ser verificadas para qualquer s_1 e s_2 , as

condições

$$(C1) \quad \frac{\partial^2 \pi_1(x_e)}{\partial x_1^2} < 0; \quad \frac{\partial^2 \pi_2(x_e)}{\partial x_2^2} < 0;$$

$$(C2) \quad \frac{\partial^2 \pi_1(x_e)}{\partial x_1^2} \cdot \frac{\partial^2 \pi_2(x_e)}{\partial x_2^2} - \frac{\partial^2 \pi_1(x_e)}{\partial x_1 \partial x_2} \cdot \frac{\partial^2 \pi_2(x_e)}{\partial x_2 \partial x_1} > 0$$

são necessárias e suficientes no caso de duopólio.

10.2.2 FUNÇÕES DE REAÇÃO

Seja x_i a variável estratégica do produtor i . A função de reação do produtor, notada $x_i = R_i(x_{-i})$, fornece a melhor resposta do produtor i ao vetor estratégico x_{-i} dos rivais. Elas são obtidas resolvendo-se as equações (1a), (1b) ou (1c); conforme o caso, na variável q_i ou p_i . Obviamente, as quantidades (preços) de equilíbrio da indústria situam-se na *interseção* das superfícies (ou curvas) de reação, de modo que temos $q_{ie} = R_i(q_{-ie})$ ou $p_{ie} = R_i(p_{-ie})$, para todo i .

As superfícies de reação podem ser analisadas mediante a derivação direcional das funções de reação ou a diferenciação total das equações (1a), (1b), (1c). Uma representação simples das curvas de reação no plano $q_i \times q_j$ ou ($p_i \times p_j$) é obtida no caso em que dois produtores disputam o mercado (*duopólio*). Nesse caso do duopólio, a inclinação da curva de reação $R_i(q_j)$ ou $R_i(p_j)$ do produtor i (notada $r_i = R'_i$) é obtida mediante diferenciação da condição de primeira ordem. Sua expressão geral é função das derivadas de primeira e de segunda ordem da função lucro do produtor: $\pi_i(x_i; x_j)$.

Diferenciando-se totalmente a condição de primeira ordem ($\partial \pi_i / \partial x_i = 0$) do programa de maximização do lucro, obtemos: $(\partial^2 \pi_i / \partial x_i^2) dx_i + (\partial^2 \pi_i / \partial x_i \partial x_j) dx_j = 0$, o que leva à seguinte derivada da curva de reação $x_i = R_i(x_j)$ do produtor i :

$$r_i \equiv \frac{dx_i}{dx_j} = -\frac{\partial^2 \pi_i / \partial x_i \partial x_j}{\partial^2 \pi_i / \partial x_i^2} \quad (2)$$

Para facilidade da análise, trataremos aqui as variações conjecturais (v) como constantes. Isto implica que as funções de reação antecipadas pelos produtores sejam lineares. O cálculo da expressão (2) levará em conta a hipótese mencionada, introduzida apenas para simplificar a análise.

10.2.2.1 Duopólio homogêneo

Identificando dR_i com dq_i obtemos a partir de (1a):

$$[P' + (1 + \nu_i)(P' + q_i P'') - C''_i] dR_i + [P' + (1 + \nu_i)q_i P''] dq_j = 0$$

ELSEVIER

Desta equação (sempre avaliada no equilíbrio), obtemos a seguinte inclinação dR_i/dq_j da tangente à curva de reação $R_i(q_j)$ do produtor i :

$$r_i = -\frac{P' + (1 + \nu_i)q_i P''}{[P' + (1 + \nu_i)(P' + q_i P'')] - C_i''] \quad (3a)}$$

Obtemos a inclinação da tangente à curva de reação $R_j(q_i)$ do produtor j permutando-se os índices i e j na equação anterior.

10.2.2.2 Duopólio diferenciado

a) Variável estratégica *quantidade*:

A diferenciação de (1b) leva à:

$$\left[2p'_i + \nu_i \left(\frac{\partial p_i}{\partial q_j} + q_i \frac{\partial^2 p_i}{\partial q_i \partial q_j} \right) + q_i p''_i - C_i'' \right] dR_i + \left[\frac{\partial p_i}{\partial q_j} + \nu_i q_i \frac{\partial^2 p_i}{\partial q_j^2} + q_i \frac{\partial^2 p_i}{\partial q_i \partial q_j} \right] dq_j = 0$$

A derivada dR_i/dq_j da curva de reação do produtor i será então:

$$r_i = -\frac{\left[\frac{\partial p_i}{\partial q_j} + \nu_i q_i \frac{\partial^2 p_i}{\partial q_j^2} + q_i \frac{\partial^2 p_i}{\partial q_i \partial q_j} \right]}{\left[2p'_i + \nu_i \left(\frac{\partial p_i}{\partial q_j} + q_i \frac{\partial^2 p_i}{\partial q_i \partial q_j} \right) + q_i p''_i - C_i'' \right]} \quad (3b)$$

onde p'_i e p''_i designam as derivadas primeira e segunda da demanda inversa com relação às quantidades q_i .

b) Variável estratégica *preço*:

Identificando dR_i com dq_i ; a diferenciação de (1c) leva à seguinte derivada da curva de reação do produtor i :

$$r_i = -\frac{\left[\frac{\partial q_i}{\partial p_j} (1 - C_i'' q'_i) + \nu_i \left\{ \frac{\partial^2 q_i}{\partial p_j^2} (p_i - C_i') - C_i'' \left(\frac{\partial q_i}{\partial p_j} \right)^2 \right\} + (p_i - C_i') \frac{\partial^2 q_i}{\partial p_i \partial p_j} \right]}{\left[q'_i (2 - C_i'' q'_i) + \nu_i \left\{ \frac{\partial q_i}{\partial p_j} (1 - C_i'' q'_i) + \frac{\partial^2 q_i}{\partial p_i \partial p_j} (p_i - C_i') \right\} + (p_i + C_i') q''_i \right]} \quad (3c)$$

onde q'_i e q''_i designam as derivadas primeira e segunda da demanda direta com relação ao preço p_i . Aqui também a inclinação da tangente à curva de reação do produtor j no equilíbrio é obtida permutando-se os índices i e j em (3b) ou (3c).

10.2.3 CONJECTURAS CONSISTENTES

Convém distinguir as reações expressas pelas curvas $R_i(q_j)$ das reações expressas por ν_{ij} . As primeiras são respostas ótimas efetivas de um dos produtores, face à oferta do outro. As outras são conjecturadas, na determinação do equilíbrio, e somente podem ser efetivas quando consistentes.

No caso do duopólio, temos $\nu_{ij} \equiv \nu_i; \nu_{ji} \equiv \nu_j$, e a consistência exigirá que a resposta efetiva de um dos produtores seja plenamente antecipada pelo outro, isto é:

$$r_i = \nu_j; i, j = 1, 2 \quad (4)$$

O sinal das derivadas (3a) – (3c) nos informa sobre o modo ótimo de reagir de um produtor face a desvios do equilíbrio nas quantidades ofertadas ou no preço praticado pelo outro.

10.2.4 SUBSTITUIÇÃO E COMPLEMENTARIDADE ESTRATÉGICAS

Seja um duopólio e x_i, x_j as variáveis de controle dos oligopolistas, as quais podem ser p ou q . As variáveis x_i e x_j são ditas *substitutas estratégicas* se $\partial^2\pi_i/\partial x_i \partial x_j < 0$, ($i \neq j = 1, 2$) isto é, se o lucro marginal de cada produtor decresce quando o rival aumenta o nível da sua variável de controle. Inversamente, elas são ditas *complementares estratégicas* quando $\partial^2\pi_i/\partial x_i \partial x_j > 0$, isto é, o lucro marginal do produtor cresce quando o rival aumenta o nível de sua variável de controle.

Note que esta derivada cruzada figura no numerador de (2) e que a negatividade do seu denominador ($\partial^2\pi_i/\partial x_i^2$) é uma condição necessária para que o equilíbrio do duopólio seja estável (condição C1).

Temos então que a estabilidade do equilíbrio requer que as curvas de reação sejam negativamente inclinadas se as variáveis forem substitutas estratégicas e positivamente inclinadas se as variáveis forem complementares estratégicas.

Frisemos que estas definições e propriedades se aplicam tanto às situações em que as variáveis de controle são os preços quanto àquelas em que as variáveis de controle são as quantidades.

10.2.5 BENS E ESTRATÉGIAS

Devemos distinguir a substituição existente *entre os bens* (expressas por $\partial p_i/\partial q_j < 0$ na competição em quantidades, ou por $\partial q_i/\partial p_j > 0$ na competição em preços), da substituição *entre estratégias*, pois uma não implica a outra.

Como variáveis de controle, as quantidades podem, por exemplo, ser substitutas estratégicas, ainda que os bens sejam complementares. Com efeito, a expressão $\partial^2\pi_i/\partial q_i \partial q_j = \partial p_i/\partial q_j + q_i(\partial^2 p_i/\partial q_i \partial q_j)$ pode ser negativa, com os bens sendo complementares: $\partial p_i/\partial q_j > 0$. As mesmas observações são válidas para a complementaridade estratégica.

Todavia, no duopólio com demanda linear (doravante, duopólio linear), os efeitos de segunda ordem sobre a demanda são nulos, de maneira que a substituição ou complementaridade dos bens e das variáveis estratégicas são conceitos equivalentes.

O desempenho industrial depende da tecnologia disponível (custos), da conduta dos produtores (variáveis estratégicas e conjecturas) e do comportamento da demanda (preferências). Neste último aspecto em particular, veremos, à frente, como a atuação de empresas envolvidas em um mesmo regime de concorrência poderá apresentar um desempenho bastante distinto, se as estratégias (ou os bens) forem substitutos ou complementares.

As equações (1a) – (1c) e as expressões (3a) – (3c) serão particularizadas na Seção 10.4, quando iremos explorar os diversos regimes de competição contemplados neste capítulo.

10.3 Excedente Econômico

O excedente econômico total (ET) gerado pela atividade produtiva compõe-se de duas parcelas. A parcela dos consumidores, chamada *excedente do consumidor* (EC) e a parcela dos produtores, chamada *lucros* (Π):

$$ET = EC + \Pi.$$

Para o cálculo do excedente econômico, os efeitos de variações na renda dos consumidores sobre o seu consumo serão ignorados.

10.3.1 INDÚSTRIA COM PRODUTO HOMOGENE

Para uma indústria com produto *homogêneo*, atendendo à demanda agregada representada pela função de demanda $p = P(x)$, o excedente econômico gerado pelo produtor i pode ser aproximado pelo *excedente Marshalliano* (W_i) definido aqui por:

$$W_i = \int_0^{q_i} [P(x + Q_{-i}) - C'_i(x)] dx \quad (5a)$$

onde C'_i é o custo marginal da empresa i ; q_i a oferta atual da empresa i .

O excedente econômico total é:

$$ET = \sum_{i=1}^n W_i$$

sendo $Q = \sum_{i=1}^n q_i$ e $p = P(Q)$ o preço atual. Podemos então escrever $P - C'_i = P - p + p - C'$; de maneira que W_i pode ser escrito como: $\int_0^{q_i} [P(x + Q_{-i}) - p] dx + \int_0^{q_i} [p - C'_i(x)] dx$. Da primeira parcela obtemos o excedente do consumidor:

$$EC = \sum_{i=1}^n \int_0^{q_i} [P(x + Q_{-i}) - p] dx \quad (5b)$$

e da segunda parcela obtemos o lucro agregado na indústria:

$$\Pi = \sum_{i=1}^n \int_0^{q_i} [p - C'_i(x)] dx = pQ - \sum_{i=1}^n [C_i(q_i) - C_i(0)] \quad (5c)$$

A Figura 10.1 ilustra a decomposição do excedente marshalliano gerado pela produção da empresa i .

10.3.2 INDÚSTRIA COM PRODUTO DIFERENCIADO

Em um oligopólio *diferenciado*, a empresa i faz face à demanda inversa $P_i(q_i; q_{-i})$, a qual supõe-se positiva e contínua sobre um poliedro limitado do \mathbb{R}_+^n , diferenciável com relação aos seus argumentos e decrescente em q_i .

Notamos W_{ij} para o excedente econômico gerado pela demanda do bem i por meio de variações na oferta do bem j (para $i \neq j$; é o efeito substituição).

Esta quantidade é calculada como:

$$W_{ii} = \int_0^{q_i} [P_i(x; q_{-i}) - \phi'_i(x)] dx$$

$$W_{ij} = \int_0^{q_i} P_i(x; q_{-j}) dx - \phi'_i(q_i) q_j; i \neq j = 1, 2, \dots, n.$$

de maneira que a contribuição da empresa i para o excedente econômico será dado por:

$$W_i = \sum_{j=1}^n W_{ij}$$

Para o excedente do consumidor os valores correspondentes são calculados por:

$$W_{ij}^c = \int_0^{q_i} P_i(x; q_{-i}) dx - P_i q_j; (i, j = 1, 2, \dots, n)$$

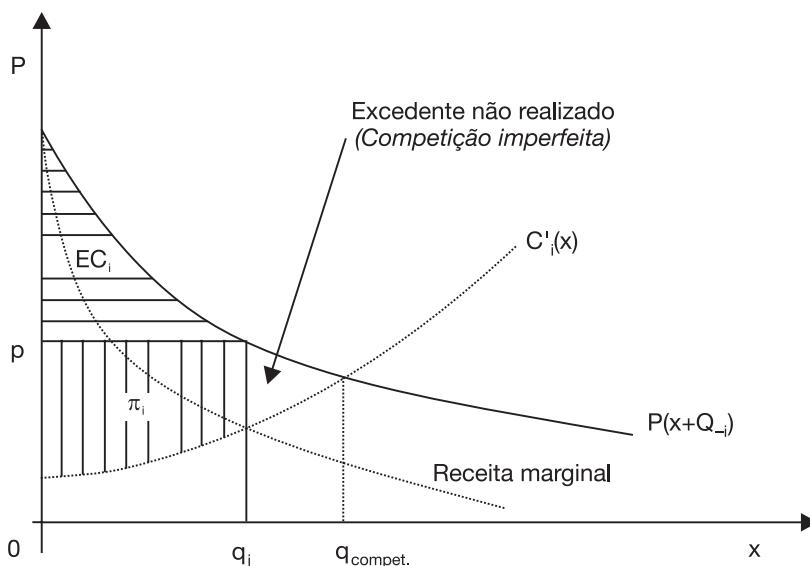


Figura 10.1 Excedente Marshalliano. EC: excedente do consumidor; π : excedente do produtor (lucro).

Entretanto, a análise de bem-estar habitual para o duopólio diferenciado não requer o uso direto das expressões anteriores, uma vez que são focalizados apenas os excedentes agregados. Nestes casos, o procedimento usual (mais simples) consiste em se avaliar as variações no excedente total medindo este pela utilidade do vetor da produção corrente, $q = [q_1, q_2, \dots, q_n]$, a partir da qual são geradas as demandas inversas P_i

$$ET = U(q).$$

As variações no *excedente do consumidor* serão então capturadas a partir das variações na diferença entre a utilidade e o lucro agregado:

$$EC = U(q) - \Pi(q) \quad (5d)$$

onde $\Pi(q) = \sum_{i=1}^n \pi_i(q)$ é o lucro agregado, o qual é igual à soma dos lucros das empresas. Evidentemente, U é uma função arbitrária, mas estritamente crescente nas quantidades (q_1, q_2).

10.4 Regimes de Concorrência

10.4.1 REGIME DE COMPETIÇÃO PERFEITA

No regime de competição perfeita, consideramos que os produtores ofertam um produto homogêneo e competem em quantidades. Em uma indústria perfeitamente competitiva, cada produtor presume que toda variação na sua oferta não altera significativamente o preço do produto no mercado. Olhando-se para a equação (1a), esta premissa é equivalente à conjectura do produtor i de que os efeitos de preço de um aumento na sua oferta $dq_i > 0$ são neutralizados por uma redução igual na oferta dos rivais $dQ_{-i} = -dq_i$. Assim, o preço é considerado pelos produtores como uma constante determinada exogenamente.

Colocando então $v_i = -1$ em (1a) obtemos a condição de equilíbrio, preço = custo marginal:

$$p = C'_i \quad (6a)$$

Logo, os custos marginais das empresas são equalizados no equilíbrio da indústria ($q_e; p_e$).

Se a empresa i tiver custo marginal mais baixo que a empresa j ($C'_i < C'_j$), produzirá no equilíbrio quantidades mais elevadas ($q_{ie} > q_{je}$). Dado que não existe interação estratégica entre as empresas, as curvas de reação entre os dois produtores são retas que se cruzam perpendicularmente, definidas por: $q_{ie} = C_i'^{-1}(p)$ e $q_{je} = C_j'^{-1}(p)$, como mostra a Figura 10.2.

Note que uma variação conjectural nula ou negativa ($v_i \leq 0$) corresponde à hipótese de uma reação “acomodativa”; face a um aumento na sua oferta, o produtor i conjectura que os rivais diminuirão suas quantidades ofertadas. O equilíbrio puramente competitivo corresponde ao caso de acomodação perfeita.

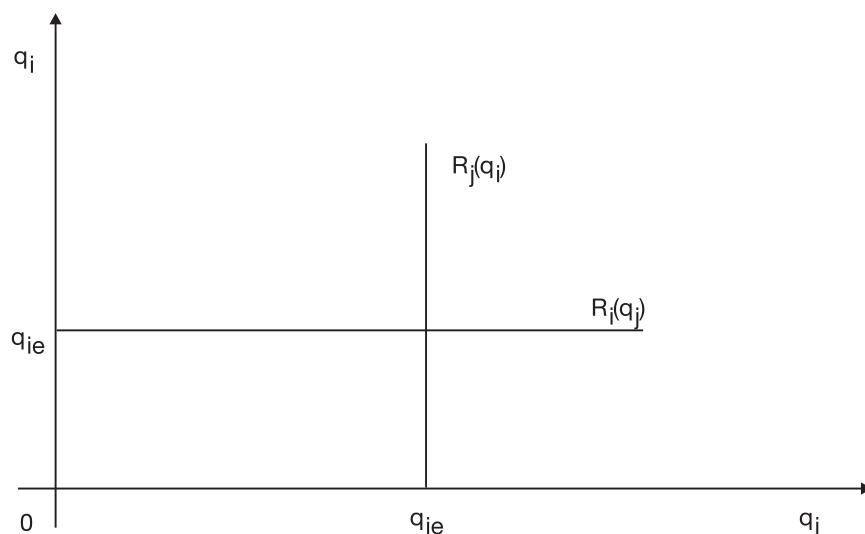


Figura 10.2 Curvas de reação. Competição perfeita.

No plano do bem-estar, o regime de competição perfeita é o que maximiza o excedente econômico total (ET) e o excedente dos consumidores (EC), definidos em (5a) – (5c). Na Figura 10.1, o excedente total gerado por cada empresa em competição perfeita inclui também o triângulo indicado pela seta, o qual representa a *perda de bem-estar* associada à competição imperfeita isto é, ao fato de a empresa não dimensionar a oferta para igualar seu custo marginal ao preço de mercado.

10.4.2 REGIME DE COURNOT (A.A.COURNOT, 1801-77)

O regime concorrencial de Cournot corresponde à competição em quantidades onde cada produtor presume que uma variação na sua oferta *não é antecipada* pelos rivais. As variações conjecturais são nulas.

Se o produto for *homogêneo*, colocamos $v_i = dQ_{-i}/dq_i = 0$ na equação (1a), de maneira que a condição de primeira ordem do produtor i torna-se:

$$P + q_i P' - C'_i = 0; \quad i = 1, 2, \dots, n. \quad (7a)$$

A maximização do lucro ocorre onde a receita marginal ($P + q_i P'$) iguala o custo marginal (ver Figura 10.1). Como esta é menor que o preço P , a oferta de equilíbrio no regime de Cournot (q_{ie}^C) é menor que a de equilíbrio em concorrência perfeita (q_{ie}^c): $q_{ie}^C < q_{ie}^c$. Em consequência, o preço do equilíbrio de Cournot será mais elevado que o do equilíbrio em concorrência perfeita $p^C > p^c$. Por outro lado, se os custos marginais forem não decrescentes, a equação (7a) mostra que os custos totais da indústria não são minimizados, pois os custos marginais não são igualados no equilíbrio (salvo no caso simétrico).

Do ponto de vista social, o regime de Cournot é dominado pelo regime de competição perfeita, visto que este último regime maximiza o bem-estar, isto é, o excedente total e o excedente dos consumidores. Entretanto, o regime de Cournot é preferível do ponto de vista privado pois, tipicamente, proporciona lucros mais elevados para as empresas individualmente.

Notando-se $\lambda_i \equiv \frac{P - C'_i}{P}$ para a margem preço-custo marginal (*índice de Lerner*), $s_i = \frac{q_i}{Q}$ para a parcela de mercado da empresa i e $\varepsilon = \frac{\partial Q}{\partial p} \cdot \frac{p}{Q}$ para o módulo da elasticidade-preço da demanda direta (todas as grandezas sendo avaliadas no ponto de equilíbrio), podemos escrever a equação (7a) como:

$$\lambda_i = \frac{s_i}{\varepsilon} \quad (8a)$$

Com $v_i = 0$, a inclinação da tangente à curva de reação $R_i(q_j)$ do produtor i no duopólio homogêneo (dada em (3a)) simplifica-se em:

$$r_i = -\frac{P' + q_i P''}{2P' + q_i P'' - C''_i} \quad (9a)$$

Pela condição (4) da consistência das conjecturas, a condição $r_1 = r_2 = 0$ somente será verificada se o numerador de (9a) for nulo, o que requererá demanda convexa ($P'' = -P' q_i \geq 0$). Uma condição suficiente para que a conjectura de Cournot seja consistente é que P seja constante.

Com efeito, é possível mostrar que no duopólio linear homogêneo, com custo marginal constante, a única conjectura (constante) consistente na competição em quantidades é: $v_1 = v_2 = -1$ (conjectura de competição perfeita).

As condições (C1) e (C2) para a estabilidade do equilíbrio dão, neste caso:

$$2P' + q_i P'' - C''_i < 0 \quad i = 1, 2 \text{ e} \quad (i)$$

$$(2P' + q_i P'' - C''_i)(2P' + q_j P'' - C''_j) - (P' + q_i P'')(P' + q_j P'') > 0 \quad (ii)$$

Depois de efetuadas as simplificações possíveis na última expressão, percebe-se que a desigualdade $P' + q_i P'' \leq 0$ é necessária para que a condição (ii) seja atendida.

Conjuntamente com (i), vem que as condições:

$$C''_i \geq 0 \text{ e } P' + q_i P'' \leq 0,$$

são suficientes para a estabilidade do equilíbrio de Cournot.

A expressão $P' + q_i P''$ é também a derivada da receita marginal do produtor i com relação a q_j . Logo, a estabilidade do equilíbrio requer que os aumentos na oferta do rival reduzam a receita marginal do produtor. A partir de (9a), note que a estabilidade do equilíbrio implica que a curva de reação seja negativamente inclinada ($r_i < 0$). Neste caso, os bens i e j serão substitutos estratégicos, porque $\partial^2 p_i / \partial q_i \partial q_j = P' + q_i P'' \leq 0$: a oferta adicional do rival reduz o lucro marginal do produtor.

Se o produto for diferenciado, colocamos $v_{ij} = dq_j / dq_i = 0$ (para todo $j \neq i$), e obtemos a partir de (1b):

$$p_i + q_i \frac{\partial p_i}{\partial q_i} - C'_i = 0 \quad (7b)$$

Notando $\frac{1}{\varepsilon_i^C} = -\frac{\partial p_i}{\partial q_i} \cdot \frac{q_i}{p_i}$ para a elasticidade da demanda inversa pelo bem i e notando λ_i^C o índice de Lerner avaliado no equilíbrio de Cournot, obtemos o equivalente de (8a):

$$\lambda_i^C = \frac{1}{\varepsilon_i^C} \quad (8b)$$

Com $v_i = 0$, a inclinação da tangente à curva de reação do produtor i no duopólio heterogêneo (dada em (3b)) simplifica-se em:

$$r_i^C = -\frac{\left[\frac{\partial p_i}{\partial q_j} + q_i \frac{\partial^2 p_i}{\partial q_i \partial q_j} \right]}{\left[2p'_i + q_i p''_i - C''_i \right]} \quad (9b)$$

Aqui, a consistência da conjectura de Cournot requer que o numerador de (9b) se anule. Então, se a demanda for linear, os bens devem ser independentes ($\partial p_i / \partial q_j = 0$).

É possível mostrar que o equilíbrio do duopólio diferenciado linear, com custo marginal constante, admite conjecturas (constantes) consistentes, cujo valor depende dos parâmetros da demanda.

As condições (C1) e (C2) para a estabilidade do equilíbrio dão, neste caso:

$$2p'_i + q_i p''_i - C''_i < 0 \quad i = 1, 2 \text{ e} \quad (i)$$

$$(2p'_i + q_i p''_i - C''_i)(2p'_j + q_j p''_j - C''_j) - \left(\frac{\partial p_i}{\partial q_j} + q_i \frac{\partial^2 p_i}{\partial q_i \partial q_j} \right) \left(\frac{\partial p_j}{\partial q_i} + q_j \frac{\partial^2 p_j}{\partial q_j \partial q_i} \right) > 0 \quad (ii)$$

Em face de (9b), vemos que a estabilidade do equilíbrio de Cournot é compatível com qualquer inclinação das curvas de reação. Por exemplo, o equilíbrio pode ser instável ainda que as curvas de reação tenham inclinação negativa. Tal acontece quando a condição (i) é violada e os bens forem complementos estratégicos, isto é, $\partial p_i / \partial q_j + q_i \partial^2 p_i / \partial q_i \partial q_j > 0$: o aumento na oferta do rival j aumenta o lucro marginal do produtor i .

10.4.2.1 Demandas lineares

Uma situação frequentemente considerada em análises teóricas é a das *demandas lineares* ($P'' = 0$ ou $p''_i = p''_j = \partial^2 p_i / \partial q_i \partial q_j = 0$, no caso heterogêneo) e custos marginais *constantes* ($C''_i = 0$). Nestas situações, as curvas de reação $R_i(q_j)$ são lineares e, como vimos, a substituição ou complementaridade estratégicas e dos bens são equivalentes.

O duopólio *homogêneo* é sempre estável, e a derivada da curva de reação é: $r_i^C = r_j^C = -\frac{1}{2}$, de maneira que as curvas de reação serão negativamente inclinadas.

Dado que $p'_i, p'_j < 0$, o *duopólio diferenciado* é estável se

$$4p'_i p'_j - (\partial p_i / \partial q_j)(\partial p_j / \partial q_i) > 0.$$

As demandas estão bem definidas quando os efeitos de *substituição* (Slutsky) são simétricos, o que requer $(\partial p_i / \partial q_j) = (\partial p_j / \partial q_i) = -\gamma$ (digamos).

Como vemos, a estabilidade do duopólio não depende da substituição ou da complementaridade existente entre os bens. A derivada da reta de resposta reduz-se neste caso a: $r_i^C = \gamma / 2p'_i$.

A inclinação das retas dependerá da relação de substituição ou complementaridade. Se os bens forem substitutos ($\gamma > 0$), a inclinação será negativa; se eles forem complementares ($\gamma < 0$) a inclinação será positiva.

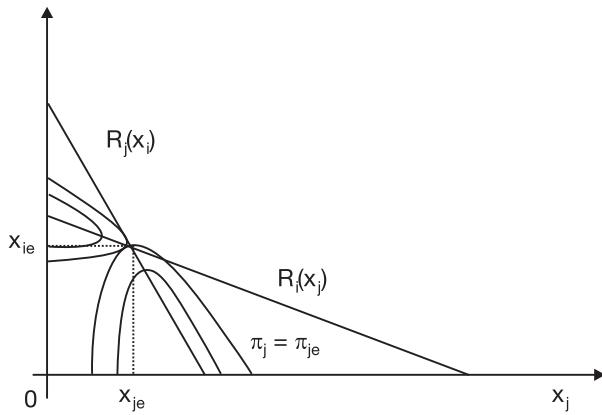


Figura 10.3A Curvas de reação e de isolucro no duopólio linear. $x = q$: Cournot (bens substitutos); $x = p$: Bertrand (bens complementares).

A Figura 10.3, no plano $x_j \times x_i$ ilustra as duas situações, para a variável estratégica quantidades ($x = q$) ou preços (x).

Geometricamente, as curvas de reação R_i formam o *locus* dos pontos q (ou p) tais que $d\pi_i = 0$. Cada ponto sobre a curva é interceptado por uma *curva de isolucro* do produtor, a qual é definida pela equação implícita: $\pi_i(q_i; q_j) = k$; $k \geq 0$.

Se os bens forem *substitutos*, as curvas de isolucro serão côncavas, como indicado na Figura 10.3A. Neste caso, quanto mais elevada a curva, menor é o nível de lucro da empresa. Na Figura 10.3A desenhamos duas curvas de isolucro (duas para cada produtor) interceptadas pelas curvas de reação respectivas. Se os bens forem *complementares*, as curvas de isolucro serão convexas, como indicado na Figura 10.3B.

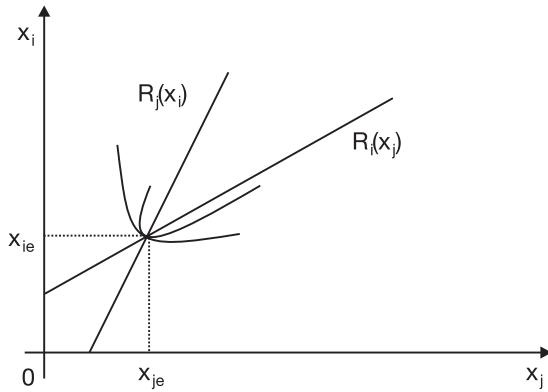


Figura 10.3B Curvas de reação e de isolucro no duopólio linear. $x = q$: Cournot (bens complementares); $x = p$: Bertrand (bens).

No equilíbrio (x_{je}, x_{ie}) , onde as curvas de reação se interceptam, as curvas de isolucro $\pi_i(x_i, x_j) = \pi_{ie}$; ($i, j = 1, 2$) são tangentes, uma à outra.

Para compreendermos o porquê da inclinação das retas de resposta representadas na Figura 10.3, consideramos que, no caso de produtos *perfeitamente substitutos*, face a um aumento na oferta do rival, o preço diminui e o produtor i evitará uma redução nos seus lucros retraindo a oferta, dado que sua demanda residual diminui: a preço menor, corresponde oferta menor. Se os dois produtos forem *substitutos imperfeitos*, a retração da demanda pelo seu produto causada pelo aumento da oferta do produto corrente e a subsequente queda no seu preço, levará o produtor i a ajustar para menos sua oferta, a fim de manter seu nível de lucro.

Enfim, se os dois produtos forem *complementares*, o aumento na demanda residual do produtor i causada pelo aumento na oferta do rival eleva o preço do seu produto: a preço maior corresponde oferta maior.

10.4.3 REGIME DE BERTRAND (J. BERTRAND, 1822-1900)

O regime concorrencial de Bertrand corresponde à competição em preço na qual cada produtor presume que alterações no seu preço não são antecipadas pelos rivais.

Se o produto for homogêneo e se todas as empresas operarem com retornos de escala constantes na produção (e não tiverem restrições de capacidade), a empresa i com menor custo marginal (c_i) monopolizará o mercado, oferecendo o produto a qualquer preço p tal que $c_i \leq p < c_j$, onde c_j é o custo marginal da segunda empresa mais eficiente. Naturalmente, o produtor i maximizará seu lucro (ver o Quadro 10.1) fixando o preço logo abaixo de c_j : $p_{ie} = c_j - \epsilon$; $0 < \epsilon < c_j - c_i$. Não terá interesse em tomar $\epsilon = 0$, pois a empresa j , ao fixar seu preço em c_j , poderá então operar reduzindo o lucro da empresa i . Quando $c_i = c_j = c$, o preço de equilíbrio será c ; as duas empresas dividirão o mercado e ambas terão lucro zero.

Se o produto for diferenciado, colocamos $v_{ij} = dp_j / dp_i = 0$ (para todo $j \neq i$); e obtemos a partir de (1c):

$$q_i + (p_i - C'_i) \frac{\partial q_i}{\partial p_i} = 0 \quad (7c)$$

Notando $\varepsilon_i^B = -\frac{\partial q_i}{\partial p_i} \cdot \frac{p_i}{q_i}$ para a elasticidade da demanda direta pelo bem i , obtemos o equivalente de (8b):

$$\lambda_i^B = \frac{1}{\varepsilon_i^B} \quad (8c)$$

A expressão (8c) é análoga à (8b) do regime de Cournot. Não são idênticas, pois os valores de equilíbrio nos dois regimes não são (em geral) os mesmos.

Todavia, revelam um aspecto comum: no equilíbrio da indústria, *os produtores obtêm margens unitárias de lucro tanto maiores quanto menor for a elasticidade-preço das suas demandas residuais*.

Com $v_i = 0$, a inclinação da tangente à curva de reação $p_i = R_i(p_j)$ do produtor i no duopólio diferenciado (dada em (3c)) simplifica-se em:

$$r_i^B = -\frac{\left[(1 - C''_i q'_i) \frac{\partial q_i}{\partial p_j} + (p_i - C'_i) \frac{\partial^2 q_i}{\partial p_i \partial p_j} \right]}{\left[(2 - C''_i q'_i) q'_i + (p_i - C'_i) q''_i \right]} \quad (9c)$$

Face à condição (4), a consistência da conjectura de Bertrand requer que o numerador de (9c) se anule. Se a demanda pelo produto i for linear nos preços, esta consistência exige custos marginais decrescentes ($C''_i = 1/q'_i < 0$) ou bens independentes ($\partial q_i / \partial p_j = 0$) muito embora a primeira condição não seja suficiente.

A condição (C1) para a estabilidade do equilíbrio dá, neste caso:

$$(2 - C''_i q'_i) q'_i + (p_i - C'_i) q''_i < 0; \quad i = 1, 2$$

Esta condição é atendida se $C''_i \geq 0$ e $q''_i \leq 0$. A condição (C2) é complexa, mas pode ser expressa mais simplesmente como: $r_i^B r_j^B < 1$, onde r_j^B é definida com (9c), permutando-se o índice i com j .

Logo, para que o equilíbrio do duopólio diferenciado de Bertrand seja estável, é suficiente que os custos marginais sejam não decrescentes ($C''_i \geq 0$), as curvas de demanda sejam côncavas nos preços ($q''_i \leq 0$) e as curvas de reação suavemente inclinadas ($r_i^B r_j^B < 1$).

Em termos de estratégias, lembremos que o equilíbrio de Bertrand é estável se as curvas de reação são positivamente inclinadas quando os preços são complementos estratégicos ($\partial^2 \pi_i / \partial p_i \partial p_j$ positivo), e negativamente inclinadas quando os preços são substitutos estratégicos ($\partial^2 \pi_i / \partial p_i \partial p_j$ negativo).

10.4.3.1 Demandas lineares

Se os duopolistas fazem face a demandas *lineares* e incorrem em custos marginais *não decrescentes*, r_i^B terá o sinal de $\frac{\partial q_i}{\partial p_j}$.

Assim, se os bens i e j forem *substitutos* ($\partial q_i / \partial p_j > 0$), as curvas de reação nos preços serão positivamente inclinadas: $r_i^B > 0$. Se os bens forem *complementares* ($\partial q_i / \partial p_j < 0$), as curvas de reação serão negativamente inclinadas: $r_i^B < 0$.

Tomando-se $x = p$, a Figura 10.3 mostrada na seção anterior ilustra o equilíbrio de Bertrand para o duopólio linear com custo marginal constante no espaço dos preços $p_j \times p_i$. As curvas de reação $p_i = R_i(p_j)$ são lineares com inclinação $r_i^B = -(\partial q_i / \partial p_j) / 2q'_i$ constante ($i = 1, 2$).

Pela Figura 10.3B vemos que o equilíbrio de Bertrand com bens substitutos imperfeitos (preços complementos estratégicos) é o dual do equilíbrio de Cournot com bens complementares. A Figura 10.3A mostra que o equilíbrio de Bertrand com bens complementares (preços substitutos estratégicos) é o dual do equilíbrio de Cournot com bens substitutos (imperfeitos).

Obviamente, os valores de equilíbrio obtidos nos dois regimes não são (em geral) idênticos ($q_e^C \neq q_e^B$ e $p_e^C \neq p_e^B$). Entretanto, no cálculo operacional, pode-se passar de um ao outro com a permutação das quantidades pelos preços e a substituição dos parâmetros da demanda inversa pelos parâmetros correspondentes da demanda direta.

10.4.4 BERTRAND × COURNOT

Para compararmos a performance do duopólio nos regimes de Bertrand e Cournot, vamos usar a hipótese de que, dados os preços p_1, p_2 e a renda m , o consumidor representativo dotado de preferências quase lineares escolhe as quantidades dos bens que maximizam a função de utilidade quadrática:

$$U(q_i, q_2) = q_0 + \alpha_1 q_1 + \alpha_2 q_2 - \frac{1}{2}(\beta_1 q_1^2 + \beta_2 q_2^2 + 2\gamma q_1 q_2) \quad (10)$$

sob a restrição orçamentária $q_0 + p_1 q_1 + p_2 q_2 \leq m$; onde q_0 é um bem numerário. As demandas inversas ótimas resultantes desta otimização, para cada bem, são dadas por:

$$p_i = \alpha_i - \beta_i q_i - \gamma q_j; \quad i, j = 1, 2 \quad (11a)$$

Sob a hipótese $\alpha_i \beta_j - \alpha_j \gamma > 0$ ($i \neq j$), o *equilíbrio em Cournot* está definido para $\delta = \beta_1 \beta_2 - \gamma^2 > 0$. Se $\alpha_1 = \alpha_2$, o termo $d = \gamma^2 / \beta_1 \beta_2$ é visto como um *grau de diferenciação* dos produtos, variável entre 0 (bens independentes) e 1 (substitutos perfeitos).

Como nosso propósito é comparar dois padrões concorrenenciais distintos, podemos supor, sem perda de generalidade, que as empresas incorrem em custos idênticos e nulos. A confrontação dos dois regimes em duopólio realizada a seguir acompanha de perto a exposição feita por Singh e Vives citados acima.

Colocando $\Delta = 4\beta_1 \beta_2 - \gamma^2$, e usando as equações (7b), o equilíbrio em Cournot dá:

$$q_{ie}^C = (2\alpha_i \beta_j - \alpha_j \gamma) / \Delta \quad (12a)$$

$$p_{ie}^C = \beta_i q_{ie}^C \quad (13a)$$

$$\pi_i^C = \beta_i (q_{ie}^C)^2 \quad (14a)$$

Para calcularmos o *equilíbrio em Bertrand*, fazemos uso das demandas diretas:

$$q_i = a_i - b_i p_i + c p_j; \quad i, j = 1, 2 \quad (11b)$$

Os parâmetros a_i, b_i e c_i exprimem-se em função dos parâmetros α_i, β_i e γ , tomando-se o inverso do sistema (11a) o que dá:

$$a_i = (\alpha_i \beta_j - \alpha_j \gamma) / \delta; \quad b_i = \beta_j / \delta; \quad c = \gamma / \delta.$$

Colocando $D = 4b_1 b_2 - c_2$; e usando as equações (7c) obtemos os seguintes valores no equilíbrio em Bertrand:

$$p_{ie}^B = (2a_i b_j + a_j c) / D \quad (12b)$$

$$q_{ie}^B = b_i p_{ie}^B \quad (13b)$$

$$\pi_i^B = b_i (p_{ie}^B)^2 \quad (14b)$$

Note que $D = \Delta / \delta^2$.

A substituição dos valores de a, b e c dados acima nas equações (12b) – (14b) do equilíbrio em Bertrand permite uma comparação direta com os valores correspondentes do equilíbrio em Cournot. Em particular, obtemos:

$$p_{ie}^C - p_{ie}^B = \frac{\alpha_i}{(4/d) - 1} > 0 \text{ e}$$

$$q_{ie}^B - q_{ie}^C = \frac{\gamma^2 (\alpha_i \beta_j - \alpha_j \gamma)}{\delta \Delta} \geq 0.$$

Vemos que os preços no equilíbrio em Bertrand são menores e as quantidades maiores do que no equilíbrio em Cournot, independentemente do grau de substituição ($\gamma > 0$) ou de complementariedade ($\gamma < 0$) existente entre os bens.

Assim, o duopólio de Cournot aparece como mais colusivo que o de Bertrand. A razão está no fato de que, dado o preço do rival, ao escolher o seu nível de preço, o produtor i percebe uma elasticidade da demanda mais elevada (coeficiente angular = b_i) que aquela percebida quando escolhe as quantidades no modelo de Cournot (coeficiente angular = $1/\beta_i = b_i - c^2/b_j < b_i$). Em consequência, no regime de Bertrand, os aumentos de preço acima do custo marginal são mais suaves do que no regime de Cournot.

O excedente econômico total ET é avaliado pela utilidade $U(q_{1e}, q_{2e})$ explicitada em (10). Como U é crescente e côncava nas quantidades, vemos que $ET^B > ET^C$. Por outro lado, o excedente dos consumidores, calculado usando-se (5d), é uma função decrescente e convexa dos preços, de modo que $EC^B > EC^C$. Assim, do *ponto de vista social*, o padrão competitivo de Bertrand domina o padrão de Cournot.

Do *ponto de vista privado*, sendo $\pi_i(q_{1e}, q_{2e})$ o lucro da empresa i ($i = 1, 2$), é possível mostrar que o resultado da comparação entre os dois regimes depende da relação existente entre os bens. Se os bens forem *substitutos* ($\gamma > 0$) temos $\pi_i^C > \pi_i^B$, de maneira que as empresas obtêm, individualmente, lucros maiores ao competirem em quantidades. Se os bens forem *complementares* ($\gamma < 0$), então $\pi_i^B > \pi_i^C$ e a competição em preço entre ambas é dominante. Enfim, se os bens forem *independentes* ($\gamma = 0$), o lucro auferido pela empresa é idêntico em ambos os regimes, de modo que a escolha do padrão competitivo torna-se irrelevante para as empresas.

Os resultados obtidos para o duopólio linear com retornos constantes, no qual as empresas concordam em escolher a mesma variável de controle (q ou p), são consistentes com aqueles obtidos em um jogo em *dois estágios*, no qual cada produtor escolhe primeiramente a variável estratégica e, em seguida, compete com o rival escolhendo o nível da variável eleita que maximiza o seu lucro [Singh & Vives, 1984]. Neste caso, são considerados também os equilíbrios dos mercados nos quais os produtores elegem variáveis de controle distintas (q, p) ou (p, q).

Não surpreendentemente, a estratégia dominante de ambos os produtores prevê, no primeiro estágio, a escolha das quantidades (q, q) se os bens forem substitutos, e dos preços (p, p) se os bens forem complementares.

Resumindo, no caso de *substituição entre os bens*, temos um conflito entre o interesse privado e o interesse social, pois, como vimos, o padrão de Cournot proporciona lucros maiores, mas menor excedente econômico que aquele gerado no padrão de Bertrand. Se os bens forem *complementares*, os interesses coincidem, pois, neste caso, o padrão de Bertrand leva a maiores excedentes, tanto para os produtores como para os consumidores. Se os bens forem *independentes*, os dois regimes são equivalentes.

A passagem do duopólio de Cournot para o duopólio de Bertrand é, portanto, *Pareto-eficiente* (há ganhos sem que ninguém perca) somente se os produtos forem *complementares*.

10.4.4.1 Oligopólio

Todavia, os resultados descritos anteriormente são sensíveis à hipótese de duopólio adotada. No caso de um número maior de empresas ($n = 3, 4, \dots$), não é possível resumir as condições para a dominância de um regime sobre o outro unicamente em função da substituição ou da complementaridade existente entre os bens. Introduzindo-se as simetrias $\beta_i = 1$; $\gamma_i = \gamma$ ($i = 1, 2, \dots, n$) entre as empresas a partir da utilidade quadrática como em (10), e mantendo-se a hipótese dos custos marginais constantes e idênticos ($= 0$), a diferença existente entre os parâmetros α ($\alpha_i \neq \alpha_j$) assinalará diferenciais de qualidade (vertical) existente entre os bens. Os seguintes resultados são então obtidos:³

- Se os bens são substitutos, a relação $p_i^C - p_i^B > 0$ se mantém, de modo que o excedente dos consumidores será maior com a competição em preços. Entretanto, se as diferenças de qualidade entre os produtos forem elevadas, as empresas i que ofertam produtos de maior qualidade [$\alpha_i > \sum_{j(j \neq i)} \alpha_j / (n-1)$] podem auferir lucros maiores no regime de Bertrand do que no regime de Cournot. Assim, a dominância deste último regime não fica estabelecida, *do ponto de vista privado*;
- Se os bens são complementares, a relação $\pi_i^B > \pi_i^C$ se mantém, de maneira que o excedente dos produtores será maior na competição em preços. Entretanto, se as diferenças de qualidade entre os produtos forem elevadas, as empresas que ofertam produtos de baixa qualidade podem vir a praticar preços maiores no regime de Bertrand do que no regime de Cournot. Em consequência, a dominância do primeiro regime não está assegurada, *do ponto de vista social*.

10.4.5 REGIME DE STACKELBERG (H. von STACKELBERG, 1934)

O regime de Stackelberg pressupõe decisões sequenciais em dois estágios. No primeiro, o produtor 1 (chamado *líder*) toma sua decisão escolhendo q_1 (ou p_1) para maximizar seu lucro. Na segunda etapa, os outros $n - 1$ produtores (chamados *seguidores*) efetuam suas melhores escolhas (q_j ou p_j) individualmente, dada a escolha do líder.

Para facilidade da exposição, trataremos o caso do duopólio ($n = 2$). O equilíbrio de modelos sequenciais é obtido por indução retroativa.⁴ Assim, o líder (que move primeiro) considerará a função de melhor resposta do seguidor $x_2 = R_2(x_1)$, obtida da condição de primeira ordem relativa à maximização do lucro do seguidor.

Aqui, temos $v_2 = 0$, com $x = q, p$.

10.4.5.1 Liderança em quantidades

No modelo de liderança em quantidades (*quantity leadership*), o líder maximizará em q_1 a função lucro:

$$\pi_1(q_1; R_2(q_1)) = q_1 p_1(q_1; R_2(q_1)) - C_1(q_1).$$

onde $q_2 = R_2(q_1)$ é a função de reação do seguidor, obtida resolvendo-se implicitamente, em q_2 , a condição de primeira ordem (7a) no caso homogêneo, ou (7b) no caso diferenciado.

Note que se o produto for *homogêneo*, a demanda inversa p_1 escreve-se: $P(q_1 + R_2(q_1))$.

A condição de primeira ordem para o líder é:

$$P + q_1 P' [1 + R'_2(q_1)] - C'_1 = 0 \quad (\text{caso homogêneo});$$

$$p_1 + q_1 \left[\frac{\partial p_1}{\partial q_1} + \frac{\partial p_1}{\partial q_2} R'_2(q_1) \right] - C'_1 = 0 \quad (7d) \quad (\text{caso diferenciado}).$$

Comparando-se a primeira destas equações com (1a) e a segunda com (1b) vemos que a variação conjectural (consistente) do líder é:

$$R'_2(q_1) = - \left(\frac{\partial^2 \pi_2}{\partial q_2 \partial q_1} \right) \left/ \left(\frac{\partial^2 \pi_2}{\partial q_2^2} \right) \right.$$

Vamos analisar o caso do *duopólio diferenciado*. O duopólio homogêneo aparece como um caso particular do primeiro, quando existe substituição perfeita entre os bens.

Se o equilíbrio for estável e os bens forem *substitutos estratégicos*, então $\frac{\partial^2 \pi_2}{\partial q_2 \partial q_1} < 0$ e $R'_2(q_1) < 0$, de modo que o líder antecipa corretamente que o seguidor “acomodará” os aumentos de oferta, reduzindo a sua.

Se os bens forem *complementos estratégicos*, então $\frac{\partial^2 \pi_2}{\partial q_2 \partial q_1} > 0$ e $R'_2(q_1) > 0$ e o líder considera a atitude “mimética” do seguidor, o qual replicará os aumentos ou contrações da sua oferta.

Em ambos os casos, se os efeitos de segunda ordem não forem dominantes, deveremos ter $\frac{\partial p_1}{\partial q_2} R'_2 > 0$; de maneira que, em face da condição de primeira ordem apresentada anteriormente, a receita marginal do líder será maior do que a percebida por este produtor em competição de Cournot.

Assim, o líder sempre produzirá quantidades maiores do que o faria no regime de Cournot: $q_1^S > q_1^C$.

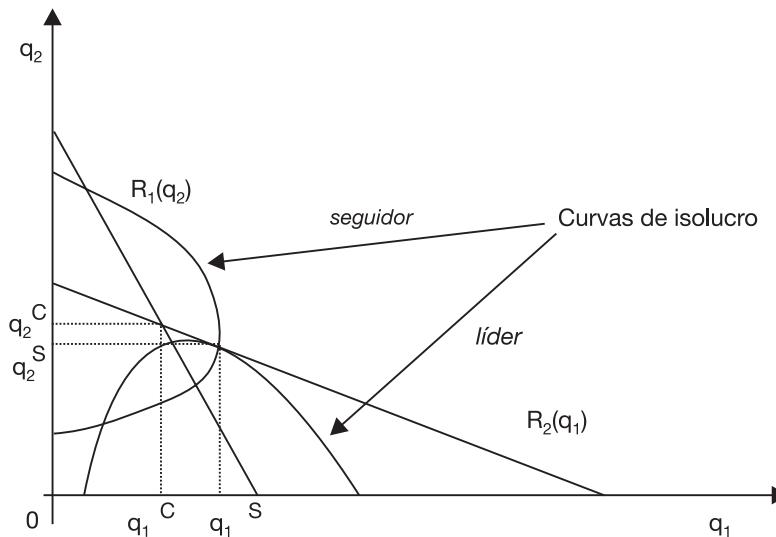
Quanto ao *seguidor*, como sua função de resposta é decrescente num caso e crescente no outro, segue-se que ele ofertará quantidades menores se os bens forem *substitutos* ($q_2^S < q_2^C$) e maiores se os bens forem *complementares* ($q_2^S > q_2^C$). Em ambos os casos, os lucros do líder são maiores do que obteria na competição independente: $\pi_1^S > \pi_1^C$. Isto é ilustrado graficamente nas Figuras 10.4 e 10.5, que mostram o equilíbrio de Stackelberg (*liderança em quantidades*) no duopólio linear quando os bens são substitutos e complementares, respectivamente.

Um argumento simples confirma este fato: como o líder seleciona seu nível de produção ótimo sobre a curva de resposta do seguidor, escolheria o nível de Cournot se este lhe proporcionasse lucro mais elevado. Assim, qualquer que seja a relação entre os bens, o lucro do líder é maior do que aquele que obteria em competição de Cournot. Para o seguidor, o argumento geométrico das Figuras 10.4 e 10.5 mostra que, com relação ao regime de Cournot, seu lucro será menor do que em Cournot se os bens forem *substitutos* ($\pi_2^S < \pi_2^C$) e maior se os bens forem *complementares* $\pi_2^S > \pi_2^C$, respectivamente.

O fato de o lucro da empresa líder ser maior que π_1^C aparece pelo nível da sua curva de isolucro (tangente à R_2) na Figura 10.4, a qual se situa em um nível mais baixo da curva passando pela interseção de R_2 com R_1 (não representada), que é o equilíbrio de Cournot. Por outro lado, a curva de isolucro da seguidora passando pelo equilíbrio de Cournot (não representada), é mais baixa que aquela que passa pelo equilíbrio de Stackelberg e, por isso, $\pi_2^S < \pi_2^C$.

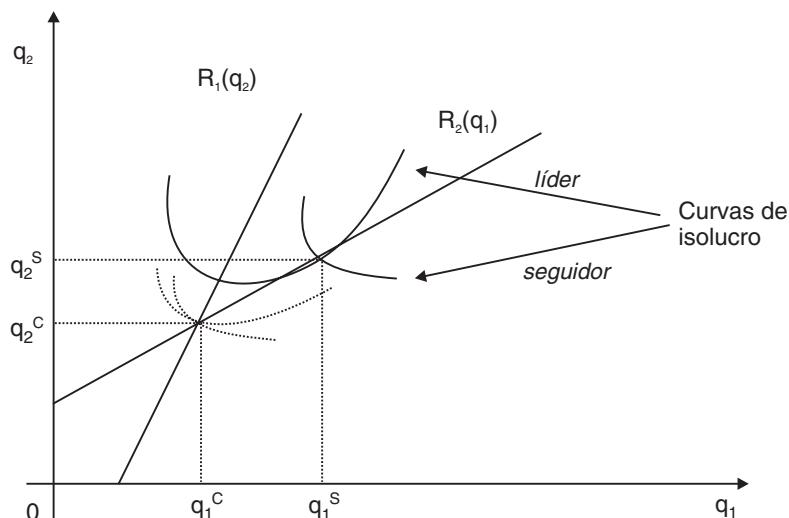
Com bens complementos estratégicos, as curvas de isolucro são convexas e, quanto mais elevadas, maiores os níveis de lucro. Na Figura 10.5, as curvas passando pelo equilíbrio de Stackelberg são superiores àquelas (pontilhadas) passando pelo equilíbrio de Cournot. Assim, ambos os produtores auferem lucros mais elevados. A Figura 10.5 também ilustra o fato de que ambos os produtores ofertarão quantidades também mais elevadas neste caso: $q_1^S > q_1^C$ e $q_2^S > q_2^C$. Obviamente, a oferta agregada será maior: $Q^S > Q^C$.

A resolução do equilíbrio para o caso do duopólio linear diferenciado e simétrico permite que se obtenha uma ilustração algébrica deste resultado. Retomemos esse caso, apresentado na Seção 10.4.3 (com custos zero). O uso das condições de primeira ordem (7b) para a seguidora e (7d) para a líder leva às seguintes quantidades e preços de equilíbrio:



empresa 1: líder; empresa 2: seguidora.

Figura 10.4 Equilíbrio de Stackelberg com bens substitutos (liderança de quantidades). Empresa 1: Líder; Empresa 2: seguidora.



empresa 1: líder; empresa 2: seguidora.

Figura 10.5 Equilíbrio de Stackelberg com bens complementares (liderança em quantidade). Empresa 1: Líder; empresa 2: seguidora.

$$q_1^S = \frac{\alpha}{4} \left(\frac{4\beta - 2\gamma}{2\beta^2 - \gamma^2} \right) \quad (\text{empresa líder});$$

$$q_2^S = \frac{\alpha}{4} \left(\frac{4\beta - 2\gamma - \gamma^2/\beta}{2\beta^2 - \gamma^2} \right) \quad (\text{empresa seguidora});$$

$$p_1^S = \frac{\alpha}{4} \left(\frac{4\beta^2 - 2\beta\gamma - 2\gamma^2 + \gamma^3/\beta}{2\beta^2 - \gamma^2} \right) \quad (\text{empresa líder});$$

$$p_2^S = \frac{\alpha}{4} \left(\frac{4\beta^2 - 2\beta\gamma - \gamma^2}{2\beta^2 - \gamma^2} \right) \quad (\text{empresa seguidora}).$$

Notemos que se os bens forem independentes, então $q_1^S = q_2^S$. Se os bens forem substitutos ou complementares ($\gamma \neq 0$), teremos $q_1^S > q_2^S$. Por outro lado, como $\beta > \gamma$, o preço praticado pela seguidora será maior que aquele praticado pela empresa líder: $p_2^S > p_1^S$.

Colocando-se $\alpha_i = \alpha$; $\beta_i = \beta$ nas equações (12a) – (14a), podemos efetuar a comparação direta do equilíbrio em Stackelberg com o equilíbrio em Cournot no caso simétrico. Os resultados obtidos são sumariados a seguir:

- a) Com *bens substitutos*, a oferta da indústria é maior do que no equilíbrio simétrico de Cournot ($Q^S > Q^C$) e o preço praticado pelo líder é menor que o praticado pelo seguidor; ambos são menores que o preço no equilíbrio de Cournot: $p_1^S < p_2^S < p^C$.

No duopólio *simétrico* ($\pi_1^C = \pi_2^C = \pi^C$), o lucro da empresa líder é maior do que o da seguidora. Temos, de fato: $\pi_1^S > \pi^C > \pi_2^S$. A dominância da posição de líder é explicada pelo fato dela mover-se primeiro. Se as empresas escolhem capacidades de produção, e a seguidora for proponente ao ingresso no mercado, a dominância da posição de líder é vista como um *prêmio auferido pelo fato de ser pioneira*.

- b) Com *bens complementares*, a oferta da empresa líder é maior que a da seguidora. Temos: $q_1^S > q_2^S > q^C$. Já o preço praticado pela seguidora é maior que o preço em Cournot, o qual é maior que o preço praticado pelo líder: $p_2^S > p^C > p_1^S$. Para os lucros, temos a seguinte hierarquia: $\pi_2^S > \pi_1^S > \pi^C$. Surpreendentemente, a seguidora aufera lucros maiores que a empresa líder. Aqui, o fato de mover-se em segundo é favorável à seguidora. Como o lucro marginal da seguidora cresce com as ofertas adicionais da líder, o fato de fixar sua oferta após a líder lhe permite ajustar as quantidades para elevar o preço a ponto de auferir lucro superior ao da empresa líder.

10.4.5.2 Excedente econômico: Stackelberg × Cournot

Referindo-nos ao duopólio diferenciado linear e simétrico, vimos que a oferta agregada em Stackelberg é sempre maior que a oferta em Cournot: $Q^S > Q^C$. Assim, como a função de utilidade é crescente nas quantidades, o padrão de Stackelberg, com liderança em quantidades, socialmente domina o padrão de Cournot. O mesmo pode ser dito com relação ao ponto de vista dos consumidores, pois o seu excedente é função decrescente do nível de preços.

Do ponto de vista privado, se os bens forem *substitutos*, a empresa líder preferirá o regime de Stackelberg e a seguidora o regime de Cournot. Se os bens forem *complementares*, ambas preferirão o regime de Stackelberg, mas a líder desejará ser seguidora. Neste caso, nenhuma das empresas tem incentivo em mover primeiro.

Resumindo, a passagem do regime de Cournot para o regime de Stackelberg (socialmente preferível) é *Pareto-eficiente* somente se os bens forem *complementares*, pois, neste caso, empresas e consumidores, no conjunto, são beneficiados. Se os bens forem substitutos, esta passagem é lesiva à empresa seguidora, que vê seu lucro diminuir.

A passagem inversa, do regime de Stackelberg para o de Cournot, também não é *Pareto-eficiente*, pois, neste caso, a líder será a empresa prejudicada.

10.4.5.3 Liderança em preço

No modelo de liderança em preço (*price leadership*) e produto *diferenciado*, o líder maximizará em p_1 a função lucro $\pi_1(p_1; R_2(p_1))$; obtida substituindo-se p_2 pelo valor da reação da seguidora:

$$\pi_1 = p_1 q_1(p_1; R_2(p_1)) - C_1(q_1(p_1; R_2(p_1))).$$

A condição de primeira ordem é:

$$q_1 + (p_1 - C'_1) \frac{\partial q_1}{\partial p_1} [1 + R'_2(p_1)] = 0 \quad (7e)$$

Comparando esta equação com (1c), vemos que a variação conjectural consistente do líder é $v_1 = R'_2$.

Se o equilíbrio for estável e os preços praticados pelos produtores forem *complementos estratégicos*, então $\frac{\partial^2 \pi_2}{\partial p_2 \partial p_1} > 0$ e $R'_2(p_1) > 0$, de maneira que o líder sabe que seus aumentos de preço não serão combatidos, mas acompanhados pelo seguidor. (Lembre-se de que no padrão de Bertrand, a conjectura da empresa é de que seus aumentos de preço não serão seguidos: $v_1 = 0$). Como a receita marginal proporcionada ao líder por um aumento do preço é maior do que no padrão de Bertrand, o preço de equilíbrio do líder será maior: $p_1^S > p_1^B$.

O seguidor define seu preço após o líder e, como sua função de resposta é crescente, fixará seu preço também acima do nível de Bertrand: $p_2^S > p_2^B$.

Tal resultado pode ser visualizado pela Figura 10.5, substituindo-se quantidades por preços. Pelo argumento racional da lucratividade, deduzimos que o lucro do líder é maior do que aquele que ele obteria no regime de Bertrand.

Todavia, no caso *simétrico*, o lucro do líder será menor que o da seguidora, porque o lucro marginal desta cresce com os aumentos de preço efetuados pelo líder, de maneira que temos, tipicamente, $\pi_2^S > \pi_1^S > \pi^B$.

Resultado, assim, paradoxal, já encontrado na liderança em quantidades com bens *complementares estratégicos*, pode ser mais bem entendido se os bens forem substitutos (tal será o caso se os efeitos de segunda ordem sobre a demanda não forem dominantes): o seguidor ajustará seu preço para mantê-lo abaixo de p_1^S , para estimular a demanda pelo seu produto, e assim auferir lucro superior ao do líder.

O cálculo algébrico do equilíbrio em Stackelberg para o duopólio diferenciado, linear e simétrico confirma este *rationale*.

Com efeito, retomemos o duopólio linear diferenciado e simétrico (com custos zero) apresentado na Seção 10.4.3. O uso das condições de primeira ordem (7c) para a seguidora e (7e) para a líder leva às seguintes quantidades e preços de equilíbrio:

$$\begin{aligned} p_1^S &= \frac{\alpha}{4} \left(\frac{4b + 2c}{2b^2 - c^2} \right) && \text{(empresa líder);} \\ p_2^S &= \frac{\alpha}{4} \left(\frac{4b + 2c - c^2/b}{2b^2 - c^2} \right) && \text{(empresa seguidora);} \\ q_1^S &= \frac{\alpha}{4} \left(\frac{4b^2 + 2bc - 2c^2 - c^3/b}{2b^2 - c^2} \right) && \text{(empresa líder);} \\ q_2^S &= \frac{\alpha}{4} \left(\frac{4b^2 + 2bc - c^2}{2b^2 - c^2} \right) && \text{(empresa seguidora).} \end{aligned}$$

Note que se os bens forem independentes ($c = 0$), então $p_1^S = p_2^S$. Se os bens forem substitutos ou complementares ($c \neq 0$), o preço praticado pela empresa líder será superior ao preço adotado pela seguidora: $p_1^S > p_2^S$. Por outro lado, como $\beta > |\gamma|$ a empresa seguidora ofertará quantidades maiores: $q_2^S > q_1^S$.

Colocando-se $a_i = a$; $b_i = b$ nas equações (12b)–(14b), podemos efetuar a comparação direta do equilíbrio em Stackelberg com o equilíbrio em Bertrand no caso simétrico. Os resultados obtidos são sumariados a seguir:

- a) Com *preços complementares estratégicos* (bens substitutos), é fácil verificar que os preços serão maiores do que em Bertrand. Temos, com efeito, a seguinte hierarquia: $p_1^S > p_2^S > p^B$ e, para as quantidades, $q_2^S > q^B > q_1^S$. Os lucros auferidos pela empresa seguidora serão maiores que os da empresa líder. Ambos são superiores aos lucros obtidos em Bertrand: $\pi_2^S > \pi_1^S > \pi^B$;
- b) Se os preços praticados pelos produtores forem *substitutos estratégicos* (bens complementares), então $\frac{\partial^2 \pi_2}{\partial p_2 \partial p_1} < 0$ e $R'_2(p_1) < 0$, de maneira que o líder antecipa corretamente que seus aumentos de preço serão sancionados por reduções de preço do seguidor. Em princípio, o líder deveria fixar seu preço abaixo do preço em Bertrand, já que sua receita marginal é menor do que o seria naquele regime. Todavia, aumentos no seu nível de preço forçam a empresa seguidora a manter seu preço em nível mais baixo, o que eleva o lucro marginal do líder.

Se os efeitos de segunda ordem não são importantes, as perdas de receita do líder (decorrentes da retração na demanda causada pela alta do seu preço) serão atenuadas pela receita induzida com o aumento na demanda do seguidor, dada a complementaridade existente entre os bens. Portanto, no equilíbrio simétrico teremos: $p_1^S > p^B > p_2^S$ (o que pode ser visualizado substituindo-se a variável q por p na Figura 10.5).

O lucro dos produtores é função crescente do preço, de modo que deveremos obter, neste caso, a seguinte hierarquia de resultados para as empresas: $\pi_1^S > \pi^B > \pi_2^S$. Os resultados obtidos no duopólio linear e simétrico confirmam esta interpretação. A oferta da empresa líder é menor que a da seguidora e ambas são menores que a oferta que obteriam no equilíbrio simétrico em Bertrand: $q^B > q_2^S > q_1^S$.

10.4.5.4 Excedente econômico: Stackelberg × Bertrand

Referindo-nos ao caso do duopólio diferenciado linear e simétrico, os resultados obtidos mostram que a oferta agregada do duopólio, com liderança em preço, é menor que a oferta no duopólio de Bertrand em ambos os casos, com *substitutos*, ou com bens *complementares* (preços substitutos estratégicos): $Q^B > Q^S$.

Como a oferta agregada é aqui um indicador suficiente para se avaliar o bem-estar, vemos que, contrariamente ao mercado com liderança em quantidades, a liderança em preços não é socialmente desejável no caso do oligopólio linear.

Do ponto de vista social temos, neste caso a hierarquia: Bertrand \succ Stackelberg \succ Cournot.

Do ponto de vista *privado*, se os bens forem *substitutos*, tanto a empresa líder como a seguidora preferirão o regime de Stackelberg. Todavia, nenhuma das empresas desejará fixar o preço em primeiro lugar, pois o lucro da líder é menor que o da seguidora. Se os bens forem *complementares*, apenas a empresa líder preferirá o regime de Stackelberg. A empresa seguidora prefere ser independente e competir em Bertrand.

Assim, a passagem do regime de Bertrand ao regime de Stackelberg não é *Pareto-eficiente* pois os consumidores serão lesados. A passagem inversa, do regime de Stackelberg para o de Bertrand, também não é *Pareto-eficiente* pois, neste caso, a líder será a empresa prejudicada.

10.4.5.5 Produto homogêneo

Quando o custo marginal da empresa seguidora não for constante, também será possível analisar o modelo competitivo de Stackelberg com liderança em preços no caso de uma indústria com produto *homogêneo*. Os produtos ofertados pelas empresas são substitutos perfeitos um do outro.

Neste caso, devemos ter: $p_2 = p_1 = p$ (o seguidor vende ao preço fixado pelo líder), de maneira que a curva de oferta do seguidor será determinada por:

$$C'_2(q_2) = p \quad (\text{equilíbrio competitivo, com preço dado}).$$

O líder considerará a *demanda residual*:

$$q_1(p) = Q(p) - C'^{-1}_2(p)$$

e escolherá o preço p para maximizar o lucro:

$$\pi_2(p) = p \{Q(p) - C'^{-1}_2(p)\} - C_1(Q(p) - C'^{-1}_2(p)).$$

Como neste caso os bens são substitutos perfeitos, ambas as empresas auferirão lucros maiores (ou iguais) aos do regime de Bertrand (sem restrição de capacidade). No caso de *custos simétricos*, o lucro da empresa líder será maior do que o da seguidora.

10.5 Conclusão: Análise do Bem-estar nos Diferentes Regimes Competitivos

Resumimos agora os resultados da análise do bem-estar obtidos a partir dos modelos de liderança em quantidades e em preços de Stackelberg, quando comparados com aqueles obtidos nos modelos de competição independente das Seções 10.4.2 e 10.4.3 Cournot e Bertrand, respectivamente.

Na Seção 10.4.4 mostramos que o regime de Bertrand domina socialmente o regime de Cournot, na medida em que provê, no equilíbrio, uma maior oferta de bens: $Q^B > Q^C$. Em seguida, mostramos que, na competição em preços, o regime de Bertrand também domina o regime de Stackelberg: $Q^B > Q^S$ -preços. Na competição em quantidades, pelo seu lado, o regime de Stackelberg domina o regime de Cournot: Q^S -quant. $> Q^C$.

Assim, a competição independente em preços (Bertrand) aparece como o melhor regime competitivo do *ponto de vista social*, e a competição independente em quantidades (Cournot) aparece como o pior. O regime de Stackelberg, com competição em preços ou em quantidades, situa-se em uma posição intermediária entre estes dois regimes.

Do *ponto de vista privado*, na competição em preços para bens substitutos, ambas as empresas preferem o regime de Stackelberg ao regime de Bertrand, o que conflita com a preferência dos consumidores, que adquirem os bens a preços mais baixos neste último regime. Na competição em quantidades para bens complementares, o regime de Stackelberg é preferível ao de Cournot. Neste caso particular, unicamente, o interesse das empresas pode ser compatível com o interesse dos consumidores, no seu conjunto.

Obviamente, estes resultados dependem da hipótese assumida para o duopólio linear e simétrico. A não linearidade das curvas de demanda e/ou de custos abre a possibilidade para a existência de equilíbrios múltiplos e para a instabilidade. A assimetria de custos entre as empresas altera não apenas a hierarquia dos lucros entre elas como também o excedente dos

consumidores. Ainda que (em condições regulares) não haja dificuldades maiores para se determinar e analisar o equilíbrio de uma indústria compreendendo um número arbitrário de empresas assimétricas, os resultados da comparação entre os desempenhos obtidos, em cada um destes regimes, são sensivelmente obscurecidos pela assimetria dos custos ou das preferências dos consumidores. A qualificação dos regimes dependerá, também, da magnitude destas assimetrias (vide confronto Bertrand *versus* Cournot, na Seção 10.4.3). A dominância de um padrão competitivo sobre o outro não pode ser estabelecida, em toda generalidade.

Notas

1. Reações antecipadas na determinação do equilíbrio foram primeiramente consideradas por A.L. Bowley (1924). A atual designação é devida a R. Frisch. (1933).
2. Como a análise à frente sugere, a consistência das conjecturas ocorre em situações um tanto excepcionais, em modelos não dinâmicos. Esta limitação reflete a dificuldade de se operar com um conceito dinâmico no interior de um mecanismo estático. Apesar de a inconsistência torná-las menos úteis do ponto de vista operacional, as reações antecipadas pelos produtores ganharão relevo na apresentação que segue por uma razão didática: os modelos teóricos mais usuais de competição entre as empresas (competição perfeita, Cournot, Bertrand, Stackelberg) ou de cartel, podem ser entendidos a partir do seu prisma, sendo gerados por valores particulares da variação conjectural.
3. Conforme Hackner, 2000.
4. Esse método é descrito no Capítulo 11.

Bibliografia

- Boff, H.P.; Werlang, S. R. C. (1998) "Cournotian Competition under Knightian Uncertainty", *Revista de Econometria*, v. 18, n. 2, p. 265-308.
- Cabral, L. M. B. (1995) "Conjectural variations as a reduced form", *Economic Letters*, v. 49, p. 397-402.
- Deneckere, R.; Davidson, C. (1985) "Incentives to form coalitions with Bertrand competition", *Rand Journal of Economics* v. 16, n. 4, p. 473-486.
- Dixit, A. (1986) "Comparative statics for oligopoly", *International Economic Review*, v. 27, n. 1, p. 107-122.
- Dockner, E. J. (1992), "A Dynamic Theory of Conjectural Variations", *The Journal of Industrial Economics*, v. XL, n. 4, p. 377-397.
- Farrell, J.; Shapiro, C. (1990) "Horizontal mergers: An Equilibrium analysis", *The American Economic Review*, v. 80, n. 1, p. 107-12.
- Hackner, J. (2000) "A Note on price and quantity competition in differentiated oligopolies", *Journal of Economic Theory*, v. 93, p. 233-239.
- Hart, O. D. (1985) "Monopolistic competition in the spirit of Chamberlin: A general model", *Review of Economic Studies*, v. LII, p. 529-546.
- Kamien, M. I.; Schwartz, N. L. (1983) "Conjectural Variations", *Canadian Journal of Economics*, p. 191-211.
- Mas-Colell, A.; Whinston, M. D.; Green, J. R. (1995) *Microeconomic Theory*, Oxford University Press, Caps. 3, 4, 5, 12.
- Perry, M. K.; Porter, R. H. (1985) "Oligopoly and the incentive for horizontal mergers", *The American Economic Review*, v. 75, p. 219-27.
- Salant, S. W.; Switzer, S.; Reynolds, R. J. (1983) "Losses from horizontal mergers: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium", *Quarterly Journal of Economics*, v. XCVII, n. 2, p. 185-189.
- Scherer, F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*, Houghton Mifflin Co., Caps. 2, 6, 7.
- Shaked, A.; Sutton, J. (1982) "Relaxing price competition through product differentiation", *Review of Economic Studies*, v. XLIXI, p. 3-13.
- Shy, O. (1995) *Industrial Organization*, The Mit Press, Caps. 3, 4, 5, 6, 7, 8.
- Singh, N.; Vives, X. (1984) "Price and quantity competition in a differentiated duopoly", *Rand Journal of Economics*, v. 15, p. 4, Winter.
- Tirole, J. (1989) *The Theory of Industrial Organization*, The MIT Press, Caps. 1, 5, 7.
- Varian, H. R. (1992) *Microeconomics Analysis*, W.W. Norton & Co., 3^a ed., Caps. 1, 4, 5, 7, 9, 10, 14, 16.
- Vives, X. (1985) "On the efficiency of Bertrand and Cournot equilibria with product differentiation", *Journal of Economic Theory*, v. 36, p. 166-175.
- Williamson, O. E. (1968) "Economies as an antitrust defense: the welfare tradeoffs", *The American Economic Review*, v. LVIII, n. 1, p. 18-36.7.



Coordenação Oligopolista

Frederico Rocha

11.1 Introdução

Nos capítulos anteriores, vimos que a presença de interação entre empresas em um mesmo mercado – inseridas em processos de concorrência que podem ser descritos por jogos não cooperativos, com formato do Dilema do Prisioneiro – pode conduzir a soluções subótimas do ponto de vista das empresas. Isto é particularmente relevante em situações em que a concorrência por preços é acirrada, podendo levar à inviabilidade das estruturas oligopolistas no longo prazo. A ideia de que as empresas assistirão passivamente sua eliminação do mercado por causa de concorrência predatória parece pouco plausível. Isto levanta, por um lado, algumas dúvidas sobre os processos concorrenceis puramente baseados em preços,¹ por outro, sugere que – caso as escolhas dos consumidores permaneçam sensíveis aos preços – as empresas deverão empreender algum acordo para não reduzir os preços.

A adoção de acordos parece ser possível, uma vez que se abandone a hipótese estática dos modelos anteriores. Podemos intuir que, uma vez observadas as sequências de movimentos que cortam preços, afetando negativamente seus lucros, as empresas notarão a inconsistência do processo concorrencial em que estão envolvidas. Devemos supor, portanto, que as empresas repensarão as hipóteses míopes – presentes nos modelos de Cournot e Bertrand – a respeito da reação de seus rivais, ou seja, não parece ser razoável que as empresas suponham que diante de uma mudança de suas estratégias de preço ou quantidade, as rivais manterão suas estratégias anteriores.

A revisão estratégica em direção à maximização conjunta de lucros parece ser uma possibilidade. Este comportamento já havia sido previsto por Adam Smith:

Pessoas da mesma profissão [do mesmo setor de negócios] raramente se encontram, mesmo que seja para momentos alegres e divertidos, mas as conversações sempre terminam em conspiração contra o público ou em algum incitamento para aumento dos preços.²

De fato, a evidência empírica de Economia Industrial e a história de empresas e relatos de casos em organismos de defesa da concorrência demonstram haver inúmeras situações em que empresas estabelecem acordos de preços tácitos ou formais. Acordos formais são aqueles em que empresas estabelecem algum tipo de comunicação e combinam entre si regras específicas de estabelecimento de preços ou formas de comportamento diante de distintas situações. Estes acordos podem ser escritos ou apenas mantidos por intermédio de palavra. Acontece que normalmente as autoridades de defesa da concorrência estão bastante atentas a essas práticas. Muitas vezes, portanto, as empresas não podem se reunir em grupos, nem estabelecer comunicação explícita. Nestes casos, quando passam a coordenar seus preços e ações, seguem regras tácitas, ou seja, que não

foram combinadas explicitamente, mas que a vivência na indústria e o processo de aprendizagem estabeleceram como prática saudável para as empresas. Um exemplo de acordos formais é a Organização dos Países Exportadores de Petróleo (OPEP), que estabelece quotas de produção para cada um de seus membros, impedindo que a oferta exceda um determinado nível e permitindo o estabelecimento de preços acima do custo marginal. Um exemplo de acordo tácito pode ser observado na praia em que é facilmente perceptível que o preço da cerveja na areia é mais caro do que o preço da cerveja nos quiosques e não costuma ser diferenciado entre vendedores atuando na mesma faixa, ou seja, entre ambulantes e entre quiosques. Neste caso, não houve comunicação entre todos os agentes presentes, mas parece que eles têm algum mecanismo que possibilita a uniformização dos preços, mesmo quando as condições de demanda se alteram.

Este capítulo se propõe a tratar alguns aspectos que facilitam ou prejudicam a coordenação oligopolística. Na primeira seção, os elementos básicos para a existência destes acordos são esboçados. A segunda seção é dedicada à discussão de situações e características de mercado que prejudicam a coordenação em preços. A quarta seção discute alguns modos de superar estes problemas.

11.2 Condições Básicas de Coordenação

Admitamos um mercado de um produto homogêneo com apenas duas empresas. Elas têm custos marginais idênticos e constantes. Neste caso, a solução de Bertrand será preço igual a custo marginal, com as duas empresas dividindo o mercado. Suponha, conforme a Figura 11.1, que a indústria mantenha custos fixos elevados. Nesta situação, as empresas estarão funcionando com prejuízo e, no longo prazo, seu destino será sair da indústria.

Uma alternativa à saída da indústria é a cooperação em preços, ou seja, a formação de um acordo tácito ou formalizado que deverá ter o estabelecimento de seus preços coordenados. O grande problema para a manutenção do acordo está no incentivo existente para reduzir os preços e conquistar todo o mercado. Suponhamos que os consumidores sejam perfeitamente sensíveis ao preço, de maneira que uma pequena (talvez mínima) redução no preço seja percebida pelo consumidor. Se as empresas chegam a um acordo inicial em que o preço é igual ao de monopólio, p^M , elas obterão o lucro de monopólio proporcionado pelo mercado, ou seja, $\pi^C = \frac{\pi^M}{2}$, onde π^C e π^M são, respectivamente, o lucro de cada empresa participante do acordo e o lucro de monopólio. Se um concorrente sair unilateralmente do acordo, aplicando preço igual a $p^M - \epsilon$, onde $\epsilon \rightarrow 0$, então seu lucro nesse período será igual a $\pi^M = 2\pi^C$, ou seja, se o oligopolista fugir do acordo, ele dobra o seu lucro. A empresa que não descumpriu o acordo e manteve seu preço em p^M permanece fora do mercado, pelo menos por este período.

O problema está em conhecer a reação de seu concorrente, ou seja, supondo que a detecção por parte do rival se dê no mesmo período em que a redução de preços ocorreu, que tipo de *retaliação* será empreendido pela empresa que se manteve no acordo. Várias hipóteses podem ser formuladas. Parece razoável para iniciar a discussão sobre coordenação de preços que se formule a hipótese de reação mais pessimista na visão do oligopolista: caso um dos parceiros não coopere, seu rival perder completa e

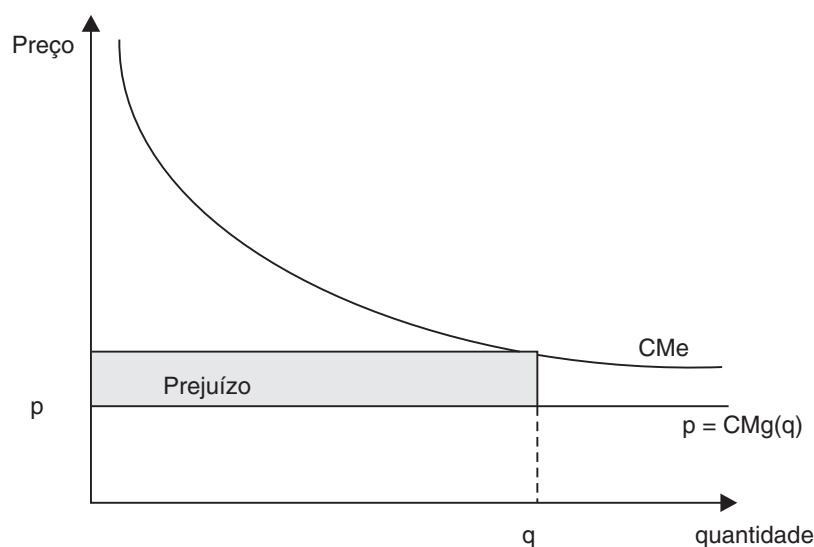


Figura 11.1 Prejuízo causado pela adoção da solução de Bertrand.

eternamente a confiança no processo de cooperação, para nunca mais aceitar entrar em acordo por preços; ou seja, em todos os períodos posteriores o preço será igual ao custo marginal e o lucro de ambos será 0. Esta é uma estratégia conhecida como gatilho de preços, muitas vezes denominada também como competição impiedosa (do inglês *grim*). Desta maneira, podemos pensar que os dois oligopolistas se veem diante das seguintes opções. Caso venham a trair o acordo, eles obterão o lucro de monopólio no primeiro período, mas receberão lucro 0 em todos os períodos subsequentes. Caso contrário, eles obterão metade do lucro de monopólio no período atual e em todos os períodos subsequentes. Suponha que o lucro dos períodos subsequentes seja descontado pela taxa de juros, de maneira que o lucro do período i seja contabilizado no primeiro período como $\delta^{(i-1)}\pi^i$, sendo $\delta = \frac{1}{1+r}$, onde δ é o fator de desconto, r a taxa de desconto e i , o período em que o lucro é percebido. Assim, a empresa decidirá cooperar desde que os retornos em caso de cooperação sejam superiores aos retornos em caso de traição, ou seja,

$$\begin{aligned} \pi^C + \delta\pi^C + \delta^2\pi^C + \dots + \delta^{(i-1)}\pi^C + \dots &> \pi^M \Rightarrow \\ \frac{1}{1-\delta} \frac{\pi^M}{2} &> \pi^M \Rightarrow \delta > \frac{1}{2}. \end{aligned}$$

Isto quer dizer que a cooperação deverá existir desde que o fator de desconto quanto ao futuro seja suficientemente elevado (ou seja, a taxa de desconto seja suficientemente pequena). Cabe ressaltar, contudo, que muitos elementos devem ser adicionados à taxa de desconto do lucro futuro. Por exemplo, rendimentos futuros também devem ser descontados pela probabilidade de o acordo terminar ou o mercado se extinguir. Supondo que esta probabilidade seja expressa por ρ , então $\delta = \frac{1}{1+r}(1-\rho)$. Suponhamos, então, que a probabilidade da relação terminar seja igual a 0. Assim, $\delta = \frac{1}{1+r}$. Se a probabilidade da relação terminar é 1, então, $\delta = \frac{1}{1+r} \cdot 0 = 0$. Isto significa que quanto maior a instabilidade do mercado maior será a dificuldade para se chegar a um acordo em preços.

11.2.1 INTENSIDADE DA RETALIAÇÃO E COOPERAÇÃO

Devemos esclarecer que os resultados dependerão do tipo de retaliação a ser adotado pelas empresas traídas (ver o Quadro 11.1). A história dos mercados não indica que, uma vez rompido o acordo, nunca mais haverá cooperação entre as empresas, como foi suposto anteriormente. As empresas tendem a relutar em ter preços reduzidos eternamente, principalmente em indústrias que detêm elevados custos fixos e a condução do preço ao nível do custo marginal implica pesados prejuízos para as empresas, como parece ser a regra geral das indústrias concentradas. Empresas em ambientes oligopolistas procuram constantemente a formação de acordos por preços. Uma vez encontrada uma violação, o mais provável é que arranjos sejam realizados para que a cooperação seja retomada. Desta maneira, a hipótese de que as empresas nunca mais se entenderão parece ser radical demais. Suponha alternativamente que a reação da empresa seja não cooperar nos dois períodos imediatamente após a traição e cooperar no terceiro período. Neste sentido, o lucro da empresa traída deverá ser determinado por:

$$\pi^M + 0 + 0 + \delta^3 \frac{\pi^M}{2} + \delta^4 \frac{\pi^M}{2} + \dots = \pi^M + \frac{\delta^3}{1-\delta} \pi^M > \pi^M.$$

Isto significa que para haver cooperação será necessário que

$$\begin{aligned} \pi^C + \delta\pi^C + \delta^2\pi^C &> \pi^M \Rightarrow (1+\delta+\delta^2) \frac{\pi^M}{2} > \pi^M \Rightarrow (1+\delta+\delta^2)\pi^M > 2\pi^M, \text{ ou} \\ \delta &> 0,618. \end{aligned}$$

QUADRO 11.1 SOLUÇÕES DO DILEMA DO PRISIONEIRO

Um torneio patrocinado por Robert Axelrod reuniu diversos estudiosos de Teoria dos Jogos com o objetivo de conhecer a melhor maneira possível de enfrentar o Dilema do Prisioneiro. Cada estudioso foi encarregado de elaborar uma estratégia com que iria enfrentar todos os demais concorrentes e suas respectivas estratégias. A estratégia que apresentou o melhor resultado final foi aquela escolhida por Anatol Rapoport, denominada *tit-for-tat* ou o bíblico *olho por olho, dente por dente*. Nesta estratégia, Anatol iniciava cada confronto com um movimento colaborativo. Caso o seu rival colaborasse, ele responderia colaboração, caso o seu rival empreendesse uma estratégia traída, ele responderia traição, ou seja, sendo $a_{i,t}$ a estratégia da empresa “ i ” no período t , $a_{i,t} = a_{j,t-1}$. É interessante perceber que muitas vezes esta não era a melhor estratégia contra cada uma das estratégias individualmente, mas, *na média*, foi a que obteve o melhor resultado.

Este resultado fornece um importante ensinamento neste estágio: sob o ponto de vista da empresa que romperá o acordo, quanto mais branda a ameaça de retaliação, menor a probabilidade de haver cooperação; quanto mais feroz a retaliação, maior o número de situações em que ocorrerá cooperação. Logo, a cooperação depende fundamentalmente da retaliação.

O mais interessante a respeito deste resultado é que, sob o ponto de vista da empresa que foi traída, conforme sugerido anteriormente, a competição impiedosa não é a melhor opção. Uma vez tendo sido traída, caso adote a estratégia impiedosa, seus lucros serão 0 em todos os demais períodos; caso retorne ao acordo, pelo menos em um período, seu lucro será superior a 0. Este problema levanta uma questão adicional: ainda que a ameaça de adoção de estratégias impiedosas seja a melhor opção para evitar a traição, sua credibilidade após a violação do acordo é pequena. A empresa traída terá muito pouco incentivo para manter seus preços iguais aos custos marginais. Isto conduz a um segundo importante ensinamento a ser extraído desta análise: a natureza indeterminada do oligopólio e de suas inúmeras variantes de interação.

11.3 Condições que Dificultam a Coordenação

Existem diversas condições que dificultam a coordenação empresarial, criando problemas adicionais. Elas estão normalmente associadas com:

1. Dificuldade de detecção da violação ao acordo
2. Assimetrias nas estruturas de custos
3. Heterogeneidade do produto
4. Número de empresas presentes no mercado
5. Estruturas de custos
6. Alterações nas condições presentes no mercado

11.3.1 SEGREDOS, DETECÇÃO E ACORDOS

Um problema parecido com o provocado pela retaliação ocorre quando há dificuldade de detecção por parte da empresa. Suponha que os preços não possam ser imediatamente observados devido à possibilidade de manutenção de contratos em segredo entre fornecedor e usuário de um determinado serviço. Neste caso, as empresas só observarão os preços no período seguinte, em decorrência de variações em suas demandas. A consequência é que as empresas venham a retaliar somente dois períodos após a ocorrência de traição do acordo por um de seus participantes. Os lucros da empresa traidora serão determinados então por $\pi^M + \delta\pi^M + 0 + 0 + \dots$. Para haver cooperação, então,

$$\frac{1}{1-\delta} \frac{\pi^M}{2} > \pi^M + \delta\pi^M \Rightarrow \delta > \frac{1}{\sqrt{2}} \cong 0,707,$$

ou seja, os requisitos para não violação do acordo são superiores ao caso da estratégia gatilho analisado anteriormente.

11.3.2 ASSIMETRIAS NOS CUSTOS

A existência de assimetrias de custos pode ser um importante obstáculo à manutenção do acordo oligopolista. Conforme pode ser observado na Figura 11.2, empresas com custos marginais diferentes, repartindo o mercado igualmente, terão percepções bastante distintas quanto ao preço a ser estabelecido para a maximização de lucros e também quanto à quantidade a ser vendida. A empresa de custos mais elevados tenderá a aplicar preços maiores e a vender quantidades inferiores. A obtenção de um acordo tácito se torna mais difícil pela falta de convergência a um preço comum a prevalecer no acordo. Supondo que as duas empresas iniciem estabelecendo preços iguais aos de monopólio, a empresa de custos inferiores, ao aplicar preços menores, atrairá demanda da empresa de custos superiores, deslocando sua curva para a esquerda e a curva de demanda da empresa de custos superiores para a direita. Isto pode ser interpretado pela empresa de custos elevados como o início de uma guerra de preços, levando a maiores cortes de preços.

Uma alternativa a este procedimento é o estabelecimento de um cartel, em que as empresas dividam o mercado de acordo com o nível de seus custos marginais, conforme exposto na Figura 11.3. O cartel estará maximizando os lucros quando os custos marginais se encontram com a receita marginal na Figura 11.3C. Nesta situação, a empresa 1 terá uma produção muito superior àquela da empresa 2. Neste caso, o excedente total do produtor será definido como a área cinza na Figura 3C e o

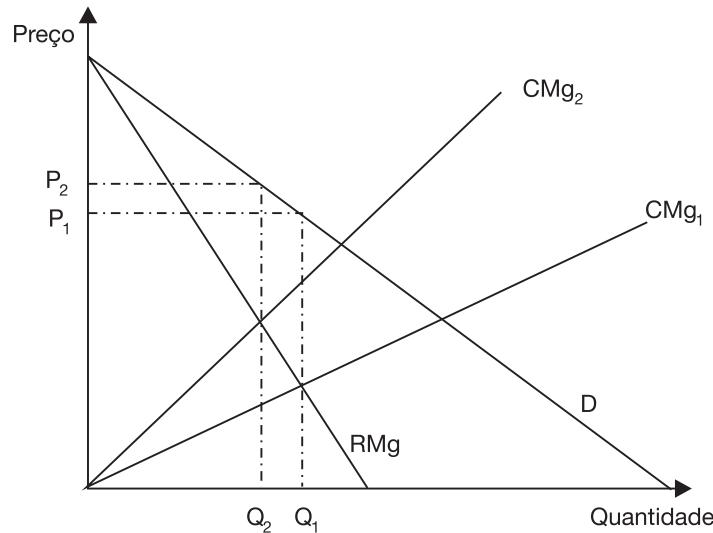


Figura 11.2 Custos Marginais Assimétricos e Determinação de Preços e Quantidades.

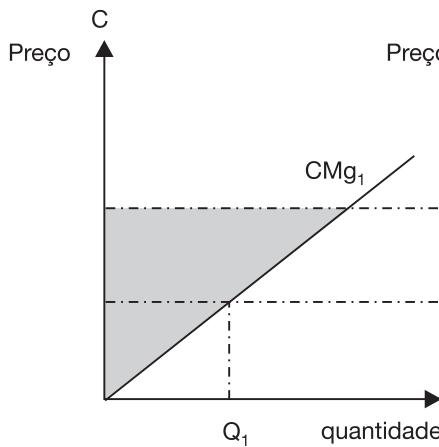


Figura 3A - Firma 1

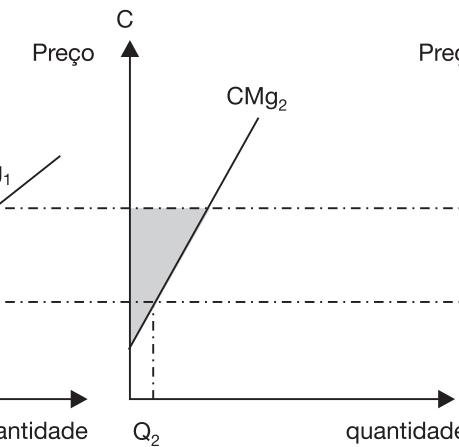


Figura 3B - Firma 2

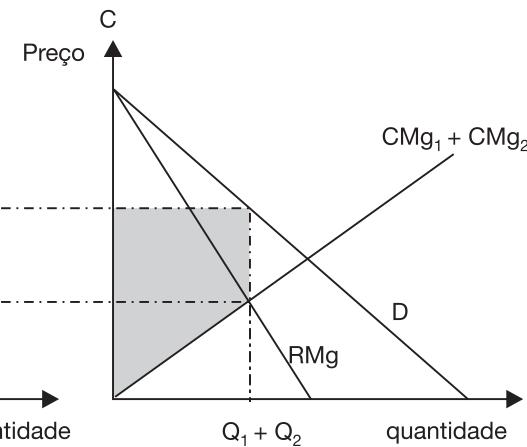


Figura 3C - Cartel

Figura 11.3 Solução de cartel de quantidades.

excedente destinado a cada uma das empresas será definido pelas áreas cinzas nas Figuras 3A e 3B. Como pode ser percebido, a empresa 1 recebe um excedente significativamente superior àquele da empresa 2. Isto pode gerar insatisfação, conduzindo a empresa 2 a solicitar partição mais equitativa. A solução requererá uma coordenação ainda superior, como, por exemplo, a formação de uma organização central encarregada de receber e distribuir lucros entre as empresas.

11.3.3 HETEROGENEIDADE DE PRODUTO

A heterogeneidade de produto pode se manifestar como obstáculo à coordenação oligopolista sob quatro formas: diferenciação de produto, diferenciação geográfica, mudanças nos parâmetros da qualidade do produto e produtos complexos.

A principal implicação da diferenciação de produto é a presença de assimetrias nas curvas de demanda individuais. Mais uma vez os preços de monopólio para as duas empresas tenderão a ser diferentes. Isto pode ser observado na Figura 11.4, que mostra que, neste caso, a empresa com a curva de demanda D_1 aplicará preços mais elevados. A discordância quanto aos preços de equilíbrio pode gerar conflitos entre as empresas. Neste caso, a situação é ainda mais complicada do que a discordância de preços em casos de diferenças de custos porque a reação usual para a empresa de preços mais altos seria diminuir seu preço, acentuando ainda mais a distorção no que se refere à partição do mercado. Uma alternativa para a empresa 2 seria aproximar seu produto em relação àquele da empresa 1, acirrando a concorrência por preços.

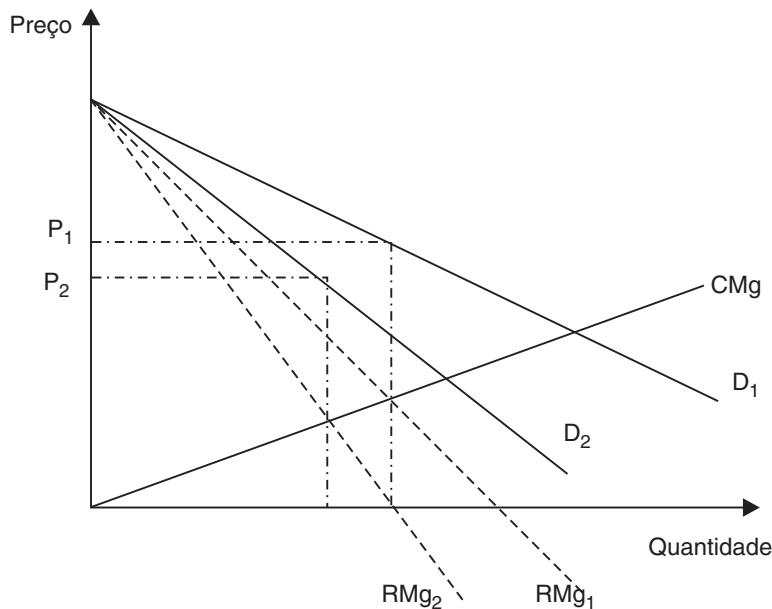


Figura 11.4 Diferenciação de produtos e assimetrias nas curvas de demandas individuais.

O aumento do número de atributos torna o produto mais complexo e tem dois efeitos. Primeiro, na medida em que aumenta a quantidade de atividades a ser coordenada, a coordenação pode se tornar muito complexa, requerendo procedimentos muito complicados e criando situações de difícil monitoração. A indústria de máquinas e equipamentos em alguns casos prepara seus produtos com requisitos específicos de cada cliente. O grau de diferenciação do produto é tão grande que exige que a equipe de vendas seja da própria fábrica e que os vendedores sejam engenheiros especializados. A combinação de preços neste setor pode ter que chegar ao nível de especificar o preço de cada componente utilizado em uma máquina, exigindo a composição de um livro de normas. Ao mesmo tempo, sistemas de financiamento da compra podem ser muito variados, exigindo o conhecimento de todos os desembolsos a serem realizados. Fica claro que monitorar um acordo neste setor se torna muito mais difícil do que monitorar um acordo no setor de refrigerantes. O segundo efeito está associado ao primeiro, mas tem outras consequências que dificultam o acordo oligopolista. Com o aumento do número de atributos do produto a serem diferenciados, o grau de especificidade do produto cresce, a ponto de ser desenhado para cada cliente. Desta maneira, cada cliente deverá ter um preço diferente aplicado sobre seu produto. Isto facilita a formação de acordos entre o fornecedor e o usuário, mantendo contratos que têm cláusulas secretas de redução de preços. Como consequência, haverá um problema de detecção de eventuais traições aos acordos firmados entre produtores. O Quadro 11.2 ilustra essa situação por meio do exemplo histórico fornecido pelo setor de transporte ferroviário nos Estados Unidos no final do século XIX.

O caso das ferrovias americanas chama a atenção para alguns elementos que devem ser considerados na análise de situações de coalizão. Primeiro, a taxa de desconto está longe de ser única entre os agentes e a heterogeneidade de objetivos entre as empresas participantes de um oligopólio deve ser um elemento de análise. Segundo, a importância da possibilidade de diferenciação de produto. E terceiro, o efeito das estruturas de custos no comportamento colusivo das empresas.

11.3.4 NÚMERO DE CONCORRENTES

Suponha que em vez de dois concorrentes no mercado, existam três. Desta maneira, o lucro das empresas no acordo por preços é $\pi^{C,3} = \frac{\pi^M}{3}$. Isto conduz a consequências importantes. Supondo que a retaliação seja a mais violenta possível, a condição para haver acordo deverá ser então que

$$\frac{1}{(1-\delta)} \pi^{C,3} > \pi^M, \text{ ou seja, } \frac{1}{(1-\delta)} \frac{\pi^M}{3} > \pi^M \Rightarrow \delta > \frac{2}{3},$$

ou seja, as condições em que o futuro é descontado têm de ser melhores do que quando só há dois concorrentes no mercado.

QUADRO 11.2 O CASO DAS ESTRADAS DE FERRO: CRIAÇÃO DE FEDERAÇÃO DE EMPRESAS

A análise da indústria de transporte ferroviário nos EUA, durante o último quartil do século XIX, ajuda a compreender as dificuldades encontradas na formação de acordos oligopolistas. A indústria era caracterizada por elevados custos afundados e quase-fixos, representados, respectivamente, pela construção da linha férrea e por sua manutenção. Em simultâneo, os custos variáveis (marginais) eram bastante reduzidos, conduzindo a uma configuração de custos semelhante àquela da Figura 11.1. A precificação no nível do custo marginal comprometia, assim, a sobrevivência da indústria no longo prazo. Acordos de preços eram, portanto, fundamentais.

Com o objetivo de coordenar o fluxo de mercadorias e estabelecer acordos, foi criada uma federação de empresas. Acordos de preços foram firmados. Havia, no entanto, uma característica na indústria que levava à negociação das tarifas caso a caso. As cargas transportadas variavam quanto aos locais de origem e destino e peso e forma de transporte. Como consequência, tarifas eram obtidas por acordos entre as estradas de ferro e seus usuários. Ainda que preços elevados fossem estabelecidos, o formato decrescente da curva de custos criava incentivos à redução de tarifas para atrair tráfego, diminuindo o custo médio. Como resultado, vários acordos foram rompidos.

Havia ainda um problema adicional que tornava mais difícil a manutenção do acordo. Enquanto alguns produtores tinham um horizonte de longo prazo, outros mantinham horizontes de curto prazo, refletindo-se em descontos muito elevados sobre os rendimentos futuros. A figura de Jay Gould parece ter sido particularmente importante a este respeito. Com forte atuação na bolsa de valores, a maior parte de seus rendimentos advinha da especulação, ou seja, da venda de ações em períodos de alta e compra em períodos de baixa. Seu controle sobre uma das principais companhias de estrada de ferro permitia a obtenção de informação privilegiada sobre o andamento de negócios. Assim, tão logo se chegava a um acordo por preços, Gould ordenava que parte de suas linhas reduzisse secretamente suas tarifas, atraindo mais tráfego. Como consequência, seus lucros se elevavam. A traição só era percebida pelos demais membros conforme o fluxo de carga ia se transferindo de suas linhas para aquelas operadas por Jay Gould. Entre o período em que os lucros se elevavam e a reação das demais empresas, o preço das ações das estradas de ferro controladas por Gould se elevava. Gould então vendia parte de suas ações. Com a quebra do acordo e o início de guerra tarifária, os lucros caíam, tendo efeito negativo sobre os preços das ações. Gould atuava, então, no mercado, comprando as ações que vendera em alta. A instabilidade dos preços nesse mercado só foi solucionada após uma onda de fusões que permitiu a concentração da indústria.

Fonte: Chandler, A. The Visible Hand. The Belknap Press, Cambridge (MA), 1974.

11.3.5 ESTRUTURAS DE CUSTOS

A existência de custos fixos elevados e o formato dos custos marginais parecem ser importantes elementos para o rompimento ou a manutenção das relações amistosas entre as empresas. Quando custos marginais são crescentes, elevações na quantidade produzida implicam aumentos nos custos. Essa característica dificulta o rompimento de acordos tácitos. Tomemos uma estrutura de custos e demanda como a presente na Figura 11.5a. A curva tracejada positivamente inclinada representa os custos marginais de cada empresa individual. A curva contínua positivamente inclinada é a curva de custos do cartel ou de oferta da indústria, que é determinada pela soma horizontal das curvas de custos marginais individuais. O preço do cartel é determinado por P_C e o custo marginal de cada uma das empresas é C_C . Se uma empresa reduzir unilateralmente o preço do bem deverá suprir todo o mercado. Suponhamos que ela escolha um preço um pouco menor do que P_C , como, por exemplo, P_T . A este preço o mercado estará demandando a quantidade Q_T . No entanto, para produzir individualmente Q_T , a empresa traidora do

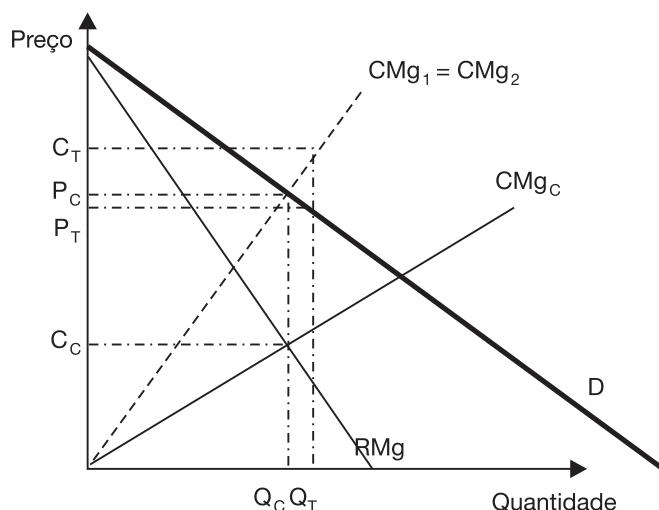


Figura 11.5a Cartel com Custos Marginais Crescentes.

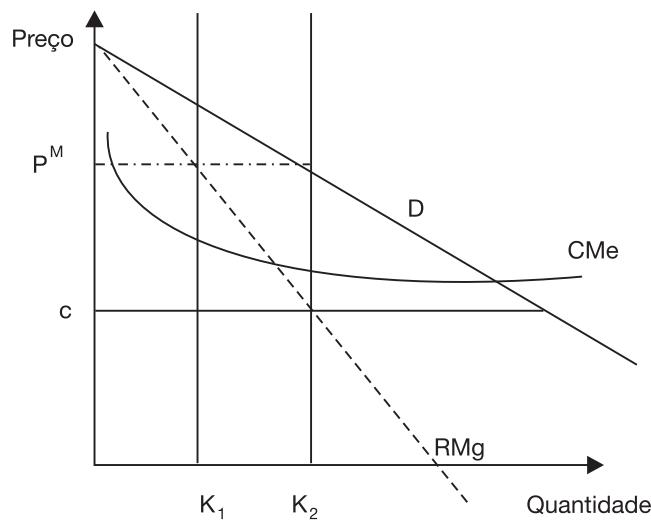


Figura 11.5b Cartel com Limitação de Capacidade Produtiva.

acordo incorrerá em um custo marginal $C_T > P_T$, o que cria incentivo para redução da quantidade produzida ou aumento do preço, facilitando a coordenação.

Um caso especial de custos marginais crescentes é a limitação de capacidade. Como pode ser observado na Figura 11.5b, pode-se imaginar a restrição de capacidade como a presença de uma curva de custos perfeitamente inelástica a partir de uma determinada quantidade K . Suponha que a empresa possa escolher dois tamanhos de plantas que apresentam custos marginais constantes até o limite da capacidade e a partir daí o custo marginal é infinito. Se houver apenas duas empresas no mercado, e as duas escolherem a capacidade K_2 , cada uma poderá isoladamente preencher todo o mercado, quando o preço for igual ao de monopólio. Se elas entrarem em um acordo ou formarem um cartel, o preço de monopólio pode ser adotado com cada uma utilizando metade de sua capacidade. Esta situação poderá comportar guerras de preços pelo incentivo que as empresas terão para ocupar a capacidade ociosa e reduzir seus custos médios. Suponha agora que as empresas tenham capacidade produtiva $K_1 = \frac{K_2}{2}$. Nesta situação, cada empresa só poderá isoladamente preencher metade da demanda quando o preço for igual ao de monopólio. Desta maneira, reduções de preço não poderão ser acompanhadas por aumento da quantidade produzida por cada empresa. Assim, se uma empresa resolver reduzir seu preço, ela não ocupará mais mercado, e mesmo que atrai consumidores da rival não poderá fornecer-lhes produtos. Pode-se concluir, portanto, que uma boa forma de garantir a coordenação tácita ou formal é a limitação da capacidade produtiva.

As dificuldades para manutenção do acordo tendem a crescer com o nível de custos fixos. Isto pode ser ilustrado pela Figura 11.6. Suponha dois mercados, ambos com a mesma demanda D e com duas empresas. No primeiro mercado, têm-se custos médios representados por CMe_1 e custos marginais CMg_1 . No segundo, os custos médios são representados por CMe_2 e os custos marginais por CMg_2 , que é a mesma curva que CMg_1 . Percebe-se, portanto, que a única diferença entre os dois mercados são os elevados custos fixos que incorrem as duas empresas que atuam no mercado 1 e os menores custos fixos que incorrem as empresas que atuam no mercado 2. Caso atuem como cartel, as empresas de ambos os mercados deverão dividir a demanda total do mercado, Q_T , com preço P , produzindo, cada uma, Q_F . Percebe-se que, no primeiro mercado, as empresas terão custos médios maiores que os preços ($C_1 > P$) e, no segundo mercado, os custos médios serão inferiores ao preço ($C_2 < P$). A principal diferença está em que, enquanto no primeiro mercado as empresas atuam com prejuízo representado pela área tracejada, no segundo mercado elas obtêm lucro representado pela área cinza. Como consequência, existe maior estímulo para as empresas do primeiro mercado violarem o acordo.

11.3.6 MUDANÇAS NAS CONDIÇÕES DO MERCADO

A exposição anterior adotou duas hipóteses que não parecem ser verdadeiras na maior parte das situações, isto é, assumiu-se que:

1. O preço adotado no acordo será igual ao de monopólio; e
2. As condições gerais do mercado e da indústria permanecem inalteradas, ou seja, privilegiamos analisar as condições que dificultam a coordenação oligopolista em situação estática.

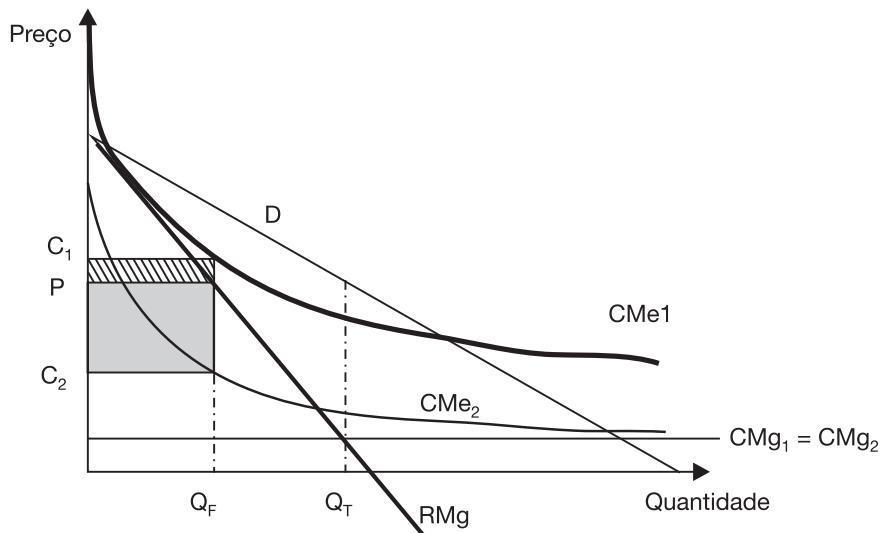


Figura 11.6 Custos Fixos Assimétricos e a Solução de Cartel.

Algumas razões podem existir para que o preço a ser adotado pela indústria em oligopólio com coordenação tácita ou formal não seja igual ao que seria estabelecido em monopólio. Inicialmente, pode haver regulação econômica (ver Capítulo 23) que impede que o preço se situe acima de um determinado nível. Segundo, conforme já ressaltado, não será em todo caso que as empresas concordarão a respeito de qual seria o preço que maximizaria os lucros em conjunto. Terceiro, existem situações em que o preço pode ser um importante instrumento para impedir novas entradas (ver Capítulos 7 e 12).

A impossibilidade de se adotar o preço de monopólio cria problemas adicionais para o cartel. Conforme ressaltado na Subseção 11.2.3, surgirá a pergunta sobre que preço adotar. A resposta pode ser bastante complexa em situações de assimetria, como as apontadas anteriormente, mas principalmente quando há desconhecimento a respeito dos custos ou condições de produção alheias. Isto pode ocorrer devido a várias razões. A comunicação entre as partes pode ser proibida legalmente; uma parte pode querer divulgar informação a respeito de seus custos, mas as condições gerais de concorrência podem ser tais que geram desconfiança quanto à informação concedida.

A consequência imediata da falta de informação é que a coordenação deverá ser alcançada pelo aprendizado. A questão está em que o aprendizado leva tempo e produz erros. Movimentos de preço podem parecer agressões entre empresas rivais, gerando situações de conflito entre as empresas. A discordância entre as partes conduz à guerra de preços.

Os problemas de aprendizado dos preços de equilíbrio parecem ser especialmente graves quando há mudanças nos custos, na qualidade e nos atributos dos produtos. A forma de coordenação de preços em indústrias de maior dinamismo tecnológico ou organizacional deve, portanto, comportar regras tácitas ou formais de mudanças nos preços quando há alterações nas condições básicas que regem as decisões de produção da empresa. As diferentes formas de coordenação e as condições que a tornam mais fácil serão exploradas na próxima seção.

11.4 Condições que Facilitam a Coordenação

As incertezas quanto aos movimentos dos rivais podem induzir as empresas a adotarem comportamentos defensivos, o que geraria resultados bastante conhecidos, como a rigidez dos preços. Neste caso, supõe-se que os oligopolistas adotam duas hipóteses: (1) elevações nos preços não serão acompanhadas por suas rivais; e (2) reduções nos preços serão seguidas pelos demais concorrentes, implicando apenas pequenas mudanças na quantidade demandada. A curva de demanda individual da empresa apresentará uma quebra, conforme expresso na Figura 11.7, que apresenta a curva de *demand quebrada* de Paul Sweezy,³ em que a curva de demanda é mais elástica para aumentos nos preços e inelástica para reduções. O resultado parece óbvio. Mudanças nos custos não deverão causar, pelo menos em um primeiro momento, alterações nos preços.

Por razões semelhantes, alterações na demanda tendem a não influenciar imediatamente os preços. O mais provável é que, em indústrias suficientemente concentradas, pequenas flutuações na demanda sejam respondidas por variações nos estoques ou atraso nas entregas de produtos do que por reduções ou aumentos nos preços. Isto significa que o ajuste em estruturas concentradas será representado basicamente por alterações nas quantidades vendidas. Este comportamento pode ser observado

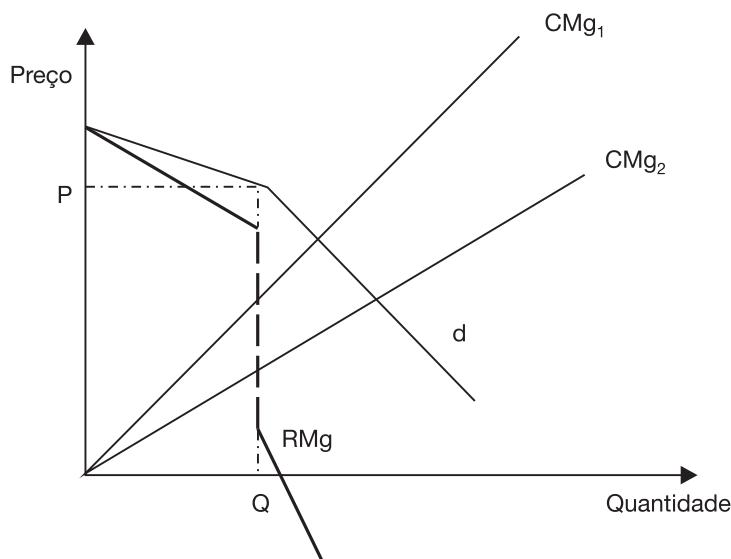


Figura 11.7 Curva de Demanda Quebrada.

em muitas ocasiões em várias indústrias. Em períodos de redução da demanda por automóveis, fruto de algum choque macroeconômico, percebe-se, por exemplo, que a primeira reação das empresas é aumentar os estoques de automóveis nos parques das montadoras, ou seja, elas mantêm a produção. Uma razão para isto está no elevado custo de dispensa ou manutenção da força de trabalho. Em um segundo momento, as empresas dão férias coletivas, impactando a produção. Somente em períodos muito longos de crise são observadas alterações nos preços.

11.4.1 REGRAS DE BOLSO E O PRINCÍPIO DO CUSTO TOTAL

Uma maneira de manter a coordenação da indústria diante de mudanças nas condições de custos é a utilização de regras de bolso⁴ para a formação de preços. Isto normalmente consiste na formulação de princípios gerais de formação de preços a partir dos custos. O formato mais usual destas regras é o princípio do custo total que se constitui em uma convenção adotada pela maior parte das empresas que atuam em oligopólio, segundo trabalho de Hall e Hitch (1939).⁵ De acordo com este princípio, as empresas calculam seus preços a partir de seus custos unitários variáveis (ou custos diretos), adicionando uma margem para cobrir os custos fixos e o lucro, ou seja, $p = v + \mu v$, onde v são os custos unitários variáveis (custos diretos) e μ é a soma da margem de lucro e o percentual dos custos fixos a ser coberto pela venda de cada unidade (*mark-up*). O principal problema na aplicação do princípio do custo total é saber qual o percentual dos custos fixos que deve ser coberto por cada unidade vendida, uma vez que as vendas tendem a variar ao longo do tempo. O mais provável, neste caso, é calcular este percentual a partir do que se acredita ser o grau normal de utilização da capacidade ou o grau de utilização planejado. Assim, apenas quando alterações no grau de utilização da capacidade forem persistentes haverá revisão da margem. Outro ponto importante é saber qual a margem de lucro a ser adotada. O artigo original de Hall e Hitch argumentava a respeito de uma margem considerada *justa*. Uma interpretação alternativa seria que o nível de barreiras à entrada deverá determinar a amplitude das margens de lucro. A adoção do princípio do custo total permite que mudanças nos preços, fruto de pequenas alterações nos custos, sejam visualizadas por todas as empresas, evitando interpretações que sugiram o início de uma guerra de preços.

O nível da margem de lucro, em decorrência das múltiplas opções que se encontram à disposição dos oligopolistas, provoca a discussão sobre possíveis situações em que haja uma confluência de opiniões sobre que preços ou que margens devem ser estabelecidos. Esta é a abordagem dos *pontos focais*, que partem do princípio de que, na presença de equilíbrios múltiplos, haverá algum elemento externo que determinará a escolha de um equilíbrio específico. Suponha que dois estudantes de economia tenham marcado ir ao jogo do Flamengo, no Maracanã, meia hora antes do início da partida, mas tenham esquecido de combinar o local de encontro. Provavelmente, a maior parte das pessoas concordaria que a estátua do Bellini é o local para onde ir encontrar o amigo. Neste sentido, a estátua do Bellini seria um ponto focal. Apesar de o encontro poder se dar em qualquer canto do Maracanã, há uma convergência de opiniões de que o Bellini é o local correto. Este tipo de procedimento pode ser adotado para determinar margens de lucro.

11.4.2 LIDERANÇA DE PREÇOS

A liderança de preços é uma importante forma de se chegar à coordenação de preços. Nesta concepção, uma empresa anunciará mudanças de preços que seriam seguidas pelas demais existentes na indústria. Esta prática, além de gerar coordenação, pode evitar suspeitas de comunicação ilegal por parte de organismos de defesa da concorrência. Existem três formas de liderança de preços: (1) empresa dominante, (2) colusão e (3) barométrica.

A liderança de preços por empresa dominante ocorre quando há uma empresa líder que tem grande parcela do mercado e alguma vantagem em relação às demais, controlando o mercado. Esta empresa adota, então, o preço que melhor lhe convém, dada a estrutura de oferta das empresas seguidoras, ou seja, ela antecipa a escolha de quantidade das empresas marginais para cada preço, calculando a demanda residual. A partir da demanda residual, ela estabelece o preço que maximiza os seus lucros, sendo seguida pelas demais empresas da indústria. A Figura 11.8 ilustra esta situação. A curva D é a demanda do mercado e S_F a oferta das empresas da franja. A empresa líder antecipa a oferta das empresas da franja, subtraindo da demanda total a quantidade que seria ofertada por estas empresas. A partir deste procedimento, ela obtém a curva de demanda residual d . Ela então maximiza os seus lucros a partir da curva de demanda individual, obtendo o preço P_L e ofertando Q_L . No entanto, a este preço, a demanda do mercado é Q_T , sendo a diferença igual à quantidade ofertada pelas empresas da franja, Q_F . Aparentemente, este é o caso dos ovos de Páscoa apresentado no Quadro 11.3.

Em estruturas mais concentradas, com elevadas barreiras à entrada, e em que não existem diferenças de custos relevantes, não há empresa dominante. Isto significa que nenhuma empresa poderá impor o preço que maximize os seus lucros individuais. Ao mesmo tempo, há necessidade de encontrar regras que possibilitem mudanças nos preços, evitando os casos de curva de demanda quebrada e guerra de preços, em um contexto em que normalmente a comunicação entre as empresas é proibida. Nestas condições, as empresas costumam convergir para a aceitação da liderança de um dos participantes do mercado. Este é o processo conhecido como liderança de preços *colusiva*. A escolha da líder pode ser feita de diversas maneiras. Algumas

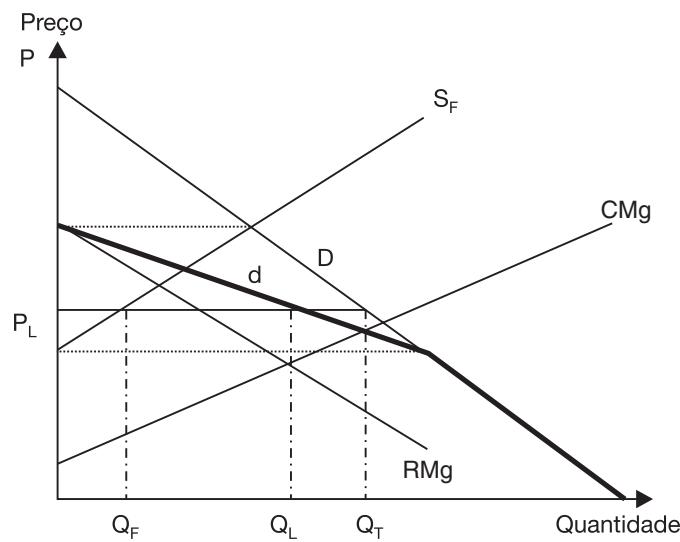


Figura 11.8 Liderança de preços com empresa dominante.

QUADRO 11.3 OVOS DE CHOCOLATE

Durante o período de Páscoa, no Brasil, há uma forte demanda por chocolate, principalmente ovos de chocolate. O mercado de venda a varejo de chocolate é composto de grandes empresas de lojas de departamento e supermercados e pequenos estabelecimentos ciliares, basicamente padarias e lojas de conveniência. A concorrência por preços neste mercado é feroz, devido ao grande número de empresas existentes e ao entendimento geral de que, caso os ovos não sejam vendidos até o domingo de Páscoa, não haverá mais mercado. Assim, as empresas têm forte incentivo para a redução de preços. Há ainda um problema adicional. Devido ao grande número de empresas no mercado, a comunicação entre elas, para se chegar a um acordo, é muito difícil. Nesta situação, uma conhecida loja de departamentos brasileira adota um procedimento bastante interessante. Ela anuncia na televisão que detém os menores preços do mercado em razão de ter os menores custos e que qualquer redução de preços por parte dos concorrentes será respondida com mais redução. Desta maneira, solicita aos concorrentes que não reduzam preços. Este tipo de procedimento pode ser entendido como liderança de preços.

vezes, ela ocorre por razões históricas. Por exemplo, ainda que a partir do início do século a parcela de mercado da US Steel na indústria de aço norte-americana tenha decrescido, sendo inclusive ultrapassada por outras empresas, ela se manteve como líder em preços durante todo o processo.

Na liderança barométrica, não há uma empresa grande que consiga impor sua vontade às demais empresas da indústria. A líder de preços é normalmente uma pequena empresa que avalia as condições de demanda e custos no mercado. Após entender que existem razões suficientes para conduzir uma mudança nos preços, ela altera os seus preços. As demais empresas da indústria, muitas vezes de tamanho superior, têm a opção de seguir ou não a alteração preconizada pela líder. Caso as demais empresas avaliem positivamente a mudança empreendida, ela será validada; caso elas entendam que a alteração não tem razão de ser, elas mantêm seus preços iniciais. Se os preços das demais empresas são mantidos, a primeira empresa retorna ao seu preço inicial. O mais importante é que suas mudanças não são entendidas como um rompimento do acordo inicial.

11.5 Conclusão

Este capítulo nos fornece alguns ensinamentos importantes a respeito do processo de concorrência oligopolista que merecem ser ressaltados. Primeiro, mostra que os modelos de oligopólio clássicos, estáticos e dinâmicos, ainda que dotados de instrumentos de teoria dos jogos, apresentam com resultados hipóteses temporais que limitam as possibilidades de reação das empresas. Na verdade, os processos de formação de preços são muito mais indeterminados do que preveem estes modelos.

Segundo, ao chamar a atenção para o grande número de possibilidades de comportamento em situações de rompimento de acordo, o capítulo acaba levantando questões que devem ser tratadas mais à frente. De um lado, aparece a ideia de credibilidade das ações, que deve ser tratada no Capítulo 12; de outro, levanta a importância da diversidade de tipos de agentes, questões que serão abordadas por teorias evolucionárias.

Terceiro, ressalta a importância de estudos setoriais e até mesmo casos de empresas individuais para se compreender a evolução e o comportamento de preços nos mercados.

Por fim, ao introduzir contextos mais dinâmicos chama a atenção para a ação estratégica das empresas. Um exemplo disto é o caso dos ovos de Páscoa, em que o comprometimento com preços mais reduzidos acaba por ser um elemento importante para garantir preços altos. Este ponto também deverá ser tratado no Capítulo 12.

Notas

1. De fato, processos de diferenciação de produto e outras alternativas de rivalidade existem e são analisados nos Capítulos 6 e 8.
2. Smith, Adam *A Riqueza das Nações: Investigaçāo sobre sua natureza e suas causas*. Coleção Os Economistas. Editora Abril, 1982, p. 139.
3. Sweezy, P. Demand under Conditions of Oligopoly. *Journal of Political Economy*, n. 47, p. 568-573, 1939.
4. Do inglês *rules of thumb*.
5. Hall, R. e Hitch, C. Price Theory and Business Behavior. *Oxford Economic Papers*, n. 2, p. 12-45, 1939.

Bibliografia

- Chandler, A. (1974) *The Visible Hand*. The Belknap Press, Cambridge (MA).
- Hall, R.; Hitch, C. (1939) Price Theory and Business Behavior. *Oxford Economic Papers*, n. 2, p. 12-45, 1939.
- Scherer F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*. Cambridge (Mass.), Houghton Mifflin.
- Smith, A. (1982) *A Riqueza das Nações: Investigaçāo sobre sua natureza e suas causas*. Coleção Os Economistas. Editora Abril.
- Sweezy, P. (1939) Demand under Conditions of Oligopoly. *Journal of Political Economy*, n. 47, p. 568-573.

Prevenção Estratégica à Entrada

Frederico Rocha

12.1 Introdução

Este capítulo visa à exposição da moderna teoria de prevenção à entrada. A motivação para o tema é fornecida por duas questões distintas associadas à teoria do preço limite, que tem origem nos trabalhos de Bain e Sylos Labini (abordados no Capítulo 7). De um lado, as hipóteses comportamentais presentes nessa corrente de autores eram limitadas, sofrendo críticas dirigidas a seu caráter *ad hoc* e à falta de consistência. De outro, os trabalhos empíricos concluíram pela deficiência explicativa da teoria com respeito a diversos fatos estilizados encontrados.

A nova teoria de prevenção à entrada enfrentou essas questões principalmente com a utilização do conceito de *custos irrecuperáveis* (*sunk-costs*), que passou a ser compreendido como importante ferramenta estratégica à disposição das empresas em processos de competição com rivalidade. As principais conclusões dessas teorias são: o questionamento teórico do paradigma estrutura-conduta-desempenho ao demonstrar que, em grande medida, posicionamentos estratégicos das empresas podem alterar a estrutura do mercado; a eleição dos custos irrecuperáveis como importante instrumento de concorrência e o retorno aos estudos de caso, em detrimento das análises longitudinais comuns no pós-guerra.

O capítulo está dividido em três seções. Na primeira, o conceito de custos irrecuperáveis é apresentado como importante gerador de assimetrias entre as empresas, ao mesmo tempo em que são levantadas as limitações do postulado de Sylos e, por consequência, do paradigma do preço limite. A segunda seção é voltada à exposição do modelo de impedimento à entrada de Dixit, formulando-se uma taxinomia de comportamento estratégico. A terceira seção se dedica à exposição de problemas onde há assimetrias de informação.

12.2 Custos Irrecuperáveis e Assimetrias de Custos

12.2.1 FATOS ESTILIZADOS

A literatura de Economia Industrial ressalta alguns fatos estilizados que levantam dúvidas sobre a suficiência do arcabouço teórico de estrutura-conduta-desempenho (ECD) explorado na Parte II deste livro para explicação das configurações industriais e seu desempenho nas economias modernas. De um lado, há comprovação empírica de relação entre elevada escala mínima eficiente da planta e alta concentração de mercado, ou seja, neste caso a estrutura de custos, influencia a estrutura do mercado. Ao mesmo tempo, a evidência empírica aponta que a ampliação do tamanho do mercado não altera os seus níveis de concentração, sugerindo que tamanho de mercado e tamanho da empresa tendem a ser positivamente

correlacionados. Estudos internacionais mostram que países com mercados de tamanho substancialmente diferente apresentam níveis de concentração semelhantes. Os estudos empíricos encontram ainda que a relação entre economias de escala e concentração sugeriria níveis de concentração muito mais reduzidos do que aqueles efetivamente encontrados nas economias desenvolvidas.

O preenchimento desta lacuna é feito, dentre outros, por Alfred Chandler¹ ao enfatizar a importância da primeira empresa a se mover na determinação da estrutura de mercado. De acordo com o autor, os primeiros a explorar economias de escala e escopo e investir em ativos específicos relacionados com organização gerencial, distribuição e pesquisa adquirem potentes vantagens competitivas ou vantagens da primeira empresa a se mover. A atenção se desloca da determinação da conduta e do desempenho pela estrutura para os movimentos estratégicos da empresa no impedimento da entrada, explicando, assim, as distintas configurações industriais.

12.2.2 DEFINIÇÃO DE CUSTOS IRRECUPERÁVEIS

O investimento em plantas e capacitação específica para determinadas atividades têm, por consequência, a aquisição de ativos (físicos ou humanos) que não podem ser transacionados sem perda total ou parcial de seu valor. Esta é a ideia relacionada a *custos irrecuperáveis*, que são despesas realizadas cujo *custo de oportunidade* de sua utilização é igual ou próximo a zero.

Suponhamos que uma empresa que atue no comércio resolva abrir uma loja em um centro comercial. A empresa deve encontrar um local onde se instalar. Para tal, ela deverá alugar um estabelecimento no centro comercial; deverá assinar um contrato, que conterá uma cláusula de resarcimento em caso de rompimento, por um prazo – digamos de um ano. Suponhamos ainda que o sistema legal do país funcione adequadamente a ponto de conseguir impor aos signatários do contrato todas as suas cláusulas e que a cláusula de rompimento preveja o pagamento por qualquer uma das partes de todas as mensalidades até o final do contrato. Suponhamos que, seis meses após ter assinado o contrato, o empresário descubra que seu negócio não é rentável. Ainda que decida fechar o estabelecimento, o empresário não poderá recuperar os gastos de aluguel comprometidos até o final do contrato. Isto quer dizer que independente de sua vontade, ele estará incorrendo em um custo equivalente a seis meses de aluguel. Esta é a essência dos custos irrecuperáveis. Uma vez que a empresa tenha se comprometido com eles, não pode voltar atrás em sua decisão. Dessa maneira, mesmo que se constitua em um custo para a empresa, a despesa realizada ou por realizar não se constitui em um *custo econômico strictu sensu* por não poder ter uso alternativo.

No exemplo citado, deve-se observar que a característica fundamental do custo irrecuperável é que o momento de decisão de gasto é distinto de seu período de utilização. O conceito de custos irrecuperáveis tem uma dimensão essencialmente temporal. Não se deve, portanto, confundir custos irrecuperáveis com custos fixos. Os custos fixos não variam com o volume de produção, ao passo que os custos irrecuperáveis não podem ser alterados com o tempo e, portanto, uma vez tendo a empresa realizado o dispêndio neles, ela não poderá alterá-los.

Em razão de serem definidos por intermédio de seu custo de oportunidade, os custos irrecuperáveis estão geralmente – não necessariamente – associados a *ativos específicos*, porque estes normalmente apresentam dificuldade para encontrar utilização alternativa. Isto pode ser observado comparando-se dois tipos de investimento em capacidade produtiva. Suponhamos que uma empresa tenha realizado um investimento na construção de uma planta de energia ao lado de uma mina de carvão. Caso o negócio não se verifique rentável, a opção de alocar os recursos mobilizados na planta para uso alternativo parece pouco provável, uma vez que ela foi desenhada em um local específico para um fim também específico. Suponhamos, alternativamente, que uma empresa compre uma loja em um centro comercial com o objetivo de utilizá-la para vender roupas. Caso o negócio não seja bem-sucedido, não é difícil encontrar utilização alternativa para o ativo comprado, como, por exemplo, convertê-lo em um restaurante. Assim, o investimento na planta de energia deve ser entendido como um custo irrecuperável e a compra da loja, não.²

12.2.3 UMA NOVA ABORDAGEM DO PREÇO LIMITE: CUSTOS FIXOS E IRRECUPERÁVEIS

A incapacidade de entrar e sair da indústria sem custos conduz ao aperfeiçoamento (ou viabilização) do que se conheceu como paradigma do preço limite apresentado no Capítulo 7 e que tinha como principal foco de análise a formação de preços em indústrias que contavam com a presença de substanciais descontinuidades tecnológicas. Ainda que a indústria contasse com substanciais economias de escala, uma empresa poderia, teoricamente, adentrar uma indústria a qualquer momento produzindo no nível de custo médio mínimo e especificar abaixo dos preços das empresas estabelecidas nesta indústria, retirando, portanto, qualquer vantagem das empresas já estabelecidas. A presença de investimentos específicos é, portanto, necessária para que as

ELSEVIER

empresas não possam sair sem incorrer em perdas, penalizando a empresa que ingressar na indústria em busca de lucros de curto prazo.

A Figura 12.1 apresenta uma situação em que esta questão pode ser visualizada com maior clareza. Existe apenas uma empresa estabelecida no mercado com uma função de custos formada por $C(q) = F + cq$, gerando uma curva de custos médios (CMe_{EST}) decrescente que se aproxima assintoticamente da curva de custos marginais (CMg), constantes ao longo de toda a região relevante. Esta empresa enfrenta uma curva de demanda inversa linear (D), formada por $P(q) = A - bQ$ também representada na figura, junto com a receita marginal (RMg). Em uma situação sem concorrente potencial, a empresa igualaria a receita marginal aos custos marginais, estabelecendo preço P_M e vendendo q_M unidades, maximizando seus lucros no curto e no longo prazos.

A possibilidade de entrada gera, no entanto, uma situação um pouco diferente. Supondo a inexistência de outro concorrente na indústria, mas a presença de concorrência potencial de outra empresa que detém a mesma tecnologia, não necessariamente maximizar os lucros no curto prazo implica maximizá-los no longo prazo. Nesta figura, a empresa estabelecida não detém qualquer vantagem de custos e os produtos não são diferenciados, impossibilitando que o consumidor faça distinção entre eles.

Adiciona-se, entretanto, o postulado de Sylos que afirma que a empresa estabelecida manterá a quantidade produzida anteriormente após o ingresso da nova concorrente, independentemente da quantidade a ser escolhida pela entrante (ou seja, ela age como uma líder de Stackelberg).³ Isto implica que mantido o preço que maximiza os lucros no curto prazo, ou seja, P_M , existe espaço para a entrada de uma nova empresa, produzindo a preços inferiores.

Aceitando-se que as empresas estejam concorrendo por quantidades, como é usual em Stackelberg, e sendo a função de lucro da empresa entrante representada por $\pi_{ENT}(q_{ENT}, q_{EST}) = [A - b(q_{ENT} + q_{EST})]q_{ENT} - (F + cq_{ENT})$, a função de reação da empresa entrante será determinada por:

$$\frac{\partial \pi_{ENT}}{\partial q_{ENT}} = A - bq_{EST} - 2bq_{ENT} - c = 0 \Rightarrow q_{ENT} = \frac{A - c - bq_{EST}}{2b},$$

sempre que $\pi_{ENT} \geq 0$, ou seja,

$$\pi_{ENT}(q_{ENT}, q_{EST}) = [A - b(q_{ENT} + q_{EST})]q_{ENT} - (F + cq_{ENT}) \geq 0, \text{ que substituindo } q_{ENT} = \frac{A - c - bq_{EST}}{2b}, \text{ resulta em } q_{EST} \leq Y \leq \frac{(A - c)}{b} - 2\sqrt{\frac{F}{b}}.$$

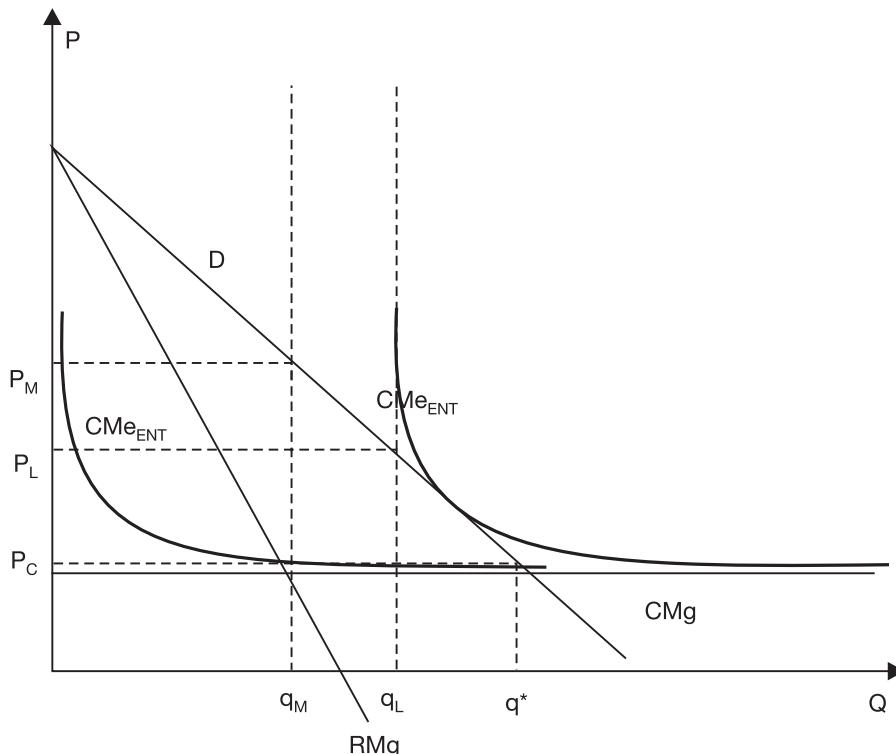


Figura 12.1 Estabelecimento do preço-limite.

Como resultado, o entrante em potencial terá a seguinte função de reação:

$$q_{ENT} = \begin{cases} \frac{A - c - bq_{EST}}{2b}, & \text{se } q_{EST} \leq Y = \frac{(A - c)}{b} - 2\sqrt{\left(\frac{F}{b}\right)} \\ 0, & \text{para outras quantidades} \end{cases}$$

A empresa estabelecida pode, então, escolher entre maximizar o lucro no curto prazo e aceitar a entrada da nova empresa, ou estabelecer a quantidade $q_{EST} = \frac{(A - c)}{b} - 2\sqrt{\left(\frac{F}{b}\right)}$, impedindo a entrada da concorrente em potencial, conforme pode ser visto na Figura 12.2. A opção é entre uma situação em que a empresa obtém lucros elevados em um primeiro período e depois obtém lucros próximos a zero e outra em que os lucros são mais estáveis nos diferentes períodos em questão. A escolha dependerá do diferencial de lucros nas duas situações e da taxa de desconto aplicada para os períodos futuros.⁴

No Capítulo 7, distinguiram-se quatro situações quanto à entrada de concorrentes potenciais. A entrada poderia estar bloqueada, eficazmente impedida, ineficazmente impedida ou livre. O primeiro e o último casos são de pouca relevância, visto tratarem-se ou de monopólio natural ou de um mercado concorrencial. As situações relevantes aparecem quando a entrada está eficazmente ou ineficazmente impedida. Neste último caso, a empresa estabelecida decidirá maximizar seus lucros no curto prazo; no anterior, a empresa estabelecida decidirá maximizar os lucros no longo prazo, prevenindo entradas.

Deve-se ressaltar que se a empresa puder vender sem perda de valor o equipamento ou dispensar os recursos humanos contratados sem custos adicionais representados em F , ela poderá ingressar no mercado por um período, vender ao preço P_C da Figura 12.1 e depois se retirar, vendendo os ativos adquiridos. *Isto sugere que mais do que as descontinuidades tecnológicas ntadas por Sylos Labini, os custos fixos devem ser irrecuperáveis para que as barreiras à entrada sejam viáveis.* Conforme ressaltado por Tirole (1988:308-11), a presença de *custos irrecuperáveis* impossibilita o ajuste mais rápido de quantidades em relação a preços.

12.2.4 CREDIBILIDADE E POSTULADO DE SYLOS

O postulado de Sylos formula uma hipótese comportamental questionável. Um dos problemas do postulado está em destituir as concorrentes em potencial da possibilidade de agir por conta própria, forçando-as a adotar posição passiva de seguidoras. Somente em situações muito especiais isto ocorreria. Isto pode ser exemplificado pelo modelo de liderança de preços apresentado no Capítulo 10. Naquela situação é razoável pensar que as empresas da margem se adaptarão às decisões tomadas pela líder por haver assimetria nos custos. Isto não parece plausível, contudo, quando as concorrentes são empresas grandes com custos

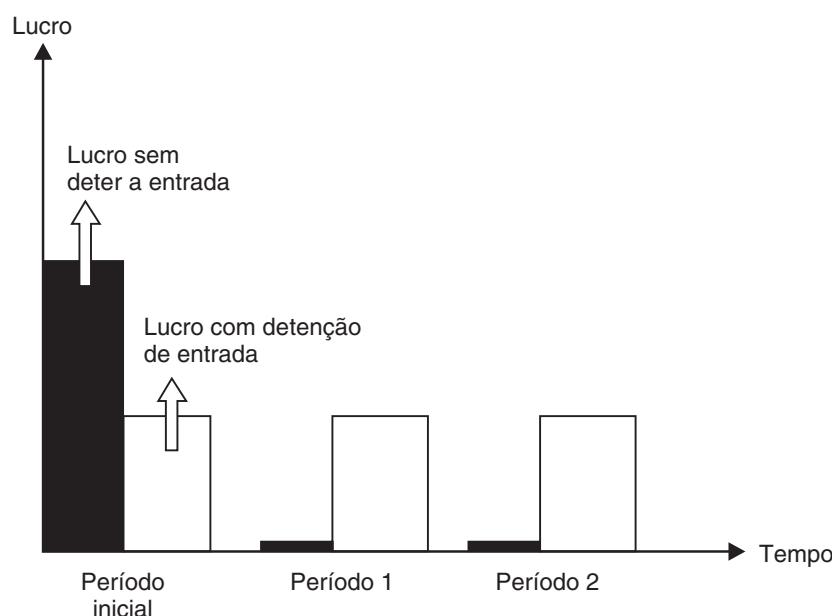


Figura 12.2 Lucro da firma estabelecida com e sem detenção da entrada.

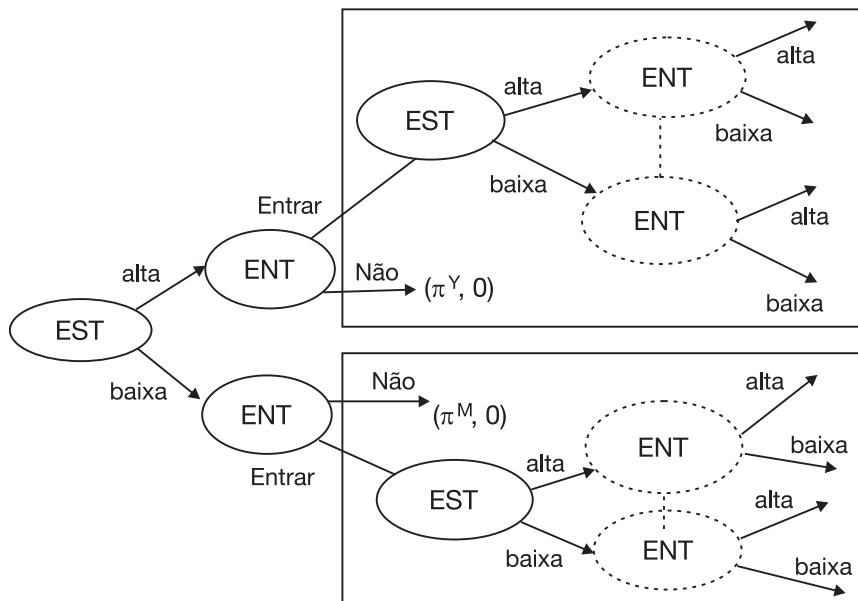


Figura 12.3 Árvore de jogo oligopolista.

simétricos aos das empresas estabelecidas. Seria esperado que a decisão de entrada não fosse tomada por uma regra *ad hoc* que não assegura, em qualquer nível, que o comportamento será ótimo. Não se pode, portanto, falar de um equilíbrio de Nash. Fica evidente a fraqueza do postulado de Sylos quando a empresa estabelecida se confronta com um pequeno número de grandes empresas potenciais entrantes. É, por conseguinte, mais plausível adotar a hipótese de igualdade de condições após a entrada com respeito ao processo decisório.⁵ A Figura 12.3 apresenta um jogo em forma extensiva em três estágios. No primeiro, a empresa estabelecida pode produzir em duas quantidades, alta, representada por $Y = \frac{(A-c)}{b} - 2\sqrt{\frac{F}{b}} > q_M$, ou baixa, representada por q_M . No segundo estágio, a potencial entrante pode escolher ingressar no mercado ou não. Caso não entre, o jogo termina com a empresa estabelecida produzindo a quantidade anteriormente escolhida; caso ingresse no mercado, o jogo caminha para o terceiro estágio, onde as duas empresas concorrem por quantidades.

Resolvendo o jogo por indução para trás, percebe-se que o resultado dos dois subjogos do terceiro estágio, apresentados dentro dos retângulos, é o mesmo. Ambas as empresas escolhem quantidades de Cournot,⁶ aqui representadas por $q_{COURNOT} = \frac{A-c}{3b}$, ou seja, não importa se o jogo prossegue pela sequência em que a empresa estabelecida define quantidades Y ou q_M , as quantidades após a entrada serão as mesmas. Isto ocorre porque a ameaça de manter a quantidade anterior à entrada não é crível.⁷

A questão está em que, apesar de apresentar problemas, é possível verificar situações em que as empresas atuam com preços inferiores aos de monopólio e superiores aos concorrentes e em regiões inelásticas da curva de demanda.⁸ Dois tipos de solução são normalmente apontados para este problema:

1. A utilização de custos irrecuperáveis na tentativa de criar comprometimento entre as ações do período após a entrada e as ameaças do período anterior à entrada; e
2. A busca de algum problema de imperfeição de informação que dê à empresa estabelecida a alternativa de tentar enganar a potencial entrante.

12.3 Custos Irrecuperáveis e Barreiras à Entrada

12.3.1 O MODELO DE DIXIT

O modelo de Avinash Dixit de 1980, assim como a seção anterior, parte de uma situação em que existe uma empresa estabelecida e uma empresa entrante, tratando-se, portanto, de um duopólio. Existem apenas dois períodos em questão: o período 0, anterior à entrada, e o período 1, posterior à entrada. A função custos apresenta uma pequena alteração. Em vez de se dividir em custo fixo e custo variável, o modelo apresenta três tipos de custo: custo fixo, custo da capacidade instalada e custo de produção. Os custos

unitários de produção e de capacidade instalada são constantes ao longo da região relevante, de maneira que a função custos da empresa “*i*” tem a seguinte formulação:

$$C_i = F + kK_i + cq_i,$$

onde F é o custo fixo, K_i é a capacidade produtiva da empresa “*i*” e q_i , sua quantidade de produção; k e c são, respectivamente, o custo unitário da capacidade produtiva e de produção. Por questão de simplificação, adota-se uma curva de demanda linear de tal maneira que a função inversa da demanda da indústria, como anteriormente, é representada por $P(Q) = A - bQ$.

No primeiro período, a empresa estabelecida escolhe uma capacidade produtiva \bar{K} , que servirá para produzir neste período e permanecerá para o seguinte. Esta capacidade poderá (ou não) ser acrescida no próximo estágio. No segundo período, a empresa entrante escolherá se ingressa na indústria, concorrendo em um duopólio de Cournot com a empresa estabelecida, ou não.

A primeira observação a ser feita, neste esquema, é a descontinuidade da função custos da empresa estabelecida no segundo período, determinada por:

$$C_{EST} \begin{cases} F + k\bar{K} + cq_{EST}, \text{ para } q_{EST} \leq \bar{K} \\ F + (k+c)q_{EST} \text{ para } q_{EST} > \bar{K} \end{cases}$$

gerando uma descontinuidade nos custos marginais em que até o nível de capacidade escolhido no primeiro período os custos marginais serão representados por c e, depois, por $k + c$. De maneira semelhante, a empresa entrante deverá ter a função custos representada por $C_{ENT} = F + (k + c)q_{ENT}$ em toda a faixa relevante, ficando clara a assimetria de custos entre as concorrentes, fruto apenas da vantagem da primeira empresa a se mover detida pela empresa estabelecida.

Esta assimetria de custos gera consequências para o jogo de Cournot após a entrada. Quando os dois custos marginais são representados por $k + c$, as duas curvas de reação são definidas como

$$R_i(q_j) = \frac{a - (k + c) - bq_j}{2b},$$

ao passo que quando a empresa estabelecida já instalou sua capacidade e incorre em custos marginais mais reduzidos, sua curva de reação é

$$R_i(q_j) = R_{EST}(q_{ENT}) = \frac{a - c - bq_{ENT}}{2b}.$$

Assim, a prévia instalação de capacidade tem como principal consequência o deslocamento da curva de reação da empresa estabelecida de R_{EST} para R'_{EST} , conforme pode ser visto na Figura 12.4. Além da descontinuidade na curva de reação da empresa estabelecida, representada pela linha grossa no gráfico, percebe-se que a empresa estabelecida na curva de menor custo marginal sempre responderá com maior quantidade às decisões de produção da empresa entrante. Isto significa que a ameaça de a empresa estabelecida utilizar maior quantidade é crível.

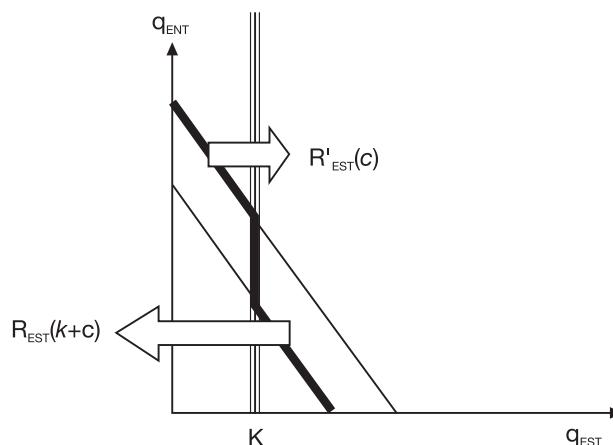


Figura 12.4 Curva de reação da empresa estabelecida.

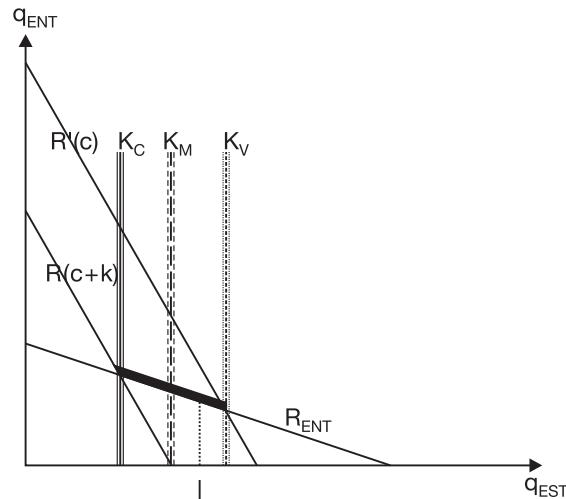


Figura 12.5 Curvas de reação da empresa estabelecida e da entrante.

A Figura 12.5 contempla, além das duas curvas de reação (com capacidade instalada no primeiro período e sem capacidade instalada no primeiro período) da empresa estabelecida, a curva de reação da empresa entrante. O segmento de linha mais grossa sobre a curva de reação da empresa entrante representa os possíveis equilíbrios de Nash para diferentes níveis de capacidade instalada no primeiro período. Assim, se o nível de capacidade escolhido pela empresa estabelecida, no primeiro período, for K_C , representado na Figura 12.5 pela linha dupla contínua, as quantidades de produção obtidas pelas duas empresas serão iguais às quantidades de Cournot; se a empresa estabelecida implantar capacidade igual à quantidade de monopólio, representada pela linha dupla tracejada na Figura 12.5, ela produzirá a quantidade de líder de Stackelberg, ao passo que a empresa entrante produzirá a quantidade de seguidora de Stackelberg; se a empresa estabelecida escolher implantar capacidade K_V , representada pela linha dupla pontilhada na Figura 12.5, ela irá produzir quantidade igual à capacidade instalada, ao passo que a entrante irá produzir uma quantidade que será denominada de V_2 ; qualquer nível de capacidade superior a K_V terá V_2 sempre como resposta por parte da empresa entrante, deixando à empresa estabelecida a alternativa de produzir quantidade equivalente a K_V , ou seja, implantação de capacidade, no primeiro período, superior a K_V gerará capacidade ociosa.

A observação importante em todos estes casos é que entre K_C e K_V a adição de capacidade por parte da empresa estabelecida dá credibilidade em níveis de produção no período após a entrada. Isto significa que se a quantidade que detém a entrada, Y , conforme colocado anteriormente, estiver situada entre estes dois pontos, a empresa estabelecida poderá deter a entrada implantando capacidade. Quando a quantidade que detém a entrada

$$Y = \frac{(A - (c + k))}{b} - 2\sqrt{\left(\frac{F}{b}\right)} > q_M$$

estiver à esquerda de K_C , a entrada não ocorrerá independente da ação da empresa estabelecida, ao passo que, se Y estiver à direita de K_V , a entrada estará *livre*. Os casos interessantes estão, portanto, entre K_C e K_V . O primeiro caso ocorre quando $Y < q_M$, ou seja, a quantidade que detém a entrada é menor do que a quantidade que a empresa estabelecida escolheria se fosse atuar como monopolista no primeiro período. Nesta situação, ela deverá escolher a capacidade produtiva K_M , detendo a entrada e maximizando os lucros no curto prazo. Assim, para qualquer Y inferior a K_M , a entrada estará *blockeada*. Logo, os casos relevantes de impedimento da entrada se darão quando Y se situar entre K_M e K_V . Suponha que se a empresa implantar a capacidade I , ela detém a entrada. Neste caso, deve-se comparar a lucratividade da *empresa estabelecida* nas duas situações alternativas. Primeiro, se ela se desvia da maximização dos lucros no curto prazo, instalando capacidade I . Nesta situação, a empresa entrante decide permanecer fora do mercado e a empresa estabelecida produz quantidade igual a I no primeiro e no segundo períodos. Alternativamente, a empresa estabelecida pode escolher a capacidade produtiva de monopólio K_M , maximizar seus lucros no primeiro período, produzindo a quantidade de monopólio; no segundo período, a empresa produzirá a quantidade de líder de Stackelberg (igual à de monopólio e, no caso específico de demanda linear $q_M = \frac{A - c}{2b}$) e a empresa entrante decidirá ingressar no mercado, produzindo a quantidade de seguidora de Stackelberg,

$$q_S = \frac{A - c}{4b}. \text{ Se } \pi_{EST}(q_M, 0) + \delta\pi_{EST}(q_M, q_S) < \pi_{EST}(q_I, 0) + \delta\pi_{EST}(q_I, 0),$$

então a entrada estará *eficazmente impedida*; caso contrário, a entrada estará *ineficazmente impedida*.

Deve-se observar, no entanto, que, no caso proposto, ainda que a entrada esteja ineficazmente impedida, é interessante a empresa implantar capacidade no primeiro período até a quantidade de Stackelberg, fazendo com que a empresa entrante estabeleça a quantidade de seguidora.

12.3.2 CONCORRÊNCIA POR PREÇOS

Dixit propõe três extensões para seu modelo: (1) o entrante adquire a condição de líder por quantidades, após a entrada; (2) custos crescentes de capacidade produtiva; e (3) concorrência por preços (Bertrand) após a entrada. As duas primeiras servem como sugestão de exercício para o estudante. A terceira é, contudo, de especial interesse para este capítulo. Quando uma empresa faz investimento em capacidade produtiva, seus custos marginais se reduzem, permitindo que ela pratique preços mais reduzidos; no entanto, a realização do investimento tem de ser no exato valor da diferença entre o custo marginal sem investimento e o custo marginal com investimento, multiplicado pela quantidade produzida. Isto pode ser visto na Figura 12.6, em que se pode identificar o custo do investimento representado pela área hachurada ($A + B$). Como resultado, ainda que a escolha com expansão de capacidade seja racional, a empresa que investiu em capacidade acima da quantidade de monopólio estará, no primeiro período, obtendo lucros inferiores aos máximos.⁹

A concorrência de Bertrand, conforme visto no Capítulo 10, gera curvas de reação positivamente inclinadas, como na Figura 12.7. Percebe-se que uma redução do custo marginal permite que a empresa estabelecida seja mais agressiva na concorrência por preços, ao deslocar sua curva de reação para a esquerda. Assim, $R(c)$ é sua função de reação quando investimentos são realizados no primeiro período, $R(c+k)$, caso contrário, R_{ENT} é a função de reação da empresa entrante. A linha grossa é a função de reação da empresa quando é feito um nível específico de investimento, \bar{K} . Percebe-se que há uma quebra na função de reação da empresa estabelecida. A região da curva de reação da entrante que está em linha dupla representa os preços que podem ser escolhidos pela empresa estabelecida e que impedirão a entrada.

12.3.3 ACOMODAÇÃO DA ENTRADA

Existe, no entanto, uma importante diferença entre os resultados obtidos por concorrência por preços e concorrência por quantidades na situação de *entrada ineficazmente impedida*. Enquanto se pode concluir que, no segundo caso, a empresa deve manter a sua agressividade em relação à concorrente, adicionando capacidade, no primeiro caso a adição de capacidade quando a entrada está ineficazmente impedida não está necessariamente assegurada. Na Figura 12.7, quando a empresa não adiciona capacidade no primeiro período, ela estará dividindo o mercado ao preço p_A^1 ; caso ela adicione capacidade, seu preço será inferior, podendo chegar, no caso do investimento máximo, a p_B^1 , com o mercado também dividido. Como os custos totais, conforme discutido na Figura 12.6, permanecem os mesmos, o resultado é uma redução dos lucros.

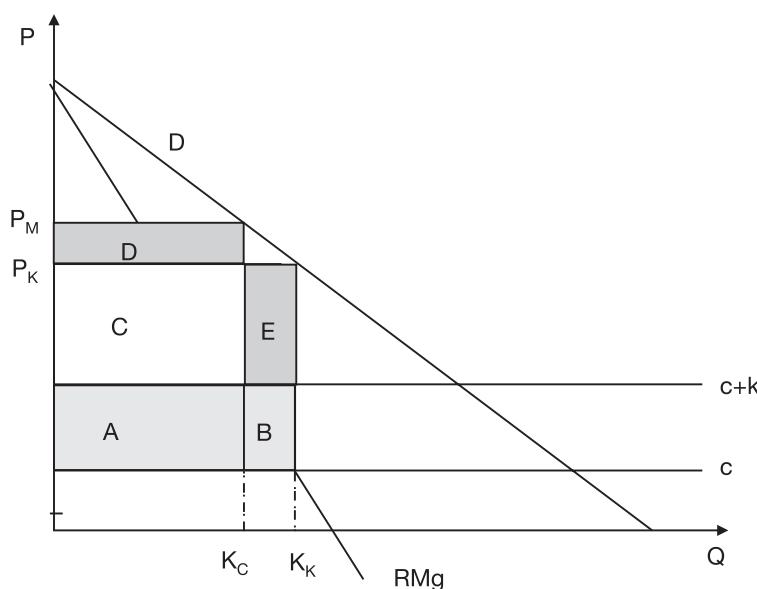


Figura 12.6 Custos no segundo período, com e sem investimento adicional em capacidade.

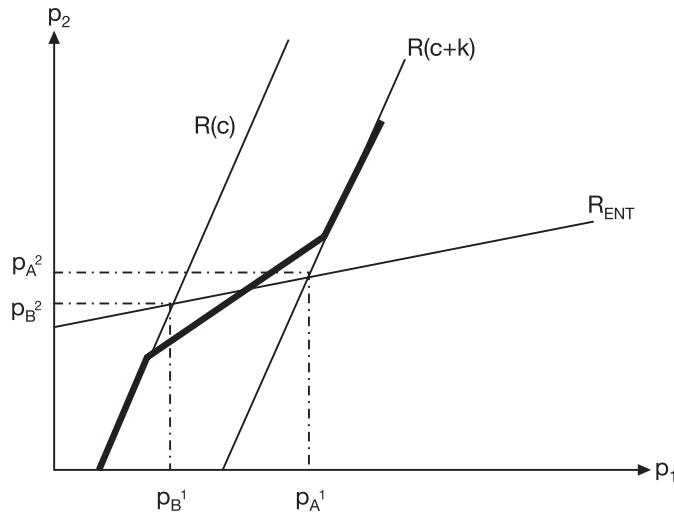


Figura 12.7 Curvas de reação com concorrência por preços após a entrada.

A explicação para a diferença de comportamento da empresa estabelecida está no efeito estratégico do investimento em capacidade. O investimento em capacidade torna menores os custos marginais da empresa estabelecida. Isto fortalece sua posição em um processo de concorrência, pois, por um lado, em concorrência por preços, custos inferiores implicam possibilidade de aplicar preços menores e, por outro lado, em concorrência por quantidades, torna consistente a escolha de quantidades mais elevadas. Em ambos os casos, o efeito estratégico sobre a empresa entrante é a redução de seus lucros. Esta redução pode ser suficiente (ou vantajosa para a empresa estabelecida) para deter a entrada de novos concorrentes, ou não. Se a entrada for *eficazmente impedida*, o investimento será realizado.

A diferença dos dois processos de concorrência (preços e quantidades) se dá quando a entrada está *ineficazmente impedida*. Nesta situação, a empresa estabelecida deve escolher a melhor estratégia para *acomodar a entrada*. Se a concorrência se der por quantidades, a elevação da quantidade da empresa estabelecida irá implicar redução da quantidade da empresa entrante, ou seja, quantidades são substitutos estratégicos, isto é, as curvas de reação são negativamente inclinadas, ou seja,

$$\frac{dq_i}{dq_j} < 0.$$

Uma redução da quantidade da empresa entrante é benéfica à empresa estabelecida, porque implica possibilidade de aumento de sua própria quantidade. Logo, em situações de concorrência por quantidade, a empresa estabelecida adotará estratégias *agressivas* em termos de capacidade, procurando produzir maior quantidade do que a entrante. Se a concorrência for por preços, a redução dos preços da empresa estabelecida, fruto da redução de seus custos marginais, implicará redução dos preços da empresa entrante. A redução de preços da empresa entrante tem efeito negativo sobre a empresa estabelecida, ou seja, os preços são estrategicamente complementares

$$R'_i(p_j) = \frac{dp_i}{dp_j} > 0.$$

Isto significa, portanto, que reduções no preço da empresa entrante também terão como resposta reduções nos preços da empresa estabelecida, resultando em impactos negativos sobre os lucros. Logo, a estratégia de investir em capacidade adicional até a quantidade de monopólio pode não ser a melhor alternativa. Isto significa que o melhor movimento estratégico para a empresa estabelecida na indústria, no caso de concorrência por preços após a entrada de uma nova concorrente, pode ser não adicionar a capacidade. De fato, no Capítulo 10 foi visto que a restrição de capacidade pode ser uma boa forma de atingir lucros altos quando a concorrência de Bertrand prevalece após a entrada. Isto sugere que a empresa estabelecida pode, no primeiro período, desviar da maximização de lucros no curto prazo, implantando capacidade abaixo da ótima no curto prazo para possibilitar preços superiores no segundo período.

É interessante perceber, no entanto, que os elementos de concorrência não se limitam a preços e quantidades. Há outras maneiras em que as empresas concorrem, como, por exemplo, diferenciação de produto. Pode-se, de uma maneira geral, pensar

que quando o elemento da concorrência for estrategicamente substituto, ou seja, apresentar curvas de reação negativamente inclinadas, a melhor estratégia de acomodação da entrada será realizar o investimento e agir *agressivamente*. De maneira alternativa, quando os elementos da concorrência forem estrategicamente complementares, ou seja, tiverem curvas de reação positivamente inclinadas, a melhor estratégia para acomodar a entrada será ser *permissivo* e investir pouco em capacidade.

12.3.4 CUSTOS IRRECUPERÁVEIS E ESTRATÉGIAS COMPETITIVAS

Devemos também ressaltar que o investimento em capacidade é apenas uma entre as diversas possibilidades de se realizar dispêndios em custos irrecuperáveis. Investimento em P&D, capacitação da mão de obra, fixação da marca, entre outros, são maneiras de realizar gastos em custos irrecuperáveis, prevenir a entrada de empresas potenciais rivais e dar consistência a posicionamentos estratégicos e ameaças.

Um exemplo de marca fortemente consolidada pode ser dado no mercado de telecomunicações no Brasil. A Embratel deteve o monopólio das chamadas interurbanas e internacionais no Brasil por muitos anos. A liberalização do mercado para novas empresas nos anos 1990 possibilitou que concorrentes surgissem. Um dos maiores concorrentes foi a Intelig. Ainda que pesadas campanhas publicitárias tivessem sido realizadas tentando fixar o nome da empresa e ensinar um número de discagem, os serviços da Intelig permanecem sendo menos utilizados do que os da Embratel pelo simples fato de que os consumidores desconheciam ou estão pouco acostumados com a nova empresa.

A decisão do CADE no processo de fusão da Brahma e da Antarctica para formação da Ambev também levou em consideração elementos referentes à marca. Os reguladores entenderam que algum espaço deveria ser aberto para a entrada de novos concorrentes, uma vez que as principais empresas presentes no setor estavam se juntando. Assim, determinaram a venda da marca Bavaria já consolidada, para que outra empresa pudesse entrar no mercado sem desvantagens adicionais. Neste mesmo processo, pode-se perceber outra fonte de assimetrias entre empresas estabelecidas e entrantes. Para comercializar a Bavaria, as autoridades de defesa da concorrência determinaram que a Ambev disponibilizasse ao novo detentor da marca sua rede de distribuição por um período de cinco anos.

A presença de diferentes instrumentos em que se possa investir em custos irrecuperáveis pode, então, influenciar pesadamente sobre a estrutura de mercado que deverá prevalecer, na medida em que funciona como uma maneira de deter a entrada. Assim, mercados em que é possível a realização constante de investimentos em custos irrecuperáveis devem apresentar estruturas mais concentradas do que mercados em que a realização de investimentos em custos irrecuperáveis é apenas ocasional. J. Sutton distingue os custos irrecuperáveis em duas categorias: (1) exógenos; e (2) endógenos. Os custos irrecuperáveis exógenos são representados por investimentos que são realizados apenas uma vez, não tendo desdobramentos. Isto ocorre, por exemplo no investimento em uma planta em processos de alta intensidade de capital. Custos irrecuperáveis endógenos geram desdobramentos porque são parte essencial do processo de concorrência da indústria. Nesta categoria se encontram investimentos em propaganda e P&D. Enquanto barreiras à entrada a partir de custos irrecuperáveis exógenos se esgotam e tendem a desaparecer ou perder sua importância com o crescimento do mercado, a presença de custos irrecuperáveis endógenos permite, em contrapartida, o estabelecimento de barreiras à entrada e à saída de longo prazo. Os resultados empíricos de Sutton (1988) para propaganda e Sutton (1998) para P&D confirmam sua hipótese, mostrando que indústrias onde estas características são relevantes têm maior grau de concentração.

Como consequência deste tipo de análise, além do processo de concorrência funcional, baseada em preços, quantidades, parcelas de mercado, qualidade, entre outros, há um processo estratégico de concorrência, a partir da realização de investimentos em custos irrecuperáveis, criador de irreversibilidades para a empresa estabelecida e de assimetrias entre elas e as empresas entrantes.

Um segundo aspecto dos custos irrecuperáveis é que eles podem ser utilizados não só como instrumentos para prevenir a entrada, mas também como forma de concorrer com outras empresas já presentes no mercado.

12.4 Informação e Barreiras à Entrada

12.4.1 PREÇO LIMITE COM INFORMAÇÃO ASSIMÉTRICA

Na seção anterior, o questionamento do postulado de Sylos, de que as empresas estabelecidas mantêm a quantidade produzida após a entrada, foi superado pelo investimento em custos irrecuperáveis, criadores de assimetrias entre a empresa estabelecida e as concorrentes potenciais. Uma alternativa a esta formulação é a suposição de assimetrias de informação entre as empresas.

O modelo de Milgrom e Roberts propõe que a empresa estabelecida utilizará o preço pré-entrada para influenciar a decisão de entrada de suas concorrentes potenciais; em contrapartida, estas últimas tentarão inferir, a partir dos preços aplicados, o nível de custo da empresa estabelecida. A principal base para este tipo de argumentação continua sendo a curva de reação que a entrante deverá se defrontar. Quando a concorrência após a entrada ocorre por quantidades, que são elementos estrategicamente substitutos, a curva de reação da empresa entrante cruza as duas curvas, gerando, respectivamente, as quantidades q_B e q_A (ver Figura 12.8a). Se a quantidade Y , que detém a entrada da concorrente potencial, estiver entre esses dois pontos, conforme a Figura 12.8a, a entrada da rival estará detida apenas no caso em que a empresa estabelecida tiver custos baixos. A Figura 12.8b apresenta análise semelhante para o caso de concorrência por preços.

Suponhamos que possam existir dois tipos de empresa. É conhecimento comum que há probabilidade ρ de a empresa estabelecida ter custos baixos e probabilidade $(1 - \rho)$ de ter custos altos.¹⁰ A empresa estabelecida conhece seus custos, mas a empresa entrante só conhece a distribuição da probabilidade de que sua rival tenha custos altos ou baixos. A empresa estabelecida deve escolher um preço. O preço alto (p_A) maximiza os lucros no curto prazo se a empresa tem custos altos, ao passo que o preço baixo (p_B) maximiza os lucros se a empresa tem custos baixos. A empresa potencial entrante, após observar a escolha de preços da empresa estabelecida, deve decidir se entra ou não.

A Figura 12.9 expõe um jogo que simula a interação das duas empresas. A linha tripla representa um movimento inicial aleatório que designa o nível de custos da empresa estabelecida. Após este movimento, a empresa estabelecida saberá em que parte da árvore do jogo estará. Se tiver custos baixos, estará no tronco de cima, caso contrário, estará no de baixo. Ciente de seus custos, a empresa estabelecida se movimenta para a esquerda, p_A , ou para a direita, p_B . As duas linhas tracejadas representam o nível de informação da empresa entrante que, apesar de conhecer o lado da figura (esquerdo ou direito), desconhece se ela está enfrentando uma empresa de custos baixos ou altos, ou seja, se está situada no tronco de cima ou no de baixo. Se a informação fosse simétrica, ou seja, se a empresa entrante tivesse conhecimento perfeito sobre os custos da empresa estabelecida, só ingressaria no mercado se a empresa estabelecida tivesse custos altos, pois $\pi_{ENT}(C_B) < 0 < \pi_{ENT}(C_A)$.

A empresa estabelecida tem maiores lucros atuando como monopolista do que como duopolista, independente de seus custos serem altos ou baixos, significando que, em caso de ter custos baixos, ela gostaria de se revelar para a empresa estabelecida. Ao mesmo tempo, se a empresa tivesse custos altos, ser-lhe-ia vantajoso evitar a entrada, de maneira que ela gostaria de enganar a empresa entrante, fazendo-se passar por uma empresa com custos baixos. Dois tipos de equilíbrio são factíveis nesta situação:

1. Equilíbrio com separação – empresas com níveis de custos diferentes lucrariam mais adotando estratégias distintas; e
2. Equilíbrio conjunto (*pooling*) – empresas adotam mesma estratégia independente de seus custos.

Na situação de equilíbrio conjunto, a estratégia anterior da empresa estabelecida não melhora as condições de informação da empresa entrante, de maneira que ela ingressará com probabilidade $1 - \rho$ e não ingressará com probabilidade ρ , sendo sua situação completamente determinada pela distribuição de probabilidade inicial quanto aos custos da empresa estabelecida, impossibilitando ação do preço limite.

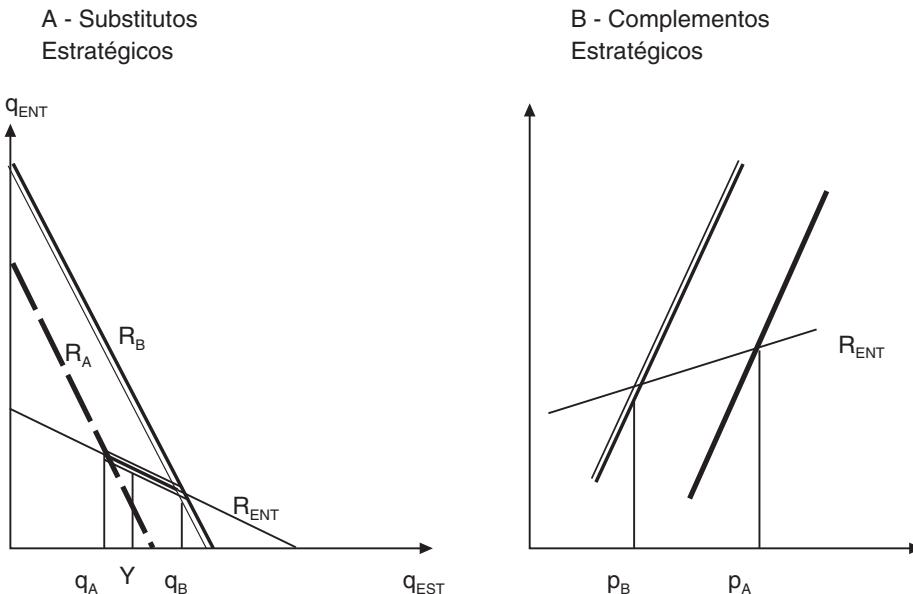


Figura 12.8 Curvas de reação com informação privada.

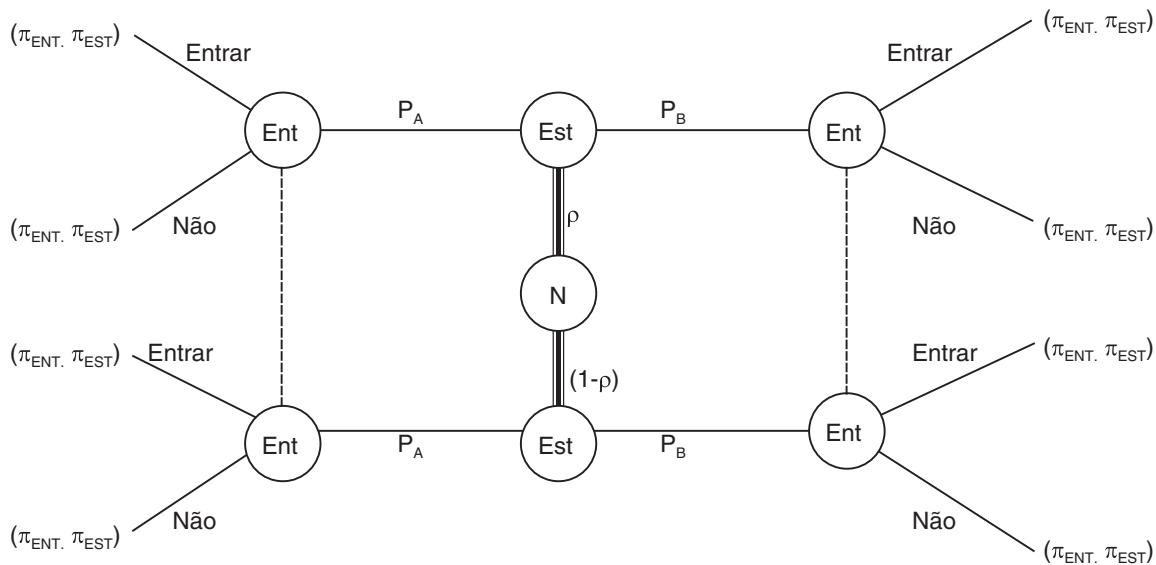


Figura 12.9 Jogo de simulação.

Suponhamos que M^A representa o lucro da empresa estabelecida com custos altos quando atua como monopolista e M^B , seu lucro como monopolista e com custos baixos; D^A é o lucro quando ela é duopolista e tem custos altos e D^B , quando duopolista com custos baixos. Suponhamos que exista um preço limite $P^L < P_M^B$, tal que

$$M^B + \delta D^B \leq M^B(P^L) + \delta M^B \Rightarrow M^B - M^B(P^L) \leq \delta(M^B - D^B) \quad (1)$$

e

$$M^A + \delta D^A \geq M^A(P^L) + \delta M^A \Rightarrow M^A - M^A(P^L) \geq \delta(M^A - D^A) \quad (2)$$

A equação (1) indica que o lucro da empresa estabelecida quando tem custos baixos e aplica o preço de monopólio p_B no primeiro período, permitindo a entrada de sua empresa rival e obtendo lucros de duopólio no segundo período, é inferior ao lucro da empresa estabelecida com custos altos, quando ela aplica o preço que detém a entrada (P^L) no primeiro período e atua como monopolista também no segundo período. A equação (2) expõe uma situação em que a empresa estabelecida tem custos altos e seus lucros quando aplica preço de monopólio no primeiro período – permitindo o ingresso no mercado da empresa entrante e atuando como duopolista no segundo período – serão maiores do que se adotar a alternativa de aplicar o preço limite (que detém a entrada) no primeiro período, mantendo-se como monopolista no segundo.¹¹ Caso as equações (1) e (2) prevaleçam, existirá um equilíbrio com *separação*, ou seja, P^L será um preço que só será adotado pela empresa com custos baixos. Em contrapartida, a empresa com custos altos utilizará seu preço de monopólio. Como consequência, apesar de a empresa estabelecida, no caso de custos baixos, manipular seu preço, a empresa entrante não é enganada. Ela aprende perfeitamente os custos da empresa estabelecida, entrando somente nas condições em que entraria se a informação fosse perfeita.

Suponha agora que a equação (1) seja válida, mas a (2) seja substituída por:

$$M^A + \delta D^A \leq M^A(P^L) + \delta M^A \Rightarrow M^A - M^A(P^L) \leq \delta(M^A - D^A) \quad (3)$$

que indica que a empresa estabelecida, quando tem custos altos, terá maior nível de lucros se adotar o preço limite, detendo o ingresso da potencial entrante, do que se maximizar os lucros no curto prazo, utilizando o preço de monopólio e permitindo o ingresso da empresa potencial entrante. Neste caso, não se distingue as estratégias da empresa estabelecida nos dois casos alternativos de custos altos e baixos, ou seja, a estratégia do preço limite será adotada independente do nível de custos da empresa estabelecida, caracterizando um equilíbrio conjunto ou *pooling*. Neste caso, a taxa de entrada se dá em condições distintas daquelas de informação perfeita. Para a determinação da intensidade de entrada é necessário conhecer as condições de lucratividade da empresa entrante. Suponhamos que

$$\rho(D_{ENT}^B) + (1-\rho)D_{ENT}^A < 0 \quad (4)$$

ou seja, que o lucro em duopólio da empresa entrante, quando a empresa estabelecida tem custos baixos, multiplicado pela probabilidade da empresa estabelecida ter custos baixos, mais o lucro da empresa entrante, quando a empresa estabelecida

tem custos altos, multiplicado pela probabilidade da empresa estabelecida ter custos altos é inferior a zero. Isto significa que a expectativa de lucro da empresa entrante caso escolha ingressar no mercado será negativa. Se as condições (1), (3) e (4) forem válidas, a implementação do preço limite por parte da empresa estabelecida, independente de seu nível de custos, garantirá um nível de entrada inferior ao que ocorre quando a informação é perfeita.

Suponhamos agora que as condições (1) e (3) prevalecem, mas a condição (4) é substituída pela equação (5):

$$\rho(D_{ENT}^B) + (1 - \rho)D_{ENT}^A > 0. \quad (5)$$

Isto significa que a expectativa de lucro da empresa entrante é positiva. Prevalecendo as condições (1), (3) e (5), a empresa entrante não aprende com o preço, mas a distribuição de probabilidade sugere que seu rendimento será superior se ela decidir entrar. Logo, a entrada ocorre com intensidade superior às condições de informação perfeita, não importando a prática de preços da empresa estabelecida. A melhor ação, tanto para empresas de custo baixo, quanto para as de custo alto, é adotar, no primeiro período, o preço que maximiza seus lucros no curto prazo. Devemos notar que, particularmente para as empresas de custo baixo, seria mais vantajoso ter informação perfeita.

12.4.2 CUSTOS DE SAÍDA, REPUTAÇÃO E COMPORTAMENTO PREDATÓRIO

Uma questão enfatizada neste capítulo se refere às hipóteses comportamentais antes da entrada de uma nova empresa no mercado. No entanto, situações relevantes podem ocorrer quando, após o ingresso no mercado, as empresas devem decidir suas estratégias. Suponhamos que, nas condições levantadas anteriormente, no caso de equilíbrio *pooling*, ou seja, prevalecem as condições (1) e (3), a empresa entrante decida ingressar no mercado porque a condição (5) é satisfeita. A empresa estará entrando no mercado desconhecendo o nível de custos da empresa estabelecida. Suponhamos que a empresa estabelecida tenha custos baixos. Nesta situação, se, antes de tomar a decisão de entrada, a informação fosse perfeita, a nova empresa não teria entrado. Uma vez que tenha ingressado no mercado, a imposição de um preço P_L por parte da empresa estabelecida expulsaria a nova concorrente, desde que os custos irrecuperáveis fossem desprezíveis. Esta estratégia é conhecida como *predatória*, que consiste na imposição de preços baixos, que reduzem os lucros de ambas as partes no curto prazo, mas que têm como objetivo elevar os lucros da empresa dominante (ou seja, que adota o comportamento predatório) no longo prazo.

Uma vez que este tipo de comportamento pode ser adotado pela empresa estabelecida com custos baixos, a manutenção de informação imperfeita, ou seja, os custos da empresa estabelecida permanecendo sua informação privada mesmo após a entrada da nova concorrente, a estratégia predatória também poderá ser adotada pela empresa estabelecida, mesmo tendo custos altos. Nestes casos, ganha importância a comparação entre a solidez financeira das duas empresas. O Quadro 12.1 ilustra uma situação de comportamento predatório.

Às vezes uma empresa pode desejar adotar um comportamento de preços baixos independente de visar à expulsão de outra empresa do mercado ou de sinalizar custos baixos, mas apenas indicar que tem comportamento agressivo a futuros concorrentes em potencial. Dessa maneira, os rendimentos que pretende usufruir não serão diretamente obtidos pela saída de uma empresa específica, mas pela manutenção de reputação ao longo dos anos.

QUADRO 12.1 COMPORTAMENTO PREDATÓRIO EM ALGUNS MERCADOS

Um exemplo de comportamento predatório em uma indústria com reduzidos custos irrecuperáveis ocorreu no mercado de transporte aéreo nos EUA durante a década de 1980, quando houve uma pesada desregulação. Após o mercado ter sido desregulado, uma série de novas empresas ingressou com serviços mais baratos, de menor qualidade, mas com preços inferiores. Isto fazia com que as novas linhas aéreas tivessem custos substancialmente inferiores às empresas já estabelecidas. A resposta das grandes linhas aéreas foi um corte imediato de suas tarifas nos trechos de entrada das novas companhias. O resultado foi a incapacidade de sobrevivência de grande parte das empresas novas, que deixaram o mercado após alguns anos. A ocorrência de uma guerra de preços na indústria de café norte-americana durante a década de 1970 também nos informa sobre comportamento predatório. A General Foods controlava 45% do mercado oriental dos EUA por intermédio de sua marca Maxwell House. A marca Folger, controlada pela Procter e Gamble e forte no mercado norte-americano ocidental, resolveu invadir o mercado oriental. A resposta da Maxwell House foram fortes cortes de preço. A guerra de preços foi instalada, concluindo pela permanência da Folger, na década de 1980.

Algumas vezes, comportamentos predatórios são adotados por empresas que ingressam no mercado, desde que tenham maior solidez financeira ou acesso fácil a insumos. Isto ocorreu durante a década de 1990 no Brasil na indústria de laticínios, quando a Parmalat adotou uma estratégia agressiva de preços para abrir espaço no mercado, eliminando as empresas da margem, formada por cooperativas e evitando uma confrontação direta com sua concorrente maior, a Nestlé.

12.5 Conclusão

Ao acentuar a importância das estratégias empresariais na detenção da entrada e nos processos competitivos posteriores, o capítulo advoga a inversão da lógica que vai da determinação da estrutura para a conduta, mostrando que também a conduta irá determinar a estrutura do mercado. Isto pode ser corroborado com importante evidência empírica.

Outro elemento a ser enfatizado é a importância do comprometimento estratégico para dar credibilidade às ações das empresas. Dessa forma, conforme ressaltado por Chandler:

Estas empresas (primeiras a se mover) junto com algumas poucas seguidoras que entraram em seguida na indústria não competiam mais com base em preços. Em vez disso, elas competiam por parcelas de mercado ou lucros por intermédio de medidas funcionais e estratégicas. Eles assim agiram funcionalmente melhorando seus produtos, processos, marketing, política de compras e relações de trabalho, e estrategicamente ao se direcionarem a mercados em crescimento com maior rapidez e se retirarem dos mercados em declínio mais rapidamente e de maneira mais efetiva do que seus competidores. (Tradução livre de Chandler, 1990:8, grifos do original.)

Os elementos funcionais ressaltados por Chandler são justamente aqueles em que os custos irrecuperáveis têm importância. Deve-se, no entanto, ressaltar que os elementos estratégicos no sentido de Chandler não são tratados neste capítulo. Esta conclusão aponta que a maior parte dos elementos de competição aqui considerada se baseia em hipóteses iniciais de assimetria entre os agentes. A partir do momento em que se assume o pioneirismo de uma empresa, ela age para realizar investimentos irreversíveis que dão credibilidade a ameaças de comportamento. Essa ênfase é voltada para situações anteriores ao pioneirismo, que determinam o movimento mais rápido em direção a mercados em crescimento. Neste caso, a análise deste capítulo é limitada, e o recurso à história e a outras teorias em que elementos de racionalidade não substantiva estão presentes parece ser importante. A Parte V deste livro tratará deste assunto.

Notas

1. Chandler, A. *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, The Belknap Press, 1990.
2. Mais uma vez, deve ser ressaltado que os custos irrecuperáveis não ocorrem apenas com ativos específicos como o exemplo do aluguel da loja torna claro.
3. Ver os Capítulos 9 e 10.
4. No geral pode-se supor que o fator de desconto no período t deverá ser igual a δ^t , onde $\delta = \frac{1}{1+r}$, sendo r a taxa de desconto. Normalmente, utiliza-se a taxa de juros como referência para determinar o nível da taxa de desconto, no entanto, outros elementos que não a taxa de juros, tais como progresso técnico e expectativas, podem afetá-la.
5. Isto não significa, portanto, que seria plausível pensar em capacitação semelhante. No entanto, conforme será facilmente perceptível, com a finalidade de simplificar a exposição, os próximos passos serão tomados com conhecimento tecnológico pleno por parte da entrante.
6. Ver Capítulo 10 a respeito de modelos de oligopólio estáticos.
7. Ver a noção de equilíbrio perfeito em subjogo no Capítulo 9.
8. Ver a noção de equilíbrio perfeito em subjogo no Capítulo 9.
9. Deve-se lembrar ao aluno que a receita marginal é negativa nas regiões inelásticas da demanda, sendo um contrassenso que empresas com poder de monopólio atuem de forma a reduzir suas receitas. Logo, não seria esperado que isto ocorresse e uma solução racional deve ser buscada.
10. Caso maximizasse os seus lucros, estaria obtendo o lucro equivalente à área C+D, contra C+E quando tenta deter a entrada.
11. Esta versão é baseada em uma simplificação do modelo de Milgrom, P. e Roberts, J. *Limit Pricing and Entry Under Incomplete Information*. *Econometrica*, v. 50, p. 443-460, 1982, presente em Tirole, J. *The Theory of Industrial Organization*. The MIT Press, Cambridge (Mass.), 1988, p. 368-371. A versão original além de informação privada sobre os custos da empresa estabelecida prevê também informação privada sobre os custos da empresa entrante.
12. Em todos os casos o lucro do segundo período está descontado por δ .

Bibliografia

- Bain, J. (1956) *Barriers to New Competition*. Cambridge, Harvard University Press.
- Chandler, A. (1990) *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, The Belknap Press.
- Dixit, A. (1980) The role of investment in entry deterrence. *Economic Journal*, v. 27, p. 95-106.
- Gilbert, R. (1989) Mobility Barriers and the Value of Incumbency. In Schmalensee, R.; Willig, R. *The Handbook of Industrial Organization*. Amsterdam, North-Holland.
- Jacquemin, A. (1987) *The New Industrial Organization*. MIT Press, Cambridge.
- Milgrom, P.; Roberts, J. (1982) Limit Pricing and Entry Under Incomplete Information. *Econometrica*, v. 50, p. 443-460.
- Scherer F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*. Cambridge (Mass.), Houghton Mifflin.
- Sutton, J. (1988) *Sunk Costs and Market Structure*. Cambridge, MIT Press.
- Sutton, J. (1998) *Technology and Market Structure*. Cambridge, MIT Press.
- Sylos-Labini, P. (1982) *Oligopólio e Progresso Técnico*. Rio de Janeiro, Forense Universitária.
- Tirole, J. (1988) *The Theory of Industrial Organization*. The MIT Press, Cambridge (Mass.).



PARTE IV

A GRANDE EMPRESA CONTEMPORÂNEA

Teoria dos Custos de Transação

Ronaldo Fiani

13.1 Introdução

Até a publicação do artigo pioneiro de Ronald Coase (prêmio Nobel em 1991) em 1937, intitulado *The Nature of the Firm (A natureza da firma)*, a teoria econômica tratava em detalhe apenas dos custos de produção. Embora se reconhecesse a existência também dos custos de transação no sentido estrito de que não apenas o ato de produzir, mas também o ato de comprar e vender acarretava custos, supunha-se em geral que os custos associados às transações econômicas eram negligenciáveis, de tal forma que os únicos custos que realmente importavam eram os custos de produção.

Em função dessa ênfase praticamente exclusiva em custos de produção, a empresa era vista fundamentalmente como uma função de produção: dadas as quantidades de vários insumos por meio de um processo produtivo qualquer, conseguia-se uma quantidade y de produto. Por exemplo, dada uma quantidade de couro, verniz, pregos, eletricidade, horas de trabalho na linha de produção etc., conseguia-se uma certa quantidade de sapatos. Ao economista competia exclusivamente conhecer a relação matemática entre os insumos e o produto para, juntamente com o preço desses mesmos insumos e do produto final, calcular a quantidade a ser produzida que maximizaria o lucro da empresa, isto é, sua quantidade de equilíbrio.

Ronald Coase em seu artigo busca uma definição da empresa “que corresponda ao que ela é no mundo real”. Para isso faz uma pergunta algo primária, mas que, surpreendentemente, nenhum dos mais importantes economistas até aquele momento havia feito: por que existem empresas? Dito de outra forma, por que existem organizações dirigindo o processo produtivo em que relações hierárquicas, definidas pela subordinação dos empregados à direção da empresa, determinam como se deve organizar a produção?

Isso porque a teoria econômica até aquele momento não tinha estudado organizações, apenas mercados. Portanto, havia um vasto material explicando o funcionamento dos mercados, mas não das empresas. Contudo, é comum que uma série de etapas do processo produtivo não passe pelo mercado, mas sejam realizadas dentro das empresas. No exemplo inicial dos sapatos, o corte do couro, sua costura, a aplicação de verniz, a colocação do solado etc. acontecem dentro da empresa, e não no mercado, não obstante sejam etapas produtivas claramente distintas. Assim, é necessário entender por que toda a produção de um produto não pode ser levada a cabo por uma série de agentes individuais, formando uma cadeia em que cada um comprasse daquele que se encontrasse em uma etapa anterior do processo de produção, até que o último produtor na cadeia vendesse o produto acabado ao consumidor final. Em outras palavras, é preciso explicar o que leva diferentes etapas do processo produtivo a serem integradas verticalmente dentro de uma empresa.

A resposta já se encontra, de certa forma, implícita na pergunta: se a cadeia imaginada anteriormente não se concretiza na produção de bens e serviços de natureza complexa, se vivemos em um mundo com grandes unidades fabris e não com indivíduos

que trocam entre si em toda etapa que a divisão das tarefas permitir, é porque estas trocas devem envolver um custo menor quando realizadas no interior das empresas. Em outras palavras, as empresas (organizações que decidem hierarquicamente a alocação dos fatores de produção no seu interior substituindo o mecanismo de mercado) existem porque os custos de transação desse tipo de organização devem ser menores do que os custos de transação no mercado para as mesmas transações.

O artigo de Ronald Coase deu início, dessa forma, ao estudo das condições sob as quais os custos de transação deixam de ser desprezíveis e passam a ser um elemento importante nas decisões dos agentes econômicos, contribuindo para determinar a forma pela qual são alocados os recursos na economia. A análise dessas condições e das consequências dos custos de transação para a eficiência do sistema constituem o objeto da Teoria dos Custos de Transação (TCT).

A esta resposta do porquê da existência de empresas, se segue uma outra pergunta: dado que existem empresas, mas que estas também não são a única forma de organização do processo produtivo, pois há um grande número de transações que se dão pelo mercado, quais são os limites das empresas para organizar a produção? Quando é mais interessante produzir sob a direção de uma hierarquia – isto é, por intermédio da organização da empresa – e quando é mais interessante deixar que o mercado, por meio do mecanismo de preços coordene a produção?

O capítulo está dividido em quatro seções, além desta introdução. Na primeira seção estudaremos a natureza e os fatores determinantes dos custos de transação. Na segunda seção serão analisadas as diferentes naturezas dos contratos. Na terceira detalham-se os tipos de transações e as estruturas de governança. Finalmente, na última seção apresentamos algumas evidências empíricas sobre os custos de transação.

13.2 Natureza e Fatores Determinantes dos Custos de Transação

13.2.1 NATUREZA DOS CUSTOS DE TRANSAÇÃO

Custos de transação são os custos que resultam da divisão do trabalho, ou seja, da divisão de tarefas no processo produtivo. Assim, os agentes enfrentam custos de transação toda vez que recorrem ao mercado para adquirir aquilo que é necessário para a produção de um bem ou serviço. Isso levou alguns teóricos de custos de transação, como Ronald Coase, a identificar custos de transação como custos de recorrer ao mercado. Nesta abordagem, custos de transação são os custos de negociar, redigir e garantir o cumprimento de um contrato, e a unidade básica de análise quando se trata de custos de transação seria o contrato.

Contudo, Oliver Williamson (prêmio Nobel em 2009) percebeu que essa abordagem dos custos de transação como custos de se recorrer ao mercado apresentava um grave problema: ela pressupunha que se não houvesse mercados, não haveria custos de transação! O próprio Ronald Coase afirmou isso, ao alegar que em uma economia socialista os custos de transação seriam nulos.

Oliver Williamson percebeu que os custos de transação possuem uma dimensão mais fundamental, que transcende a presença dos mercados. Custos de transação são os custos de se organizar o funcionamento do sistema econômico, dada a divisão do trabalho – onde por divisão de trabalho entende-se a divisão das tarefas produtivas, seja no interior de uma mesma unidade produtiva (a divisão de tarefas no interior de uma empresa), seja entre diferentes unidades produtivas (uma empresa fornecendo insumos para outra). Ou seja, custos de transação são os custos que resultam quando um ativo – um recurso produtivo – atravessa uma *interface tecnológica* – passa de uma etapa do processo de trabalho para outra. Isso acontece tanto no interior de uma empresa como em uma transação no mercado. Assim, há custos de transação tanto quando uma empresa decide internalizar uma etapa do processo produtivo – produzindo ela mesma o insumo – como quando ela decide adquirir o insumo já pronto de outra empresa. Os custos de transação em um caso e em outro serão, todavia, diferentes, pois a forma de organizar o processo produtivo será diferente em cada caso (em um caso a firma produz ela mesma, em outro recorre ao mercado).

Mas por que os custos de transação serão diferentes de acordo com a forma que se organize o processo produtivo? Pelo fato de que as transações apresentam diferenças nas suas características que afetam os custos de transação, fazendo com que em determinadas transações os custos de transação sejam mais importantes do que em outras. Nos casos em que os custos de transação são elevados, o mercado não se mostra uma forma adequada para se organizar o processo de produção, e outras formas de organizar a produção têm de ser consideradas, como a própria empresa.

A TCT elabora, assim, um conjunto de hipóteses que, caso se verifiquem em uma dada transação, tornam os custos de transação significativos: racionalidade limitada, complexidade e incerteza, oportunismo e especificidade de ativos. Essas hipóteses são os fatores determinantes da existência de custos de transação e serão tratadas em seguida.

13.2.2 FATORES DETERMINANTES DE CUSTOS DE TRANSAÇÃO

13.2.2.1 Racionalidade limitada, complexidade e incerteza

O ponto de partida no tratamento da questão dos custos de transação pela TCT é o reconhecimento, a partir dos trabalhos de H. Simon, de que o comportamento humano, ainda que sendo intencionalmente racional, enfrenta limitações. Estas limitações possuem fundamentos neurofisiológicos (que limitam a capacidade humana de acumular e processar informações) e de linguagem (que limitam a capacidade de transmitir informações). Caso a racionalidade humana fosse ilimitada, qualquer contrato poderia incorporar cláusulas antecipando qualquer circunstância futura e a forma de organizar as transações perderia importância: não haveria empresas, apenas transações pelo mercado.

Mas a presença de racionalidade limitada em si não basta para que custos de transação se tornem importantes: racionalidade limitada não teria qualquer interesse analítico se o ambiente econômico onde se processam as transações fosse absolutamente previsível e simples. Dito de outra forma, racionalidade limitada só se torna um conceito relevante para a análise em condições de *complexidade e incerteza*.

Ambientes simples, mesmo com racionalidade limitada, não oferecem dificuldades, porque as restrições de racionalidade dos agentes não são exigidas. Já no caso de transações de elevada complexidade, a descrição de todos os desdobramentos envolvendo a transação e a especificação das decisões em cada circunstância pode se tornar extremamente custosa, impedindo os agentes de especificar antecipadamente o que deveria ser feito a cada circunstância. Outro fator a pressionar a racionalidade limitada dos agentes é a existência de incerteza. Há incerteza quando eventos futuros não podem ser previstos com exatidão. Assim, há incerteza quando se trata de um fenômeno aleatório ao qual se atribui uma dada probabilidade (incerteza no sentido convencional de risco), e também quando sequer é possível atribuir uma probabilidade ao fenômeno (por exemplo, é impossível atribuir com exatidão uma probabilidade a uma guerra futura em uma região). A presença de incerteza, *mesmo que seja no sentido convencional de risco*, combinada com racionalidade limitada, dificulta definir e distinguir as probabilidades associadas às diferentes circunstâncias que podem afetar a transação. Dito de outra forma, mesmo que seja possível, em princípio, calcular as probabilidades dos eventos, isso pode ser muito difícil, dada a racionalidade limitada dos agentes que participam da transação.

Racionalidade limitada, complexidade e incerteza têm como consequência gerarem *assimetrias de informação*. Assimetrias de informação nada mais são do que diferenças nas informações que as partes envolvidas em uma transação possuem, particularmente quando essa diferença afeta o resultado final da transação.

13.2.2.2 Oportunismo e especificidade de ativos

Racionalidade limitada, ambiente complexo e incerteza criam as condições adequadas para os agentes adotarem iniciativas oportunistas. Por oportunismo entende-se a transmissão de informação seletiva, distorcida,¹ e promessas “autodesacreditadas” (*self-disbelieved*) sobre o comportamento futuro do próprio agente (neste último caso, o agente em questão estabelece compromissos que ele mesmo sabe, *a priori*, que não irá cumprir). Como não se pode distinguir *ex-ante* a sinceridade dos agentes, há problemas na execução e renovação do contrato.

O conceito de oportunismo na TCT, portanto, possui um sentido diverso daquele que se utiliza na linguagem corrente, em que um comportamento “oportunista” é muitas vezes definido como a habilidade por parte de um agente de identificar e explorar as possibilidades de ganho oferecidas pelo ambiente. É importante ter clareza de que oportunismo, neste último sentido, *não é* oportunismo para a TCT. Oportunismo na TCT está essencialmente associado à manipulação de assimetrias de informação, visando à apropriação de fluxos de lucros.

Para se entender melhor o sentido do oportunismo na TCT, considere-se o seguinte caso: uma empresa solicita ao seu fornecedor uma mudança na especificação de um determinado insumo. Seu fornecedor informa que a mudança pretendida na especificação provocará um aumento no custo do insumo superior ao aumento que efetivamente ocorre.

Trata-se, então, de uma atitude oportunista da empresa fornecedora, uma vez que, dada a racionalidade limitada de seu cliente, assim como a complexidade na fabricação do insumo, impedem que o comprador do insumo possa conhecer as particularidades da produção daquele insumo, e, portanto, avaliar a exatidão do aumento de custo informado pelo fornecedor. A disputa em torno do fluxo de lucros que aí ocorre se dá, pois quanto maior o preço do insumo anunciado pelo fornecedor, *coeteris paribus*, menor o lucro da empresa compradora.

O outro tipo de oportunismo ocorre quando há problemas na execução de uma transação contratada. Por exemplo, quando uma empresa fornecedora de um insumo a um preço fixo reduz o nível de qualidade para reduzir seus custos. Este segundo tipo de oportunismo é conhecido na literatura por problema moral (*moral hazard*).

Contudo, mesmo racionalidade limitada, complexidade, incerteza e oportunismo não bastam ainda para gerar problemas no funcionamento dos mercados. Uma última condição se faz necessária. Esta condição é designada como a de transações que envolvem ativos específicos, isto é, ativos cujas transações ocorrem em pequeno número (*small numbers*), por haver poucos vendedores e compradores para os ativos em questão. Neste tipo de transação apenas um número limitado de agentes está habilitado a participar: a especificidade dos ativos transacionados reduz, simultaneamente, os produtores capazes de ofertá-los e os demandantes interessados em adquiri-los.

O problema associado com a especificidade de ativos é que uma vez que o investimento em um ativo específico tenha sido feito, comprador e vendedor passam a se relacionar de uma forma exclusiva ou quase exclusiva. Se um dado fornecedor é o único capaz de produzir um insumo com as particularidades desejadas por uma empresa específica, tanto o fornecedor está ligado àquela empresa, pois é a única que compra seu produto, como a empresa cliente está vinculada ao fornecedor, que é o único capaz de produzir o insumo de que necessita.

Esse vínculo entre produtor e comprador, derivado da especificidade dos ativos envolvidos na transação, pode dar origem ao que a literatura convencionou chamar “problema do refém” (*hold-up*). Esse problema ocorre quando uma das partes que realizou um investimento em um ativo específico torna-se vulnerável a ameaças da outra parte de encerrar a relação. Essa ameaça pode permitir a essas partes obter condições mais vantajosas do que as do início da transação.

O problema do refém pode se verificar tanto na relação entre o vendedor e o comprador, como vice-versa. Considere o caso de uma empresa geradora de energia elétrica que disponha apenas de uma outra empresa de transmissão para vender sua energia, ao passo que esta última compre a energia que vende de várias geradoras. Caso a empresa geradora em questão realize investimentos no aumento de sua capacidade de geração, ficará refém da empresa de transmissão para a venda da energia gerada pela sua capacidade adicional. A empresa transmissora poderá barganhar melhores preços pela energia comprada, simplesmente ameaçando não comprar a energia adicional produzida. O mesmo exemplo poderia ser pensado da forma inversa, com uma empresa geradora que atendesse a diferentes mercados, dentre eles uma empresa transmissora que comprasse exclusivamente sua energia.

A especificidade de ativos é uma condição necessária para que o risco associado a atitudes oportunistas seja significativo; caso contrário, a própria rivalidade entre os numerosos agentes aptos a participarem da transação, tanto no papel de vendedores como de compradores, reduziria a possibilidade de atuações oportunistas. No limite dessa situação de pequenos números, podemos vir a ter uma situação de monopólio bilateral, com apenas um vendedor e um comprador, caso que a teoria convencional sempre teve dificuldade em tratar.

Aqui a TCT identifica um problema interessante. Uma transação que inicialmente se caracterizava como tendo muitos agentes habilitados, isto é, como uma transação de grandes números (*large numbers*), à medida que se desenvolve no tempo pode se converter em uma transação de *small numbers*.

Isto se daria pelo que a TCT convencionou chamar “vantagens da primeira empresa a se mover” (*first-move advantages*): aqueles que vencem as ofertas iniciais terão vantagens não triviais sobre seus concorrentes potenciais, baseadas em conhecimento acumulado (*learning by doing*) sobre seus clientes. Obviamente, isto só pode ser relevante no caso de barganhas recorrentes envolvendo *ativos específicos*, e em um ambiente dinâmico em constante mudança. A esse processo, em que transações de grandes números acabam por se converter em transações de pequenos números, a TCT chamou *transformação fundamental*.

Racionalidade limitada, complexidade, incerteza, oportunismo e ativos específicos acarretando pequeno número de transações geram dificuldades significativas no momento de se contratar uma transação, isto é, adquirir um insumo ou serviço por meio do mercado. Essas dificuldades se refletem em custos de contratação. Para entender melhor essa questão, é preciso considerar a natureza dos contratos, para compreender as limitações dos mercados no que diz respeito a transações com ativos específicos.

13.3 A Natureza dos Contratos

As condições descritas anteriormente tornam bastante problemática a contratação de uma transação, especialmente se ela será concretizada em uma data futura. Racionalidade limitada, complexidade e incerteza, oportunismo e especificidade de ativos geram dificuldades no momento de negociar e redigir um contrato, assim como quando for necessário garantir sua execução mais tarde.

Contudo, há diferentes tipos de contratos, sendo preciso, então, analisar em que medida cada tipo de contrato se ajusta a uma dada configuração de custos de transação. Podemos discernir quatro tipos básicos de contratos:

1. Contratos que especificam no presente uma determinada performance no futuro;
2. Contratos que especificam no presente uma determinada performance no futuro, condicionada à ocorrência de eventos definidos antecipadamente no futuro, isto é, contratos de cláusulas condicionais;

3. Contratos de curta duração, realizados apenas nos momentos em que as condições necessárias para a realização da transação efetivamente se concretizam, isto é, contratos de curto prazo sequenciais;
4. Contratos estabelecidos hoje com o direito de selecionar no futuro uma performance específica dentro do conjunto de performances estipulado previamente, isto é, estabelecer uma *relação de autoridade*.

O contrato do tipo (1) pode ser descartado para transações que envolvam complexidade e incerteza. Na medida em que não permite nenhuma flexibilidade para ajustes a mudanças nas circunstâncias futuras, é adequado apenas para as transações mais simples, que não envolvem custos de transação significativos. Os três tipos seguintes de contrato merecem atenção mais detalhada.

13.3.1 CONTRATOS DE CLÁUSULAS CONDICIONAIS

Neste tipo de contrato as partes estabelecem um dado desempenho dependendo do que ocorra no futuro. Uma rede de lanchonetes pode estabelecer um contrato com o seu fornecedor de sorvetes em que é admitido um fornecimento extra, caso, no verão, o consumo de sorvete ultrapasse determinado volume diário.

Este tipo de contrato enfrenta duas dificuldades. A primeira delas diz respeito à dificuldade de redação à medida que a complexidade aumenta: em função da racionalidade limitada dos agentes torna-se muito difícil antecipar todas as circunstâncias futuras, assim como as providências que devem ser tomadas.

Uma segunda dificuldade diz respeito à necessidade de garantir o cumprimento do contrato. Em primeiro lugar, para garantir que o contrato está sendo executado da forma prevista é preciso determinar quais as condições, dentre as que definem as medidas a serem tomadas, estão efetivamente se realizando a cada momento. Em outras palavras, é preciso que sejam verificadas com *precisão e a baixo custo* quais as circunstâncias que estão vigorando em um dado instante. Quando as condições vigentes não podem ser identificadas com precisão e a baixo custo, abre-se espaço para que cada parte identifique como condição vigente aquela que lhe é mais conveniente, tomando atitudes oportunistas.

Em segundo lugar, ainda que as condições vigentes possam ser discernidas a baixo custo e com precisão, ainda há o problema de identificar se a parte responsável pela execução das tarefas realmente adotou as medidas adequadas de acordo com aquelas condições. Em ambientes muito complexos, isto mais uma vez pode dar margem a atitudes oportunistas.

Resulta, assim, que esse tipo de contrato é o mais indicado quando há o interesse de se preservar o vínculo entre comprador e vendedor, dada a existência de ativos com algum grau de especificidade, e em um ambiente em que a complexidade não tem consequências severas em termos de custos de negociação e garantia dos contratos. Trata-se, assim, de um contrato limitado a situações não muito complexas e incertas, e a ativos que possuem moderada especificidade.

13.3.2 CONTRATOS DE CURTO PRAZO SEQUENCIAIS

Neste tipo de contrato, vendedor e comprador não possuem um vínculo contratual duradouro. As ofertas são feitas a cada momento, em um mercado à vista (*spot*), cada comprador adquirindo o que deseja apenas no momento em que a necessidade se faz sentir. Trata-se de um tipo de contrato que reduz expressivamente qualquer problema de adaptação entre as partes envolvidas, pois não há necessidade de se antever as circunstâncias futuras que irão afetar a transação: ela somente é realizada no momento em que se faz necessária e as condições futuras já são conhecidas.

Esse gênero de contrato, portanto, também enfrenta limitações. Inicialmente, ele exige a existência de um mercado *spot*, onde os custos de acesso para vendedores e compradores sejam baixos, de tal forma que os agentes possam recorrer a este mercado sem ônus significativo. Isto exige que a transação não envolva, fundamentalmente, ativos específicos. O objeto da transação tem de ser homogêneo, de tal forma que não faça diferença tanto a identidade do comprador como a do vendedor.

Por outro lado, esse tipo de contrato também não está livre de atitudes oportunistas. Embora antes das transações se efetivarem haja realmente um caso de grandes números, após a efetivação da transação pode ser que ocorra a transformação fundamental vista anteriormente, isto é, que um processo de aprender fazendo leve o vendedor a adquirir informação privilegiada sobre o comprador, passando a ter vantagens na competição com os demais vendedores. Estabelecida assim a vantagem, o vendedor pode adotar atitudes oportunistas.

Este tipo de problema é comumente observado nas licitações do setor público: apesar de, a princípio, todos os licitantes estarem em igualdade, o vencedor da licitação acaba adquirindo informação sobre o serviço prestado, o que o coloca em vantagem diante dos demais nas licitações seguintes. O mesmo também ocorre no mercado de trabalho. Embora no momento da admissão o trabalhador frequentemente se encontre em igualdade com os demais candidatos no mercado de trabalho, o

exercício de sua função acaba conferindo conhecimentos que dão vantagem em relação a outros candidatos no mercado, caso a empresa decidisse contratar trabalhadores em contratos sequenciais de curto prazo. Por sinal, essa é a razão dessa forma de contrato não ser comum no mercado de trabalho.

Os contratos de curto prazo sequenciais são adequados, dessa maneira, em situações onde não há o interesse em preservar os vínculos entre comprador e vendedor, e onde a transformação fundamental não se verifique.

13.3.3 A RELAÇÃO DE AUTORIDADE

A característica da relação de autoridade de um agente sobre outro é que ao primeiro é facultado definir o que o segundo deverá executar, dentre um conjunto de ações possíveis. Assim, um gerente administrativo pode determinar, dentre as várias ações admitidas para um funcionário administrativo, qual é a mais adequada para determinada situação. Da mesma forma, quando o gerente de produção decide que determinado insumo tem de passar de uma etapa na fábrica para a etapa seguinte em determinadas condições, ele está escolhendo, dentre as várias ações previamente determinadas, qual o setor responsável por aquela etapa tem de adotar e, portanto, está exercendo uma relação de autoridade sobre o referido setor.

A primeira vantagem da relação de autoridade é o fato de que não é necessário antecipar todas as circunstâncias futuras, bem como as ações que devem ser executadas para cada uma delas. Em outras palavras, não é necessário gerar a *árvore de decisões* antecipadamente e, dessa maneira, não há problemas em relação à racionalidade limitada dos agentes, como no caso dos contratos de cláusulas condicionais em ambientes complexos.

A segunda vantagem diz respeito ao fato de que, sob a relação de autoridade, não é necessário recontratar sucessivamente a cada mudança na situação, o que reduz significativamente os custos de transação em circunstâncias onde existe especificidade de ativos, representando, nesse caso uma vantagem comparativamente aos contratos de curto prazo sequenciais.

Contudo, deve ficar claro que a relação de autoridade é *incompatível* com a transação no mercado, pois pressupõe que uma das partes envolvidas abra mão de sua autonomia. Em outras palavras, não há na relação de autoridade a autonomia que normalmente se pressupõe em uma oferta e uma demanda independentes. Por isso trata-se de uma relação adequada a transações com ativos de elevada especificidade, em ambientes de complexidade e incerteza, sendo o tipo de relação que se encontra no interior de uma empresa. Toda a discussão até aqui se deu em termos de especificidade de ativos como se fosse uma característica puramente qualitativa: os ativos ou são ou não são específicos. Na verdade, há graus de especificidade, como teremos oportunidade de ver. A necessidade ou não de integrar verticalmente para poupar custos de transação deve ser verificada em relação ao grau de especificidade dos ativos. Em outros termos, devem-se tipificar as transações pelo grau de especificidade dos ativos. Esse será o próximo assunto a ser abordado.

13.4 Tipos de Transações e Estruturas de Governança

Todos os problemas citados (racionalidade limitada, complexidade, incerteza, oportunismo e pequeno número de transações) ganham maior ou menor destaque de acordo com o tipo de transação em análise ou, na terminologia de Oliver Williamson, de acordo com o tipo de investimento realizado. Portanto, é preciso, de acordo com a TCT, classificar os diferentes tipos de transações. Inicialmente é necessário definir em que grau os ativos envolvidos são específicos: caso sejam ativos com mercados muito limitados, as transações com esses ativos são designadas pela TCT *transações com ativos específicos*.

Para os agentes envolvidos nesse tipo de transação, assegurar a sua continuidade é condição indispensável para estimular a decisão de investimento. Tanto para o fornecedor do ativo específico, como para o seu comprador, há o interesse em que a relação não seja interrompida, para preservar os investimentos feitos, por parte do comprador e por parte do vendedor. Obviamente, isto pressupõe que se trate de transações frequentes ao longo do tempo, ou seja, *recorrentes*. Em transações *ocasionais* tal necessidade (de continuidade da transação ao longo do tempo) não se mostra tão significativa.

O outro caso polar no que diz respeito ao caráter das transações seria o de *transações não específicas*, envolvendo equipamento ou materiais padronizados, em geral submetidos à normalização técnica. Na situação de fronteira teríamos as *transações mistas*.

Vistas as características das transações de acordo com o grau de especificidade dos ativos, o passo seguinte é examinar as diferentes formas de implementar as transações, em função dessa especificidade. Para isso, temos de conhecer as estruturas de governança.

Como afirmamos, uma situação de racionalidade limitada, complexidade, incerteza, oportunismo e pequenos números (*small numbers*) pode envolver graves problemas, de acordo com o tipo de transação envolvida. Para assegurar a realização

das transações desenvolveram-se vários tipos de estruturas de governança. Uma estrutura de governança define-se como o arcabouço institucional no qual a transação é realizada, isto é, o conjunto de instituições e tipos de agentes diretamente envolvidos na realização da transação e na garantia de sua execução. Vejamos como a TCT classifica estas estruturas e como elas se relacionam com o tipo de investimento realizado pela empresa.

1. *Governança pelo Mercado*: forma adotada em transações não específicas, especialmente eficaz no caso de transações recorrentes. Não há esforço para sustentar a relação, e na avaliação de uma transação, as partes precisam consultar apenas sua própria experiência. É o caso que mais se aproxima da noção ideal de mercado “puro”.
2. *Governança Trilateral*: aqui é exigida a especificação *ex-ante* de uma terceira parte, tanto na avaliação da execução da transação quanto para a solução de eventuais litígios. É a mais adequada em transações ocasionais, sejam elas de caráter misto ou mesmo específico.
3. *Governança Específica de Transação*: neste caso, o fato de os ativos transacionados não envolverem padronização aumenta significativamente o risco da transação e a possibilidade do surgimento de conflitos de solução custosa e incerta. Ao mesmo tempo, quanto maior o grau em que as transações forem recorrentes, maior a possibilidade de cobrir os custos derivados da constituição de um arcabouço institucional específico para a transação.² Ou seja, se a transação acontece com frequência vale a pena constituir uma estrutura de governança para lidar com ela, apesar dos custos que isto envolve. Dois tipos de estruturas podem então surgir: (a) um *contrato de relação*, onde as partes preservam sua autonomia; e (b) uma estrutura unificada e hierarquizada, isto é, uma empresa. A probabilidade da opção por uma estrutura unificada e hierarquizada cresce com o caráter idiossincrático do investimento.

Podemos caracterizar um contrato de relação, de acordo com Paul Milgrom e John Roberts, da seguinte maneira: as empresas envolvidas não se preocupam em elaborar contratos detalhados, que estipulem de forma exaustiva todos os procedimentos a serem adotados. Ao contrário, as empresas envolvidas estabelecem metas e objetivos a serem alcançados. Da mesma forma, estabelecem condições gerais de execução do contrato, especificando critérios para circunstâncias imprevistas os quais definem quem tem a autoridade para agir e limites para essas ações. Também são estabelecidos mecanismos para a resolução de conflitos, caso eles ocorram.

Como exemplo de contrato de relação teríamos a pesquisa e desenvolvimento de projetos realizados conjuntamente por empresas, cuja evolução é incerta e não pode ser prevista com antecipação.

A alternativa ao contrato de relação, quando a especificidade da transação juntamente com os demais problemas que geram custos de transação significativos (racionalidade limitada, complexidade e incerteza e oportunismo) tornam até mesmo o contrato de relação inviável, é a estrutura hierarquizada da empresa, e, nesse caso, a empresa integra verticalmente a etapa que ela transacionaria no mercado.

A Tabela 13.1 sintetiza a relação entre o tipo de investimento e a estrutura de governança.

Quanto mais “caminhamos” da esquerda para a direita e de cima para baixo, maior é a tendência no sentido de que as transações por meio do mercado sejam substituídas por transações em que estruturas de governança mais específicas – como a governança trilateral, em que há uma terceira parte regulando ou arbitrando a transação, ou as partes estão vinculadas por um contrato de relação – sejam constituídas. Isto prossegue com o aumento da especificidade dos ativos e da recorrência da transação até o limite em que a transação passa a ser realizada no interior da empresa, que cria uma estrutura de governança específica para isto (seja uma divisão, departamento ou filial), isto é, que ocorra o que se convencionou chamar de um processo de *verticalização*.

Em termos gerais, o que está acontecendo é que o caráter crescentemente específico das transações – com a consequente redução no número de vendedores e compradores do ativo – reduz progressivamente as vantagens que o mercado oferece, em especial a vantagem de se obter o ativo a um preço mais baixo, pois o mercado precisa agregar um grande volume de demandas, para realizar, assim, economias de escala. Ao mesmo tempo os custos derivados de negociar, redigir, implementar e verificar a execução adequada das cláusulas contratuais crescem, e com isto os custos de transação. É da comparação destes dois termos

TABELA 13.1 Característica das Transações

Frequência das Transações	Não Específicas	Mistas	Específicas
Ocasionais	Governança de Mercado	Governança Trilateral	Governança Trilateral
Recorrentes	Governança de Mercado	Governança Bilateral (Contrato de Relação)	Organização Interna (Empresa)

Fonte: Williamson (1986, p.117).

(economias de escala *versus* custos de transação) que caberá a decisão final quanto à forma institucional mais adequada para a organização da transação: via mercado ou via estrutura hierárquica da empresa.

Este ponto é importante e deve ser ressaltado. A condição para a obtenção de economias de escala é que o mercado seja suficientemente extenso para que os ganhos de escala possam ser aproveitados. Isto nada mais é do que a existência de um grande número de transações, o que, por sua vez, significa que o ativo transacionado possui um grau de especificidade bastante baixo. Se o ativo não é específico, o mesmo ativo pode ser oferecido a um número razoavelmente grande de clientes potenciais e, assim, o produtor ganha escala, e com isso pode oferecer o ativo em questão a um preço mais baixo do que o custo que o demandante teria se produzisse o ativo apenas para si mesmo.

Ao mesmo tempo, por ser um ativo de baixo grau de especificidade, a soma dos custos de negociar, redigir e garantir a execução de um contrato deixa de ser expressiva. Sendo o ativo pouco específico, não envolve complexidade e incerteza. Não envolvendo complexidade e incerteza, a negociação e a redação dos contratos não pressionam a racionalidade limitada dos agentes. Adicionalmente, sendo o ativo pouco específico, temos transações em grande número, o que significa muitos vendedores e compradores no mercado, limitando, assim, as vantagens de qualquer uma das partes em adotar atitudes oportunistas: sempre se pode encontrar um outro fornecedor ou um outro cliente para transacionar o ativo.

Dessa forma, o empresário compara dois custos: o custo de produzir ele mesmo o insumo e com isso integrar verticalmente, e o custo de adquirir o insumo no mercado. No custo de produzir ele mesmo o insumo, o empresário considera que a perda de escala (pois ele estará produzindo apenas para si mesmo) vai tornar o insumo mais caro do que se fosse adquirido do mercado, onde o insumo é produzido não apenas para o empresário em questão, mas para todos os demais que desejam adquiri-lo. Por outro lado, o empresário tem de considerar, caso decida não produzir ele mesmo o insumo, os custos de transação de obter o insumo no mercado.

Se o insumo é pouco específico e as transações no mercado são de grande número, isto significa que provavelmente economias de escala significativas são realizadas pelos produtores desse insumo no mercado, e o empresário terá de considerar um aumento de custo significativo caso ele mesmo decida produzir o insumo, pois não terá a mesma escala dos produtores no mercado. Por outro lado, como o insumo é pouco específico, os custos de transação não devem ser expressivos. Como sairá muito mais caro (por perda de escala) produzir o insumo do que adquiri-lo no mercado (já que os custos de transação são baixos), o empresário deverá optar por comprar o insumo com baixo grau de especificidade no mercado, não integrando verticalmente sua produção.

O mesmo já não ocorre caso a transação sob análise seja do gênero *small numbers*. Neste caso, o ativo é produzido tendo em vista as necessidades específicas do demandante, logo, há pouca ou nenhuma possibilidade do ativo ser produzido para outros demandantes e com isso as vantagens em termos de escala de produção (com a consequente redução de custos) se perdem, ao mesmo tempo em que os problemas de negociação e implementação do contrato aumentam. Pode passar a ser mais barato, então, para o demandante produzir ele mesmo o ativo em questão do que recorrer a uma transação via mercado em termos de custo final. Isto porque, produzindo ele mesmo, não há perda de escala significativa, uma vez que sendo o insumo específico seu mercado é muito limitado e ele economiza em custos de transação.

Obviamente, este último tipo de dificuldade só tende a piorar na medida em que a frequência da transação aumente: os custos de negociar, implementar e verificar os contratos aumentam na mesma razão da continuidade das transações, desde que elas sejam específicas.

Deve estar clara a importância da especificidade dos ativos para a determinação da relevância dos custos de transação em um caso específico. Mas quais as causas de especificidade de ativos? São identificadas quatro fontes de especificidade dos ativos:³

1. Especificidade de localização: decisões prévias, visando minimizar custos de estocagem e transporte, podem gerar ativos com especificidade de localização, que uma vez estabelecidos o transporte pode ser difícil ou impraticável. Uma subestação de distribuição de energia elétrica ou um gasoduto seriam exemplos.
2. Especificidade física: características de *design* podem reduzir o valor do ativo em uma aplicação alternativa. Equipamentos sob encomenda se enquadram nesse caso.
3. Especificidade de capital humano: este tipo de especificidade surge fundamentalmente por meio de processos de “aprender fazendo” (*learning-by-doing*) dos empregados de uma empresa. Isto é especialmente verdadeiro para a mão de obra alocada nos laboratórios de pesquisa e desenvolvimento das empresas.
4. Especificidade de ativos dedicados: surge nos casos em que o fornecedor faz um investimento que, exceto pela promessa da venda de uma quantidade expressiva de produto para um determinado cliente, não seria feito. Como exemplo, temos os investimentos de fornecedores de autopartes para atender a uma montadora.

Esses quatro itens listam as principais fontes de especificidades de ativos. Resta agora verificar se as especificidades acima relacionadas, juntamente com assimetria de informação, realmente determinam o grau de verticalização das indústrias. Para isso temos de considerar os trabalhos que examinaram a evidência empírica.

13.5 Evidência Empírica sobre Custos de Transação

O primeiro esforço sistemático de examinar empiricamente a TCT foi publicado em 1982, por Kirk Monteverde e David Teece.⁴ Em seu trabalho, aqueles autores examinaram a escolha entre integrar verticalmente ou recorrer ao mercado das empresas Ford e General Motors nos Estados Unidos. Utilizou-se uma lista de 133 componentes automotivos, para os quais se observava se eram produzidos internamente pelas empresas ou comprados de fornecedores. A hipótese era de que o fator determinante na escolha era o esforço de engenharia aplicado a cada um desses componentes, o que identificaria uma especificidade de ativos humanos. Dessa forma, quanto maior o esforço de engenharia em um dado componente, maiores as chances desse componente ser produzido internamente.

Também foram distinguidos componentes que eram específicos de cada companhia, em relação aos que eram comuns ao mercado automobilístico. A hipótese era de que componentes específicos de uma das empresas teriam especificidades físicas, derivadas de um *design* particular, o que também aumentaria a probabilidade desses componentes específicos serem produzidos dentro da empresa, em vez de adquiridos de fornecedores no mercado.

A hipótese nula, contra a qual a hipótese de que custos de transação gerados pela especificidade dos ativos acarretaria integração vertical era testada, consistia simplesmente na hipótese de que a especificidade dos ativos não determinaria a integração vertical da empresa na produção daqueles itens. É importante notar que, além da especificidade dos ativos, tratava-se de transações recorrentes, pois as empresas precisavam adquirir repetidamente os componentes para montar os veículos.

Os resultados obtidos por Monteverde e Teece foram consistentes com as hipóteses da TCT, permitindo rejeitar a hipótese nula. Isso deu a primeira sustentação empírica para toda a discussão acerca dos custos de transação determinando as etapas da cadeia produtiva integradas verticalmente dentro da empresa.

Scott Masten publicou, em 1984, um estudo empírico referente às compras de uma empresa aeroespacial, acerca dos componentes que ela tinha se comprometido a fornecer para o governo americano. Um questionário foi fornecido ao pessoal de compras da empresa para que os componentes fossem classificados de acordo com os seus diferentes graus de especificidade e complexidade. Masten encontrou evidências de que aumentam as chances dos componentes serem produzidos internamente quanto maiores os seus graus de complexidade e especificidade.

Outros estudos foram feitos, envolvendo, por exemplo, a decisão da empresa entre雇用其 ownários vendedores ou firmar contratos com representantes de vendas autônomos. Em todos eles verificou-se evidência empírica de que a verticalização tende a ocorrer quanto maior a especificidade e complexidade da transação.

13.6 Aplicações da Teoria dos Custos de Transação

As duas principais aplicações da TCT têm sido, afora análises de Economia Industrial acerca de estruturas verticais, nas áreas de defesa da concorrência e regulação econômica. Essas duas aplicações serão discutidas brevemente a seguir.

13.6.1 DEFESA DA CONCORRÊNCIA

A TCT foi responsável, notadamente a partir dos trabalhos de Oliver Williamson, por uma revisão significativa dos conceitos que norteavam a política norte-americana de defesa da concorrência no final dos anos 1960. Com efeito, naquele momento qualquer movimento de integração vertical era avaliado negativamente do ponto de vista da defesa da concorrência: considerava-se que o único motivo por trás da decisão de uma empresa de se integrar verticalmente seria subtrair um fornecedor de seus concorrentes, dificultando a competição no setor, e com isso aumentando indiretamente seu poder de mercado.

Ocorre, porém, que, muitas vezes, um movimento de uma empresa no sentido da integração vertical com um fornecedor pode ter como objetivo economizar em custos de transação, em função de elevado grau de especificidade e complexidade das transações. Nesse caso, o processo de integração vertical representa um aumento na eficiência da economia, e não necessariamente uma tentativa de solapar a competição subtraindo um fornecedor, até mesmo porque, se for elevada a especificidade dos ativos envolvidos, o fornecedor em questão provavelmente não encontraria outros compradores para sua produção.

Outro exemplo de aplicação da TCT à defesa da concorrência diz respeito às chamadas restrições verticais. Um tipo bastante comum de restrição vertical é a franquia, em que o agente franqueado se compromete a adquirir seus produtos apenas de um dado fornecedor, ostentando a marca desse.

Suponha agora que uma empresa tivesse um produto de qualidade superior, pelo qual os consumidores estariam dispostos a pagar um preço maior, mas que para isso o produtor tivesse de cumprir com três condições: (1) tornar reconhecidos os atributos

superiores de seu produto, (2) manter controle de qualidade sobre esses atributos e (3) manter um controle de custos para evitar que o preço do produto se torne proibitivo.

Em uma situação em que os consumidores tivessem racionalidade ilimitada, isso é, fossem capazes de imediatamente identificar que estava sendo oferecido um produto superior, assim como verificar a qualidade efetiva do produto, não haveria qualquer problema. Bastaria o produtor oferecer seu produto no mercado através de qualquer revendedor, os consumidores perceberiam as diferenças e pagariam o preço adequado pelo produto de melhor qualidade.

Todavia foi visto que a suposição de que os consumidores têm perfeito conhecimento de tudo o que é necessário não é uma hipótese adequada para enfrentar a realidade do mercado. A hipótese mais adequada nesse sentido é a de *racionalidade limitada*. Assim, nosso hipotético produtor enfrenta não apenas a dificuldade de informar que seu produto tem qualidade superior, como também o problema de fazer com que os consumidores acreditem nessa informação, pelas restrições em função da racionalidade limitada dos indivíduos.

O nosso produtor terá de, simultaneamente, atrair a atenção dos consumidores para os atributos de seu produto, garantir que esses atributos não sejam afetados até a aquisição pelo consumidor e, finalmente, monitorar todo o processo, de uma forma economicamente conveniente.

A forma economicamente adequada é a franquia: na medida em que o produtor obriga um revendedor a comprar insumos com exclusividade dele, e a revender o produto nas condições que especifica (que podem ir desde a forma de estocagem e manuseio até ao *layout* do local de revenda e visualização da marca do produtor), garante que um produto de qualidade superior chegue ao consumidor sem descharacterizar seus atributos, ao mesmo tempo em que reduz a possibilidade de atitudes oportunistas por parte do revendedor, tais como exibir a marca do produtor que oferece qualidade superior, mas vender um produto inferior e mais barato, aumentando sua margem de lucro.

13.6.2 REGULAÇÃO ECONÔMICA

Em termos de regulação econômica, a contribuição mais importante da TCT se deu no sentido de reavaliar as vantagens de concessões de serviços públicos em relação à regulação direta de uma empresa por um órgão governamental.

Serviços públicos se enquadram frequentemente na categoria de “monopólio natural”. Dito de forma muito resumida, um setor é considerado monopólio natural quando o custo de produzir com uma única empresa é menor do que produzir com mais do que uma empresa. Isto ocorre em distribuição de energia elétrica, por exemplo: é mais barato ter uma única empresa distribuindo energia elétrica do que duas empresas, com suas linhas uma ao lado da outra.

A recomendação corrente da teoria é que, nesses casos, o governo regulamente a empresa monopolista para estabelecer seu lucro no valor mínimo para cobrir o custo de oportunidade do investimento. Alguns autores, contudo, começaram a argumentar que a regulação governamental seria desnecessária, mesmo sendo a empresa monopolista. O argumento é que, para tanto, bastaria que a concessão fosse renovada periodicamente, estando a empresa vencedora sujeita a concorrer a cada renovação com outras empresas, por meio de leilão.

A TCT forneceu os instrumentos para demonstrar a precariedade desse argumento. Seria necessário, inicialmente, definir o prazo de duração da concessão. Se fosse por um prazo curto, isto é, se a concessão fosse renovada por intermédio de contratos de curto prazo sequenciais, o concessionário não teria o estímulo para fazer os investimentos de longo prazo de maturação. Isto porque haveria muitos problemas no momento de se transferir essa concessão de uma empresa para outra: quais os valores que a nova concessionária iria pagar pelos ativos da concessionária anterior? Pelo seu valor histórico, o que em caso de inflação impediria a empresa que perdeu a concessão de recuperar o investimento feito, ou pelo seu valor de mercado, obrigando a nova concessionária a pagar preços de ativos novos por ativos já utilizados?

Se o contrato da concessão for de longo prazo, outro tipo de custo de transação surge. Mesmo que o prazo da concessão seja suficiente para o concessionário amortizar todos os investimentos, agora surgem os problemas característicos de contratos de longo prazo, anteriormente vistos: são necessárias cláusulas condicionais que antecipem todas as circunstâncias relevantes no futuro. Isso em ambiente de complexidade e incerteza pode dar margem a atitudes oportunistas por parte tanto da empresa regulada como do regulador.

A TCT sugere, então, que concessões não devem ser pensadas como algo a substituir a regulação, uma vez que, embora a regulação tenha problemas, a concessão renovada periodicamente não seja isenta de dificuldades de natureza transacional. É necessário, assim, que, mesmo com os leilões de concessão, o agente regulador continue monitorando e, caso necessário, controlando preços, custos e investimento da empresa concessionária.

Notas

1. Transmissão seletiva de informações consiste em transmitir apenas as informações que favorecem a parte na transação, omitindo aquelas que reduziram os seus ganhos. Transmissão de informações distorcidas consiste em falsear informações para com isso obter ganhos que de outra forma não seriam possíveis.
2. É importante notar que a criação de um arcabouço institucional gera custos fixos. Por exemplo, uma divisão de uma empresa responsável pela produção de um insumo utilizado pela própria empresa representa um custo fixo, pois a despesa com folha de pagamento, material de consumo etc., será a mesma independentemente do volume de insumo produzido. Dessa forma, a criação de um órgão desse tipo só se justifica quando as transações em que sua participação é necessária atingem determinado volume mínimo.
3. É importante não esquecer que a especificidade dos ativos não é a única condição para a integração vertical. Assim, caso o equipamento sob encomenda seja adquirido em transações ocasionais, como foi visto anteriormente, isto não deverá levar à integração vertical.
4. Para uma resenha dos vários testes empíricos sobre a TCT ver Joskow (1991).

Bibliografia

- Coase, R. H. (1937) "The Nature of the Firm". *Economica*, s/n, p. 386-405.
- Joskow, P. L. (1991) "Asset Specificity and Structure of Vertical Relationships: Empirical Evidence", in Williamson, O.; e Winter, S. G.(orgs.): *The Nature of the Firm: Origins, Evolution, and Development*. Oxford: Oxford University Press.
- Milgrom, P.; e Roberts, J. (1992) *Economics, Organization and Management*. New Jersey: Prentice Hall.
- Williamson, O. E. (1975) *Markets and Hierarchies: Analysis and Antitrust Implications*. Nova York: The Free Press.
- Williamson, O. E. (1985) *The Economic Institutions of Capitalism; Firms, Markets, Relational Contracting*. Nova York: The Free Press.
- Williamson, O. E. (1986) *Economic Organization: Firms, Markets and Policy Control*. Nova York: N.Y. University Press.
- Williamson, O. E. (1987) *Antitrust Economics*. Oxford: Basil Blackwell.
- Williamson, O. E. (1996) *The Mechanisms of Governance*. Oxford: Oxford University Press.

Organização das Grandes Corporações

João Luiz Pondé

14.1 Introdução

A partir dos anos 1970, diferentes áreas da ciência econômica foram afetadas pela proliferação de trabalhos, teóricos e empíricos, orientados por enfoques que podem ser caracterizados como institucionalistas. Partindo da convicção de que “as instituições importam”, várias contribuições importantes têm procurado analisar como as características internas desta instituição fundamental para a sociedade em que vivemos – a empresa capitalista – afetam os processos econômicos; além disso, a aplicação da teoria econômica para explicar a gênese e o desenvolvimento de uma ampla variedade de instituições também se estendeu para a investigação dos determinantes dos tipos de empresas e formas de organização predominantes, quase sempre procurando identificar a lógica de eficiência a estas subjacente.

O estudo dos determinantes e efeitos econômicos da organização interna da empresa não havia se mostrado, até então, um tema de pesquisas atrativo para os economistas, mesmo para aqueles dedicados à área de Economia Industrial. Certamente, a desconsideração deste conjunto de fenômenos como objeto de estudo relevante não refletiu a sua real importância para a operação do sistema econômico, dado que grande parte das atividades produtivas é dirigida por decisões tomadas no contexto de organizações empresariais, dentro das quais os sinais emitidos pelo mercado atuam apenas após a mediação de uma diversidade de procedimentos gerenciais.

Em parte, as causas para esta omissão estavam na crença disseminada entre seus membros de que o desafio maior da ciência econômica estava em compreender os mecanismos relacionados à interação dos agentes nos mercados, em especial no que se refere à aparente capacidade das economias capitalistas de gerar resultados agregados com um relativo grau de ordem, ainda que cada indivíduo atue movido apenas por seus interesses particulares e egoístas. Em suas tentativas de desenvolver teorias do funcionamento dos mercados, os economistas acabaram dispensando um tratamento, no mínimo, superficial à empresa, frequentemente reduzindo-a a uma unidade decisória moldada à maneira de um indivíduo capaz de maximizar uma função objetivo bem definida ou a um conjunto de relações técnicas representadas nas funções de produção. Mesmo quando a empresa em si foi objeto de algum esforço de teorização, como no caso de modelos de crescimento da empresa que são apresentados no Capítulo 2, esta foi tomada como uma entidade individual e sua estrutura interna, que determina os padrões de interação entre seus membros, foi deixada em segundo plano.

Neste capítulo, discutiremos três abordagens econômicas da organização interna da grande empresa. A primeira delas é composta de uma diversidade de modelos teóricos que se caracterizam por estender o alcance do instrumental teórico da microeconomia tradicional para a análise das formas específicas de organização da empresa, partindo de uma redefinição conceitual desta última, que passa a ser considerada como uma rede de contratos. A segunda abordagem é aquela oferecida pela teoria dos custos de transação na versão proposta por Oliver Williamson, cujos principais conceitos já foram apresentados no Capítulo 13.

Como veremos mais a seguir, esta teoria não apenas procura explicar os movimentos de integração vertical, como também se aplica a processos de mudança da organização interna das empresas ao longo do tempo. Por último, será apresentada uma abordagem que enfatiza o papel da estrutura organizacional da empresa em mobilizar os conhecimentos necessários ao aprendizado tecnológico. Trata-se de um conjunto de contribuições que resulta, principalmente, de trabalhos do programa de pesquisa neo-schumpeteriano, embora influência de ideias geradas de fora da Economia Industrial – basicamente de teorias das organizações de natureza mais sociológica – também se faça presente.

14.2 A Empresa como um Nexo de Contratos

Esta abordagem da organização da empresa tem como marco o artigo publicado por Arman Alchian e Harold Demsetz, em 1972, ao qual se seguiu uma vasta literatura que tem como característica central o uso da teoria microeconômica neoclássica para explicar propriedades e características de uma ampla variedade de formas organizacionais e relações contratuais. As principais inovações teóricas que permitiram este movimento foram uma redefinição conceitual da empresa em termos contratuais, o relaxamento da hipótese de perfeita informação e a admissão do fato de que os agentes estão propensos ao oportunismo pós-contratual. (Ver o Capítulo 13 para essa discussão.)

14.2.1 UM NOVO CONCEITO DE EMPRESA

Em vez de conceberem a empresa como uma unidade decisória ou uma unidade produtiva cujas fronteiras e principais características organizacionais são determinadas pela tecnologia em uso, os autores desta abordagem vão defini-la como uma ficção legal que serve como um “nexo” para um conjunto de relações contratuais entre os indivíduos. Tais relações ainda se caracterizam pela existência de direitos sobre os ativos e receitas da organização, que são de propriedade de alguns indivíduos e podem ser vendidos a terceiros sem a permissão das demais partes dos contratos firmados.¹ Em outras palavras, a empresa nada mais é do que uma rede de contratos entre os proprietários dos recursos produtivos utilizados nos seus processos produtivos; a entidade jurídica correspondente a esta consiste apenas em um artifício criado para centralizar as relações contratuais em torno de uma parte contratante, em vez de organizá-la em um agregado de relações bilaterais.

Como visto no Capítulo 13, em um mundo onde não existissem custos de coletar informações, a configuração específica destes contratos seria irrelevante para a eficiência do sistema econômico, dado que os agentes seriam capazes de elaborar contratos aptos a garantir, sem qualquer custo adicional, a efetivação das transações desejadas. Contudo, se a informação não é um bem livre, os contratos jamais poderão incluir termos para solucionar conflitos entre as partes e empreender ajustes na distribuição de custos e benefícios para todas as eventualidades previstas. Além disso, caso uma das partes suspeite do não cumprimento de algum compromisso pela outra parte do contrato, a comprovação de que isto está efetivamente ocorrendo e o uso das salvaguardas legais para corrigir o problema exige o comprometimento de recursos adicionais para este fim.

Abre-se, então, a possibilidade da emergência de comportamentos que podem ser caracterizados como uma manifestação do que a literatura denomina de “risco moral” (*moral hazard*) – uma conduta oportunista pela qual uma das partes de um contrato muda sua conduta após este ter sido pactuado, se aproveitando do fato de que só ele tem acesso a algumas informações para obter ganhos em detrimento de um ou mais agentes com os quais a relação contratual foi estabelecida.² Um exemplo de uma possível manifestação do risco moral na empresa está em um empregado que cumpre sua função – estabelecida contratualmente – de forma desleixada e reduz sua produtividade, sabendo que o contratante não poderá facilmente detectar sua conduta.

A empresa é caracterizada, então, como um conjunto articulado de contratos, que especifica os direitos de propriedade vigentes para as condutas e interações dos agentes que desta participam. Os direitos de propriedade – entendidos como regras socialmente definidas quanto aos usos que os agentes podem dar aos recursos econômicos – são, em parte, definidos pela legislação e pelo sistema legal em vigor, mas têm seu conteúdo estendido, alterado e/ou especificado pelas relações contratuais livre e voluntariamente estabelecidas pelos indivíduos e grupos de indivíduos. Assim, a legislação vigente de direitos de propriedade e os contratos em uso resultam em uma estrutura de oportunidades e restrições, bem como recompensas e penalidades, cujo efeito é fixar uma matriz de custos e benefícios associados às várias ações possíveis. Em outras palavras, os direitos de propriedade determinam uma *estrutura de incentivos* para as decisões e ações. Definida no âmbito do nexo de contratos de cada empresa, esta estrutura de incentivos condiciona os comportamentos dos agentes e delimita a extensão em que os compromissos contratuais firmados serão vulneráveis a comportamentos oportunistas.

A definição de empresa como um nexo de contratos tem duas implicações importantes:

1. A distinção entre as interações que se processam entre os agentes dentro do mercado e aquelas que se processam dentro da empresa é diluída, já que, em ambos os casos, teríamos apenas relações contratuais livremente pactuadas entre agentes econômicos. O contrato de trabalho entre um trabalhador e seu empregador e o contrato de fornecimento de insumos entre empresas independentes são vistos como diferindo apenas em grau. Por isso, vários autores desta abordagem vão repudiar a concepção de que as organizações empresariais se caracterizam por possuírem relações entre os agentes baseadas em mecanismos de autoridade que configuram estruturas hierárquicas. Ao contrário, vão defender que a própria noção de uma fronteira entre o que está “dentro” e o que está “fora” da empresa deve ser abandonada – o que existiria seria apenas uma variedade de relações contratuais, cujas características permitem analisá-las como um conjunto contínuo. As distinções de classes que permitam delimitar precisamente o que faz parte e o que não faz parte da empresa não são consideradas relevantes;
2. Na medida em que as relações entre os agentes no nexo de contratos que define a empresa são da mesma natureza que aquelas verificadas no mercado, abre-se a possibilidade, para os teóricos neoclássicos, de modelar a organização da empresa da mesma maneira como são modeladas as relações entre os agentes econômicos em um mercado, como um resultado de equilíbrio para uma complexa rede de interações entre indivíduos autônomos. Assim, a abordagem do nexo de contratos desenvolve diferentes modelos que, a partir de condições iniciais determinadas, permitem a identificação de resultados de equilíbrio e uma avaliação de suas propriedades de eficiência.

Portanto, explicar a organização da empresa é, nesta abordagem, explicar o equilíbrio particular observado em um nexo de contratos. O pressuposto na definição do equilíbrio é que os agentes renegociarão os termos dos contratos sempre que, com isso, sejam auferidos ganhos líquidos, ou seja, pelo menos um agente possa auferir um ganho sem gerar perdas para os demais. Dessa forma, a organização de equilíbrio resultante possui a propriedade de ser uma solução ótima para o problema de coordenar as interações dos agentes dentro da empresa.

14.2.1.1 Custos de agência e as empresas de capital aberto

Em uma importante vertente desta abordagem, a eficiência dos contratos é analisada a partir dos conceitos de *relação de agência* e *custo de agência*. Uma relação de agência ocorre sempre que exista um contrato, formal ou informal, pelo qual um indivíduo ou grupo de indivíduos – o *principal* – contrata um ou mais indivíduos – o(s) *agente(s)* – para desempenhar alguma atividade de seu interesse, delegando aos contratados algum poder de decidir de que maneira a atividade será executada. O problema básico de qualquer relação de agência é que, se o comportamento dos indivíduos envolvidos é pautado pela busca do interesse próprio, o principal poderá encontrar dificuldades em induzir o agente a se comportar de maneira que maximize o ganho do principal – o agente pode preferir executar as atividades para as quais foi contratado de uma forma que incremente o seu ganho em detrimento do ganho do principal.

Assim, o nexo de contratos que define uma empresa pode ser analisado como um conjunto de relações de agência, cuja eficiência, por sua vez, pode ser analisada a partir do conceito de *custo de agência*. Tais custos envolvem as perdas sofridas pelo principal devido a uma conduta oportunista dos agentes e o dispêndio de recursos em evitar que isso ocorra, o que envolve:

1. O custo residual para o principal gerado pela relação de agência, ou seja, o valor das perdas impostas ao principal devido aos agentes tomarem atitudes divergentes daquelas que maximizam ganho do principal;
2. As despesas de monitoramento incorridas para verificar se os agentes estão atuando da maneira desejada pelo principal;
3. Os dispêndios realizados pelos agentes para assegurar seu comprometimento com os interesses do principal (*bonding costs*).³

Em outras palavras, os custos de agência são gastos de recursos econômicos decorrentes da existência de dificuldades em fazer com que os agentes tomem decisões ótimas do ponto de vista do bem-estar do principal. Dado que a coleta de informações sempre acarreta um gasto de recursos, os custos de agência serão um dado da realidade econômica, da mesma maneira que os custos de transporte das mercadorias.

Uma importante aplicação da noção de custos de agência está nos modelos teóricos que investigam a organização das empresas de capital aberto, mais especificamente as estruturas organizacionais que emergem da necessidade de garantir que os interesses dos proprietários das ações sejam levados em conta nas decisões tomadas pelos executivos da empresa. Desde os anos 1930, após a publicação do estudo de Bearle e Means, *The Modern Corporation and Private Property*, o reconhecimento de que a gerência profissional de uma empresa de capital aberto era capaz de tomar decisões em seu benefício e em detrimento dos interesses dos proprietários foi frequentemente considerado um forte indício de que a teoria da firma exposta nos livros-texto de microeconomia não era aplicável a este tipo de organização. Na medida em que os executivos profissionais tomavam decisões

de acordos com sua própria função objetivo, argumentava-se, o comportamento da empresa não mais refletiria a maximização de lucros – ou, mais precisamente, a maximização de seu valor presente. Por outro lado, isso implicaria que a alocação de recursos resultante não seria ótima, ou que, em outras palavras, os custos incorridos em gerar uma determinada quantidade de produto seriam maiores do que os verificados no caso de uma empresa de capital fechado.

A abordagem do nexo de contratos proporciona uma reinterpretação da natureza e da avaliação de eficiência das empresas de capital aberto, que pode ser resumida nas seguintes proposições:

1. A hipótese de comportamento maximizador não deve ser aplicada à empresa como organização, mas sim a cada um de seus membros. Assim, mesmo uma empresa dirigida pelo seu proprietário não apresentará um comportamento que maximiza seus lucros ou seu valor presente. O proprietário tomará as decisões relativas ao destino da empresa para maximizar seu bem-estar e este resulta de benefícios oriundos tanto de ganhos pecuniários na forma de rendas monetárias quanto de ganhos não pecuniários – estes últimos incluindo o ambiente de trabalho (escritórios, secretárias), relações pessoais com empregados, comprar insumos de amigos, ter um computador de último tipo para brincar com ele etc. Um equilíbrio entre estes dois tipos de ganhos ocorre quando a utilidade marginal de uma unidade monetária gasta em cada item não pecuniário é igual à utilidade marginal do aumento da riqueza do proprietário em uma unidade monetária;
2. Já se a empresa é administrada por executivos assalariados, o que antes era um problema de decisão entre ganhos pecuniários e não pecuniários por parte do proprietário-gerente, agora se torna uma relação de agência. A divergência entre os interesses entre os acionistas e os executivos resulta do fato destes últimos poderem auferir ganhos não pecuniários sem “pagar” por isso, já que a redução dos lucros da empresa não afeta seu bem-estar na mesma magnitude;
3. Por sua vez, os custos de agência inerentes às empresas de capital aberto podem ser minimizados por uma reelaboração dos contratos entre executivos e os acionistas, que corresponde à criação do aparato de monitoramento e controle de decisões que caracterizam a moderna sociedade por ações – envolvendo a formação de conselhos de administração, procedimentos de coleta e processamento de informações, especificação de mecanismos de prestação de contas aos acionistas etc.

Dessa maneira, a organização da empresa de capital aberto é analisada como uma solução contratual ótima para uma relação de agência não eliminável em um mundo no qual a informação é custosa. Sua implementação permite que os ganhos oriundos da abertura do capital das empresas e da constituição de mercados de capitais – em termos de uma gestão de portfólio mais eficiente dos proprietários de ativos – sejam auferidos, ao mesmo tempo em que os custos de agência são minimizados. Por conseguinte, a existência destes custos de agência não constitui, em si, uma evidência de que a empresa de capital aberto seja uma organização que opere em condições subótimas quando uma solução contratual de equilíbrio é obtida.

14.3 Custos de Transação e Mudança Organizacional

O tratamento oferecido por Oliver Williamson, dos determinantes da organização interna da empresa, é uma extensão natural da sua explicação de processos de integração vertical vista no Capítulo 13. Se a internalização de um conjunto de atividades oferece um aparato administrativo que proporciona economias de custos de transação, também acarreta novos custos inexistentes nos mercados. Do mesmo modo como a existência de custos de transação relativamente mais elevados em transações realizadas no mercado determina a substituição desta estrutura de governança pela organização hierárquica da empresa, a possibilidade de auferir economias de custos de transação pode resultar em alterações nas características desta organização, desencadeando um processo de *mudança organizacional*.

O argumento central aqui é que a organização das transações dentro de uma empresa não elimina o oportunismo, mas procura controlá-lo por meio de uma gama de incentivos que o mercado não possui e por meio de mecanismos mais desenvolvidos para o monitoramento e o controle das ações – a hierarquia. Ao suprimir as relações mercantis e o processo competitivo a estas associadas, as hierarquias geram uma série de distorções burocráticas, entre as quais devem ser destacadas:

1. A propensão de diferentes estratos gerenciais a utilizar os recursos das empresas para perseguir metas próprias,⁴ o que pode ser agravado pelo desenvolvimento de padrões de reciprocidade e manutenção do *status* dos gerentes, em um esquema do “viva e deixe viver” que distorce as tomadas de decisões. Muitas vezes, decisões internas operacionais e de investimento estão sujeitas à politização, o que coloca o desempenho global da empresa em segundo plano diante dos conflitos de interesses dos diferentes membros da organização;

2. A ocorrência do que Williamson denomina de “comportamento persistente”, derivado da inércia interna que dificulta rápidas mudanças. Os pactos que naturalmente emergem da citada politização da organização podem torná-la pouco flexível, na medida em que reestruturações mais profundas na sua forma de atuar podem ameaçar a posição de subgrupos específicos, o que resulta em processos de seleção mais lenientes do que os encontrados nos mercados – as redes de compromissos tornam as penalidades sujeitas à negociação.⁵

A incorporação teórica dos custos relacionados à burocracia no funcionamento das empresas significa que a organização interna das transações não elimina as dificuldades contratuais, apenas as situa em um novo patamar, retirando a intermediação do mercado. De fato, a sua internalização em estruturas hierárquicas requer a criação de incentivos alternativos aos oferecidos pelo mercado – na função de sancionar, penalizar ou premiar a conduta dos agentes – e mecanismos de controle para coordenar a interação entre as distintas atividades envolvidas.⁶ Esta abordagem da organização empresarial, enfatizando a relevância das capacitações necessárias para a construção das estruturas de governança apropriadas, permite conceber a empresa como o produto de uma série de inovações organizacionais que têm a intenção e o efeito de economizar em custos de transação.

Um exemplo de como a teoria dos custos de transação explica as mudanças na organização interna das empresas está na explicação de Williamson para o surgimento e proliferação das empresas multidivisionais a partir dos anos 1930. O Quadro 14.1 detalha esse exemplo. Neste contexto, deve-se notar que, não obstante estas constituam uma inovação organizacional geradora de ganhos de eficiência, a teoria de Williamson não concebe qualquer estrutura organizacional poupadora de custos de transação como constituindo, necessariamente, uma forma institucional “ótima”, ou ainda uma decorrência direta das necessidades de gerenciar as atividades produtivas dentro de uma base técnica e evolução da estrutura industrial específicas. Trata-se, isto sim, de uma “solução” historicamente datada, cuja análise a partir da teoria dos custos de transação se dá no sentido de identificar a lógica subjacente à sua criação e difusão.⁷

QUADRO 14.1 CUSTOS DE TRANSAÇÃO E A PROLIFERAÇÃO DAS EMPRESAS MULTIDIVISIONAIS

Até os anos 1930, as grandes empresas eram organizadas em uma estrutura funcional, na qual as atividades eram reunidas em departamentos delimitados por funções específicas. Já na empresa multidivisional, as atividades são agrupadas em divisões que reúnem todas ou quase todas as funções necessárias para atuar em um determinado mercado ou linha de produtos; cada divisão passa a se subordinar a um escritório central dedicado à formulação da estratégia de médio e longo prazos da empresa.

Do ponto de vista da teoria dos custos de transação, a empresa multidivisional é uma inovação organizacional, o produto de tentativas de reproduzir comportamentos competitivos mercantis entre as divisões e, simultaneamente, aumentar a eficácia do controle interno sobre as transações por meio da sua decomposição em níveis estratégicos e operacionais. Sua natureza pode ser mais claramente entendida a partir das dificuldades enfrentadas pelas empresas funcionais. Estas últimas são estruturadas a partir de departamentos que reúnem atividades com características semelhantes (por exemplo: vendas, produção etc.), os quais estão subordinados a um escritório central, que concentra os principais processos de tomada de decisões. Com o crescimento da empresa e a diversificação das suas áreas de atuação, tal forma organizacional é submetida a tensões crescentes no que se refere à gestão das transações. A sobrecarga de decisões sobre o escritório central, somada à insuficiência dos canais de comunicação entre este e as várias partes funcionais, repercutem em dificuldades de implementar um planejamento estratégico eficaz, bem como na emergência de submetas por parte dos departamentos, já que os mecanismos de controle mostram-se limitados. O dilema colocado por esta situação – enfrentado por empresas como a General Motors e a Du Pont na década de 1920 (Cf. Chandler, 1962) – está em descentralizar a tomada de decisões sem permitir que o oportunismo venha a gerar divergências nas condutas dos diversos grupos constituintes da empresa e uma elevação dos custos de transação. A empresa multidivisional apresenta-se como uma inovação organizacional que oferece novos instrumentos para enfrentar estas dificuldades transacionais, na medida em que altera fundamentalmente os sistemas de gestão introduzindo três elementos básicos:

1. A separação das responsabilidades na tomada de decisões, distinguindo entre as *estratégicas*, a cargo do escritório central, e as *operacionais*, a cargo das divisões;⁸
2. A disponibilidade, por parte do escritório central, de um aparato de controle capaz de avaliar o desempenho das divisões, incluindo um *staff* adequado, do poder de manipular os mecanismos de incentivo e da possibilidade de realizar auditorias internas;
3. A centralização dos fluxos de caixa e a sua realocação, por parte do escritório central. Os recursos são distribuídos entre as divisões a partir das suas perspectivas de expansão e sua importância estratégica para os objetivos globais da corporação, o que faz com que, no que se refere às características da alocação destes recursos, a empresa multidivisional funcione como uma miniatura do mercado de capitais.

Dessa forma, a multidivisionalização representa uma alternativa encontrada para, no âmbito de uma estrutura hierárquica, descentralizar a tomada de decisões em um sistema complexo de agentes que executam ações interdependentes, permitindo que o conjunto da organização ganhe a necessária capacidade de se adaptar a novas circunstâncias sem que as condutas sejam excessivamente desviante devido ao oportunismo.

14.4 Organização Empresarial e Processos de Aprendizado

A terceira abordagem a ser estudada enfoca a empresa como uma *acumulação de conhecimentos produtivos*, e, principalmente, como uma entidade capaz de ampliar estes conhecimentos e produzir inovações. Consequentemente, os autores que a desenvolvem dão à estrutura organizacional da empresa um importante papel na preservação deste conhecimento ao longo do tempo e na sua articulação e integração, proporcionando um contexto institucional dentro do qual se processa o aprendizado tecnológico e organizacional.

Apesar de não conceber a grande empresa como uma estrutura de governança que coordena as interações entre os indivíduos que a compõem e atenua as manifestações de condutas oportunistas, esta abordagem compartilha duas características com a teoria dos custos de transação proposta por Williamson. Em primeiro lugar, concebe a empresa como configurando um conjunto de arranjos institucionais qualitativamente distintos daqueles presentes nos mercados, no sentido de que estabelecem um padrão de interação entre os agentes que possuem uma natureza distinta daquele que se verifica nos mercados, seja entre compradores e vendedores ou entre concorrentes. Isso significa que tais organizações possuem uma dinâmica própria, que não pode ser tratada como um resultado de equilíbrio das iniciativas dos agentes individuais na negociação de contratos. Em segundo lugar, adota uma abordagem essencialmente *comparativa*, no sentido de que as proposições sobre os condicionantes da mudança observada da organização das empresas ao longo do tempo são o resultado de uma análise, teórica e empírica, que compara as propriedades de diferentes formas organizacionais e procura identificar os efeitos das diferenças encontradas para o desempenho competitivo das empresas.

Para melhor compreender como a organização da grande empresa está relacionada a processos de aprendizado é necessário, antes de mais nada, atentar para a grande diversidade de atividades envolvidas nos processos de geração de inovações. O desenvolvimento de um novo produto, por exemplo, requer o preciso monitoramento da evolução das necessidades dos prováveis consumidores e a identificação de oportunidades de mercado não aproveitadas por empresas rivais, a combinação das capacitações incorporadas nas equipes de P&D com informações técnicas e científicas obtidas externamente, a transformação dos protótipos em bens com qualidade e baixo custo, e a adaptação do processo produtivo e das características da mão de obra fabril ao novo produto. Além disso, a estratégia de marketing e os canais de distribuição devem ser adequados e eficazes, serviços pós-venda de suporte e manutenção devem ser criados quando necessários e, muitas vezes, é crucial o acompanhamento do produto em condições reais de utilização perante os usuários, para garantir um incremento gradativo do seu desempenho. Em suma, as inovações tecnológicas constituem o resultado, o ponto de convergência de uma ampla gama de processos de aprendizado, muitas vezes tratados parcialmente pela literatura – quando são estudados os trabalhos das equipes de P&D em departamentos formalizados, o *learning by doing* ou o *learning by using*.

Estes processos de aprendizado caracterizam-se, fundamentalmente, por:

1. Incerteza quanto aos resultados dos esforços de aprendizado, bem como dos seus desdobramentos que podem se mostrar necessários antes de obtido o resultado final. Como atividade criativa, o aprendizado é sempre um empreendimento em aberto, obviamente com graduações decorrentes da magnitude do avanço tecnológico a ser empreendido e do seu afastamento diante das soluções já existentes – por exemplo, a incerteza tende a ser maior nas atividades de pesquisa básica do que no desenvolvimento de novos modelos de produtos já existentes. Consequentemente, os participantes diretos não possuem condições de especificar previamente as capacitações necessárias, de modo que o intercâmbio de informações e a troca de experiências, dentro e fora do grupo diretamente envolvido, devem ser intensos. Pode mostrar-se necessário um esforço contínuo e flexível de resolução de problemas que envolve uma ampla variedade de indivíduos, impedindo que se estabeleça uma divisão de tarefas em compartimentos estanques, onde não haja porosidade para os fluxos de conhecimentos;⁹
2. Presença de conhecimentos tácitos, cuja transmissão e compartilhamento não se fazem por meio de uma linguagem formal codificada, exigindo – tanto para a sua difusão quanto para a integração de subconjuntos de informações – relações interpessoais duradouras, acumulação de experiências compartilhadas e aquisição de capacitações pela demonstração da sua aplicação efetiva. Os elementos tácitos de boa parte do conhecimento acumulado nas empresas é uma das causas dos limites de linguagem que tornam as informações contextuais e locais difíceis de serem repassadas além do círculo de alguns indivíduos ou grupos;
3. Abrangência das capacitações necessárias para gerar novos produtos e processos. Em qualquer empresa, as inovações resultam de um conjunto amplo de conhecimentos, muitos dos quais bastante especializados, abarcando não apenas áreas e subáreas técnicas, mas também habilidades e informações “econômicas”. Nestas últimas estão incluídos conhecimentos acerca de mercados, consumidores, mão de obra etc., bem como as experiências de pessoas alocadas em distintos departamentos funcionais das empresas, que resultam em capacitações específicas e úteis para o aprendizado tecnológico. Por exemplo, os engenheiros dos departamentos de assistência técnica geralmente adquirem conhecimento sobre o desempenho

QUADRO 14.2 UMA TAXONOMIA DOS PROCESSOS DE APRENDIZADO

O aprendizado que leva ao acúmulo de capacitações nas empresas pode ocorrer em diferentes formas, às quais usualmente correspondem soluções organizacionais específicas. Ao propor uma taxonomia dos processos de aprendizado, Malerba (1992) identifica seis tipos básicos:

1. O *learning by doing* consiste em uma forma de aprendizado que ocorre no processo de manufatura, após as atividades de P&D terem se completado, materializando-se no desenvolvimento de uma habilidade crescente na produção, que reduz os custos com mão de obra por unidade de produto, ou ainda diminui a incidência de problemas de qualidade. Também se incluem no *learning by doing* aquelas inovações incrementais no produto que resultam de capacitações obtidas por meio do envolvimento direto da empresa no processo produtivo.
2. O *learning by using* ocorre com a utilização do produto pelo seu usuário final, podendo resultar tanto em práticas de operação e manutenção mais eficazes quanto em informações que, repassadas para a empresa produtora, repercutem na introdução de melhorias incrementais no produto.
3. O *learning by advances in science and technology* consiste na absorção e na utilização de novos desenvolvimentos da ciência e da tecnologia, gerados por instituições de pesquisa externas à empresa.
4. O *learning from inter-industry spillovers* envolve atividades voltadas para a absorção de informações e conhecimentos relacionados ao que outras empresas – frequentemente concorrentes – estão fazendo. Um exemplo clássico está em esforços de engenharia reversa, quando uma empresa desmonta e analisa produtos de concorrentes para tentar imitar alguma solução tecnológica que lhe pareça interessante.
5. O *learning by interacting* ocorre quando uma empresa troca informações e realiza alguma forma de cooperação tecnológica com outras empresas, sejam estas fornecedores e usuários situados ao longo da mesma cadeia produtiva ou empresas que atuam em outras indústrias.
6. O *learning by searching* engloba aquelas atividades de busca de novas tecnologias que são internas à empresa e, na maioria dos casos, formalizadas em departamentos ou equipes de pesquisa e desenvolvimento.

dos equipamentos dos clientes e as dificuldades encontradas na sua manutenção, que não estão diretamente disponíveis para seus colegas do departamento de engenharia do produto.

O Quadro 14.2 representa uma taxonomia dos processos de aprendizado.

Esta variedade de processos de aprendizado torna patente o equívoco de relacionar as inovações apenas a atividades de P&D, pois aquelas requerem a junção das esferas técnica e econômica de uma maneira que possa ser desempenhada pela organização ao mesmo tempo em que atenda também às necessidades do mercado, o que implica fusão e cooperação entre muitas atividades nas áreas funcionais de *marketing*, P&D, produção, manutenção etc.¹⁰

Dado que as inovações, radicais ou incrementais, resultam de processos de aprendizado marcados por estas características, um condicionante decisivo da intensidade e eficácia do seu desenvolvimento está nas relações sociais estabelecidas entre os direta e indiretamente responsáveis por gerá-las. Portanto, a inovação não deve ser vinculada apenas à ação criadora de um agente, ou de um grupo de agentes que possa ser tomado como unidade autossuficiente, mas à *conduta interativa* de vários atores, sejam estes empresas, grupos dentro destas ou indivíduos.¹¹ No nível microeconômico, torna-se relevante investigar as instituições que sustentam tais processos coletivos e sociais, proporcionando uma teia de inter-relações entre diversas unidades econômicas e delimitando as possibilidades de trocas de informações, cooperação e compartilhamento de experiências. Em parte, tais instituições são proporcionadas pela estrutura interna da empresa, que oferece os canais necessários para a integração de distintos conhecimentos em uma diversidade de processos de aprendizado.

Por conseguinte, uma proposição central desta abordagem da empresa é que existem relações passíveis de serem identificadas entre os padrões setoriais de introdução e difusão de progresso técnico, de um lado, e certas características organizacionais das empresas, de outro. A depender da base de conhecimentos e informações com a qual opera um dado setor produtivo, determinadas estruturas organizacionais se mostram mais funcionais para o aprendizado tecnológico do que outras. Assim, se em alguns casos uma maior especialização e compartmentalização das atividades tecnologicamente criativas – com o estabelecimento de laboratórios de pesquisa e desenvolvimento relativamente isolados do resto da empresa – promove grandes rupturas e avanços técnicos em uma indústria cuja tecnologia é mais “científica”; em outros, a difusão de conhecimentos por meio de toda a organização e a descentralização podem resultar em progressivas inovações incrementais. O desafio de cada empresa está em, por intermédio do seu aperfeiçoamento organizacional, adequar sua estrutura interna aos requerimentos interativos do aprendizado, os quais, pela própria natureza dinâmica das tecnologias, tendem a se alterar com o tempo.

Note-se que os setores produtivos podem se diferenciar no tocante a que áreas funcionais devem interagir mais intensamente para gerar inovações – em certas condições, os canais de comunicação entre os departamentos de P&D e a produção podem ser cruciais para garantir produtos de baixo custo e boa qualidade; em outras, o primeiro deve interagir mais proximamente

com o *marketing*, para adequar precisamente os projetos dos produtos às necessidades dos clientes. Em certo sentido, pode-se dizer que a base de conhecimentos e informações com que a empresa opera possui não só uma *abrangência*, mas também uma *hierarquização*, que define a importância relativa e o grau de interdependência entre conjuntos de capacitações.

Assim, a consideração deste condicionante organizacional do aprendizado tecnológico torna-se um fator importante para se compreender por que algumas empresas enfrentam dificuldades significativas em realizar a transição para um novo paradigma tecnológico, ainda que tenham usufruído de posições de liderança no passado. Nesses casos, a estrutura organizacional desenvolvida em função das características dos processos de aprendizado associado a uma determinada tecnologia pode se mostrar menos apropriada no contexto de uma nova tecnologia, de maneira que a empresa terá dificuldades em competir com novos concorrentes, que estão comprometidos com soluções organizacionais bem-sucedidas no passado e conseguem explorar com mais eficácia os desafios tecnológicos que estão surgindo.

14.5 Conclusão

As três abordagens discutidas apresentam diferenças importantes, seja no que se refere às hipóteses teóricas básicas que são propostas, seja na opção de enfatizar um ou outro aspecto da organização das empresas. Tais diferenças não implicam uma óbvia incompatibilidade entre elas, mas sim apontam a necessidade de uma avaliação cuidadosa de possíveis complementaridades ou eventuais divergências, sugerindo, respectivamente, a possibilidade de integrá-las em um esforço de teorização mais geral ou o caminho de confrontá-las como competidoras na geração de explicações adequadas da organização das empresas.

Como vimos, a abordagem da empresa como um nexo de contratos e a teoria dos custos de transação proposta por Williamson compartilham uma perspectiva contratual, na qual formas organizacionais são analisadas como respondendo às necessidades de coordenar e conter condutas oportunistas. No entanto, esta convergência no que se refere a um foco comum em um determinado aspecto da interação dos agentes dentro das organizações é acompanhada de divergências teóricas relevantes no que se refere a:

1. Que hipóteses comportamentais são adotadas, pois a abordagem da empresa como um nexo de contratos supõe agentes que tomam decisões para maximizar alguma função objetivo, ao passo que a teoria dos custos de transação é construída a partir da noção de racionalidade limitada de Herbert Simon;
2. Em que medida a estrutura hierárquica da organização das empresas possui uma natureza específica que deve ser tratada teoricamente. Para Williamson, as hierarquias configuram arranjos institucionais irredutíveis às transações mercantis, ao passo que vários autores da abordagem do nexo de contratos acabam por diluir a distinção teórica entre relações contratuais internas à empresa e relações de compra e venda estabelecidas nos mercados de bens e serviços;
3. Se é adequada a opção de modelar a construção de soluções organizacionais como um processo que conduz a um resultado de equilíbrio que pode, de alguma forma, ser caracterizado como ótimo. Os autores da abordagem do nexo de contratos assumem que sim, enquanto a teoria dos custos de transação apenas pressupõe uma tendência à difusão das soluções organizacionais relativamente mais eficientes, dado um conjunto limitado e historicamente predefinido de alternativas.

Por outro lado, a abordagem da empresa como uma acumulação de conhecimentos produtivos enfatiza a funcionalidade da organização das empresas em processos de aprendizado e inovação, o que desde já a diferencia das teorias contratuais. Para alguns defensores desta abordagem, o fato da dinâmica de inovação ser um elemento central da operação das economias capitalistas reduz significativamente a relevância das abordagens puramente contratuais da organização econômica, já que estas se limitariam a analisar as dificuldades contratuais que podem surgir a partir da interação dos agentes com uma base técnica dada.

Além disso, sua inspiração schumpeteriana – e, no caso de alguns dos autores que a desenvolveram, filiação explícita ao programa de pesquisa neo-schumpeteriano ou evolucionista – faz com que seja fácil identificar uma incompatibilidade com a abordagem do nexo de contratos, dado que esta se desenvolve estritamente dentro dos limites da corrente tradicional (*mainstream*). A relação com a teoria dos custos de transação, contudo, é mais complexa, sendo possível vislumbrar um espaço para a busca de complementaridades e, possivelmente, alguma integração teórica.

Uma dificuldade nesta direção resulta de Williamson formular sua proposta teórica quase sempre na forma de uma estética comparativa, na qual decisões quanto à implementação de uma dada solução organizacional são afetadas por uns poucos parâmetros que determinam sua eficiência relativa. A superação desta dificuldade exige algo que o próprio Williamson faz de forma implícita ao discutir exemplos históricos de mudanças organizacionais – a reformulação da teoria dos custos de transação como um instrumental analítico capaz de auxiliar na compreensão de processos de inovação organizacional, na medida em que estabelece teoricamente relações entre propriedades de tipos organizacionais e suas eficiências relativas. Isso significa subordinar as contribuições da teoria dos custos de transação a uma teoria mais geral – e schumpeteriana – do processo de concorrência.

Além disso, o estabelecimento desta relação de complementaridade exige o reconhecimento de que as empresas são instituições complexas e que apresentam múltiplas dimensões e propriedades. Sua organização interna constitui tanto o espaço institucional em que se processa o aprendizado e o acúmulo de capacitações, quanto um conjunto de arranjos institucionais que coordena as interações dos agentes para conter as manifestações do oportunismo. Certamente a importância relativa – para a competitividade da empresa – de um melhor desempenho em termos de aprendizado e inovação e da eficiência em termos de custos de transação pode variar amplamente, inclusive a partir das características estruturais de cada setor produtivo.

Existe uma ampla literatura que permite o aprofundamento deste e de outros temas relacionados à organização da empresa, ainda que raramente traduzida para o português. Sobre a abordagem dos custos de transação, o livro de Oliver Williamson, *The Economic Institutions of Capitalism*, é uma referência clássica e indispesável. Uma exposição mais ampla da organização econômica a partir de uma perspectiva contratual é oferecida no livro-texto de Paul Milgrom e John Roberts, *Economics, Organization and Management*. Uma comparação entre três abordagens apresentadas é feita por Sidney Winter, no artigo “On Coase, Competence, and the Corporation”. Além disso, o periódico *Industrial and Corporate Change* frequentemente apresenta artigos sobre a discussão realizada ao longo deste capítulo.

Notas

1. Por “ficação legal” entende-se aqui uma entidade organizacional construída dentro de uma determinada legislação e que esta permite tratar, para certos fins, como um indivíduo.
2. Nunca é demais relembrar que o oportunismo consiste em uma propensão a adotar um comportamento de busca do interesse próprio de forma maliciosa, ou seja, incluindo esforços no sentido de enganar um outro agente, frequentemente por meio da omissão ou da deturpação de informações. Assim, o risco moral é uma manifestação do oportunismo, cujo exemplo clássico é o do agente que, após contratar um seguro para proteger algum ativo valioso, adota um comportamento mais desleixado no que se refere a precauções contra a ocorrência de algum sinistro, o que faz com que a probabilidade deste ocorrer seja maior após a realização do seguro do que antes. Do ponto de vista das seguradoras, a existência do risco moral significa que, se o valor do prêmio for calculado a partir da probabilidade do sinistro antes do seguro ser vendido, ou mesmo de uma promessa de cada segurado em manter sua conduta inalterada, o resultado será uma dificuldade de cobrir o pagamento dos sinistros ocorridos a partir do valor arrecadado com os prêmios.
3. Isso significa que são gastos recursos – pelo agente – para produzir uma garantia de que não serão tomadas ações que causem danos ao principal, ou que este será compensado. Um exemplo destes custos está no pagamento de taxas – pelo franqueado – em contratos de franquia.
4. Chamadas também de submetas, cuja busca (*subgoal pursuit*) é definida como um esforço de indivíduos no sentido de manipular o sistema administrativo com vistas ao atendimento dos interesses individuais e coletivos, em detrimento do desempenho da organização como um todo.
5. Em princípio, é mais fácil deixar de comprar de um fornecedor do que vender uma divisão em um movimento de desverticalização.
6. Trata-se, em última análise, de desenvolver sistemas decisórios e de ajuste da organização a um ambiente em transformação que sejam eficientes, além de coibir submetas de grupos específicos que colidam com a estratégia global da empresa.
7. Outra inovação nas organizações empresariais, cuja natureza pode ser mais bem compreendida a partir da Teoria dos Custos de Transação, encontra-se no surgimento das empresas multinacionais. Cf. Williamson (1985, p. 290-294) e Teece (1985).
8. Como aponta Simon (1979), uma das propriedades da organização como tal está em permitir a abordagem de decisões complexas a partir da sua decomposição entre indivíduos que assumem responsabilidades específicas, ao mesmo tempo em que são conectados por uma estrutura de comunicação e por relações de autoridade.
9. Assim, no que se refere aos processos de aprendizado, a divisão social do trabalho deve ser acompanhada de intensos fluxos de troca de conhecimentos, para o que são exigidos os canais apropriados.
10. Consultar Capítulos 8 e 19 a respeito.
11. Significando que o aprendizado não é o resultado de um conjunto homogêneo de atividades e relações sociais condensadas em um departamento de P&D.

Bibliografia

- Alchian, A. A.; Demsetz, H. (1972) “Production, Information Costs, and Economic Organization”, *American Economic Review*, v. 62, dezembro, p. 777-795.
- Chandler, A. D., Jr. (1990) *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge (Mass): The Belknap Press of Harvard University Press.
- Jensen, M.; Meckling, W. (1976) “Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure”, *Journal of Financial Economics*, v. 3, pp. 305-360.
- Malerba, F. (1992) “Learning by Firms and Incremental Technical Change”, *Economic Journal*, v. 102, julho, p. 845-859.
- Malerba, F.; Orsenigo, L. (1993) “Technological Regimes and Firm Behavior”, *Industrial and Corporate Change*, v. 2, n. 1.
- Milgrom, P.; Roberts, J. (1992) *Economics, Organization and Management*. Englewood Cliffs: Prentice-Hall.
- Nelson, R. R.; Winter, S. (1982) *An Evolutionary Theory of Economic Change*. Cambridge (Mass.): Harvard University Press.
- Williamson, O. E. (1981) “The Modern Corporation: Origins, Evolution, Attributes”, *Journal of Economic Literature*, v. 19, dezembro, p. 1537-1568.
- Winter, S. (1991) “On Coase, Competence, and the Corporation”, In: Williamson, O. E.; Winter, S. (orgs.) (1991) *The Nature of the Firm: Origins, Evolution, and Development*. Oxford: Oxford University Press, p. 179-195.

Diversificação, Competências e Coerência Produtiva

Jorge Britto

15.1 Introdução

O conceito de empresa elaborado nas visões institucionalista e schumpeteriana, abordadas nos Capítulos 2, 13 e 19, identifica-a como um organismo em contínuo crescimento, associado, por um lado, à sua capacidade de geração interna de lucros e, por outro, às possibilidades que a estrutura da indústria em que ela atua coloca para a “realização” do potencial de acumulação gerado, por meio do processo de investimento. O ponto de partida dessa formulação reside no fato de que a empresa, de uma forma ou de outra, está “obrigada” a crescer, isto é, a reinvestir produtivamente os lucros gerados, podendo ser percebida como um *locus* de acumulação de capital. Nesse processo, a empresa se esforça continuamente para encontrar escoadouros para sua produção atual e futura, mobilizando recursos gerados a partir de sua acumulação interna e/ou de fontes externas de financiamento.

O fenômeno da diversificação refere-se à expansão da empresa para novos mercados distintos de sua área original de atuação. A diversificação é uma alternativa extremamente interessante para viabilizar o crescimento da empresa, na medida em que lhe permite superar os limites de seus mercados correntes, ao mesmo tempo em que possibilita, por meio da gestão de um conjunto de diversas atividades, ampliar o “potencial de acumulação” que influencia a dinâmica do crescimento empresarial. Este capítulo aborda as conexões entre diversificação e crescimento da empresa, enfatizando a discussão das possíveis direções que o processo de diversificação pode tomar – analisadas em termos do grau de similaridade com as atividades originais da empresa – bem como dos seus condicionantes internos e externos. Essa discussão é desenvolvida em seis seções. A primeira delas procura caracterizar de forma sintética o fenômeno abordado, ressaltando alguns problemas presentes no tratamento analítico e na mensuração do fenômeno. A segunda seção aborda as principais direções do processo de diversificação mencionadas na literatura e fornece evidência empírica ao tema investigado. A terceira seção discute a relação mais geral entre diversificação e crescimento da empresa, incorporando uma sistematização dos principais ganhos advindos da diversificação na viabilização do processo de crescimento. A quarta e a quinta seções discutem, respectivamente, os condicionantes do processo de diversificação internos e externos à empresa, atribuindo particular importância à maneira como o perfil de competências das empresas e a dinâmica competitiva das diferentes atividades industriais influenciam o processo. A sexta seção discute as principais formas para viabilizar o processo de diversificação: o investimento em nova capacidade ou a realização de fusões e aquisições.

15.2 Caracterização e Mensuração do Fenômeno

Ao contrário do que propõe a análise da microeconomia tradicional, não é comum que, no processo de concorrência intercapitista, as empresas limitem suas atividades a um único tipo de produto ou mercado. Na verdade, essa concorrência muitas vezes obriga as empresas a se expandirem para diferentes mercados, diversificando seus campos de atuação. Entretanto, o tratamento desse fenômeno não é uma questão trivial. Do ponto de vista analítico, a complexidade do fenômeno da diversificação decorre precisamente do fato dele se constituir em um processo que se confunde com outras dimensões do crescimento empresarial, relacionadas às múltiplas dimensões das interações da empresa com o meio ambiente no qual está inserida. Na literatura de Economia Industrial, a discussão deste fenômeno encontra-se integrada a um corpo teórico mais amplo que aborda, como questão central, o processo de crescimento da empresa, analisado a partir das interações que se estabelecem entre as noções de Estrutura-Conduta-Desempenho. Desse modo, particular ênfase costuma ser atribuída aos impactos da dinâmica concorrencial sobre a lógica do processo de expansão da empresa para novos mercados. No entanto, o processo de diversificação, ao ser viabilizado, amplia o espaço da concorrência para mais além de mercados particulares, convertendo-se em instrumento que permite às empresas sustentar suas posições relativas diante de outras empresas também diversificadas, em um contexto associado à evolução do sistema econômico como um todo.

A complexidade da grande empresa industrial como agente econômico introduz um elemento adicional a ser considerado na discussão do fenômeno da diversificação. Em particular, há indícios de que uma série de fenômenos que ocorrem no interior dessa empresa – relacionada à maneira como se estrutura um determinado perfil de competências e como são equacionados diversos tipos de conflitos intraorganizacionais – afeta a formulação de estratégias capazes de orientar o seu posicionamento diante da evolução do meio externo, as quais poderão, ou não, incorporar um projeto de diversificação para novas atividades. Além disso, o vínculo que se estabelece entre a diversificação e o comprometimento de recursos líquidos da empresa faz com que esta esteja articulada à maneira como é operacionalizada a sua política de investimentos, envolvendo uma determinada sistemática de avaliação de oportunidades e uma determinada forma pela qual o risco e a incerteza são incorporados ao horizonte das decisões empresariais.

No plano mais concreto da realização de estudos empíricos, o tratamento do fenômeno da diversificação também não é trivial. Uma primeira questão refere-se à caracterização em si do fenômeno. É possível pensar que uma empresa *A* que fabrica um único produto pode ser caracterizada como “não diversificada”, ao passo que uma empresa *B* que fabrica três produtos poderia ser caracterizada como “diversificada”. Todavia, essa caracterização pode não se mostrar adequada, se o produto fabricado pela empresa *A* for oferecido em diversos modelos distintos entre si, ou se os três produtos fabricados pela empresa *B* nada mais forem do que pequenas variações de um produto principal. Observa-se, portanto, que a caracterização e a mensuração do processo de diversificação são afetadas pelos elementos utilizados para diferenciar o conjunto de atividades realizadas pelas empresas, podendo-se, a princípio, estabelecer uma distinção entre características relacionadas à base técnico-produtiva e à base comercial dessas atividades.¹ Nesse ponto, uma avaliação mais precisa da amplitude do processo de diversificação depende do sentido atribuído àquelas características no tratamento empírico do fenômeno.

Outro aspecto importante, também relacionado a investigações empíricas do fenômeno da diversificação diz respeito à possibilidade de mensuração desse fenômeno utilizando-se as fontes de informações estatísticas disponíveis. Nesse sentido, é comum, na realização de análises empíricas, a caracterização dos processos de diversificação produtiva das empresas a partir da utilização, como fonte básica de informações, de dados levantados a partir de classificações setoriais de atividades industriais realizadas por órgãos estatísticos oficiais. No entanto, essas classificações setoriais estão normalmente carregadas de arbitrariedades em sua definição, sofrendo os defeitos associados à dicotomia entre definição pela categoria de mercado ou pela semelhança técnica, o que muitas vezes dificulta a compreensão do processo de diversificação em toda a sua amplitude e complexidade.

Apesar dos problemas mencionados, é comum a utilização de determinados indicadores que procuram captar o nível de diversificação de empresas particulares. Dessa forma, é comum a identificação de uma medida da diversificação produtiva *D*, construída tendo como base o número e a importâncias das *n* atividades em que a empresa em questão está presente. Geralmente, esta medida é construída ponderando-se a importância quantitativa de cada uma das atividades realizadas em relação ao valor total da produção da empresa diversificada, definindo-se, assim, um indicador análogo ao índice de Herfindahl-Hirschman utilizado para avaliar o grau de concentração industrial em setores específicos:

$$D = 1 - \sum_i^n P_i^2$$

Onde *P_i* corresponde à proporção do valor da produção da empresa na atividade *i* (que varia de 1 até *n*) em relação ao valor total da produção da empresa em suas diversas áreas de atuação. Desse modo, se a empresa restringe sua atuação a uma única

atividade, D assume valor zero, ao passo que uma repartição igualitária da produção entre diversas atividades conduz a que D assuma um valor $D = 1 - 1/n$. Além disso, a medida utilizada é pouco sensível às atividades de pequeno peso relativo no total do valor da produção da empresa, o que a torna uma medida apropriada para realização de estudos com um recorte setorial.

Outro indicador utilizado para mensurar o grau de diversificação de uma empresa particular é aquele baseado em um índice de entropia dado por:

$$D = \sum_i^n P_i \cdot \log \frac{1}{P_i}$$

Nesse caso, o indicador D assume um valor nulo quando a empresa restringe sua atuação a uma única atividade e o valor $\log n$ quando sua produção se reparte igualmente entre as diversas atividades em que está presente. A principal vantagem desse indicador refere-se à possibilidade de decompor a diversificação total da empresa em dois ou mais componentes. Desse modo, torna-se possível a utilização de informações provenientes de bases estatísticas com níveis distintos de detalhamento quanto à classificação de atividades industriais, decompondo-se a diversificação total em componentes interclasses e intraclasses.

15.3 Direções Possíveis do Processo de Diversificação

Na literatura sobre diversificação de empresas, é comum identificar determinadas direções que orientam o processo. A discussão sobre as possíveis direções do processo de diversificação, incorporadas nas estratégias empresariais, pode ser sistematizada considerando-se dois critérios básicos. O primeiro deles refere-se à proximidade existente entre as atividades originais da empresa e as novas atividades para as quais ela está se expandindo, particularmente em termos da posição desta nos diferentes estágios do processo de transformação de insumos em produtos ao longo das cadeias produtivas. A partir desse critério, é possível distinguir movimentos de diversificação horizontal (ou lateral) daqueles relacionados à diversificação (ou integração) vertical. O segundo critério diz respeito ao grau de similaridade existente entre as atividades originais da empresa e as novas atividades em termos das competências produtivas e gerenciais necessárias para operá-las de forma eficaz. Nesse caso, é possível estabelecer uma distinção entre processos de diversificação concêntrica – nos quais essa similaridade é explicitamente explorada como fonte de vantagens competitivas – e processos de diversificação em conglomerado, nos quais tal aspecto não é considerado pelas empresas na definição das suas estratégias de expansão para novos negócios. A seguir, são discutidas algumas características dessas alternativas, ressaltando-se os principais desdobramentos de cada uma delas em termos do processo de crescimento da empresa.

15.3.1 DIVERSIFICAÇÃO HORIZONTAL

Esse tipo de diversificação consiste na introdução de produtos que, de alguma forma, estejam relacionados aos produtos originais da empresa em termos do mercado atingido e que possam ser vendidos pelos canais de distribuição já estabelecidos ou a partir da extensão destes. Desse modo, a expansão para novos segmentos de mercado associa-se a uma extensão da área de especialização da empresa, seja no que se refere à base tecnológica, seja no que diz respeito à área de comercialização explorada em suas atividades originais. Basicamente, o processo de diversificação visa a possibilitar a exploração de economias de escopo e dos canais de comercialização disponíveis para a empresa.

As características específicas das tecnologias de produto e processo utilizadas pela empresa influenciam esse tipo de diversificação. Alguns exemplos ilustram essas alternativas. No caso de processos contínuos, é importante considerar se estes estão baseados em equipamentos especializados ou em instalações do tipo multipropósito. Neste último caso, torna-se possível diversificar a linha de produtos a partir da utilização dos mesmos equipamentos básicos, como ocorre no caso de empresas atuantes em diversos segmentos da denominada indústria de química fina (envolvendo-se com a produção de fármacos, fertilizantes, aditivos etc.). Mesmo no caso de indústrias associadas a processos contínuos que utilizam equipamentos especializados, é possível que, ao atingir determinado estágio da cadeia de transformação via integração vertical “para frente” (discutida na próxima subseção), as empresas se defrontem com múltiplas alternativas em termos de diversificação horizontal ou lateral. Como exemplos, é possível citar empresas petroquímicas de terceira geração (particularmente as especializadas na produção de plásticos) e empresas siderúrgicas ou de alumínio que se expandem para a etapa de laminação.

No caso de indústrias de montagem, também é possível identificar características tecnológicas que favorecem processos de diversificação horizontal. Como exemplo, é possível mencionar a consolidação de uma “arquitetura modular” do produto, relacionada a um tipo particular de *design* que se baseia numa complexa estrutura de componentes e subsistemas intercambiáveis

e interdependentes. Nesse caso, identifica-se a possibilidade de obtenção de uma maior variedade de produtos, por meio de modificações na maneira como se combinam os diferentes componentes modulares, o que amplia as possibilidades de diversificação horizontal. Como exemplos, é possível citar as indústrias automobilística e eletrônica. Nessas indústrias, a flexibilização dos processos de produção, envolvendo a combinação de um maior grau de automatização com a adoção de novas técnicas organizacionais (células de produção, sistemas *just-in-time* etc.) também favorece movimentos de diversificação horizontal.

Os processos de diversificação horizontal podem ser favorecidos não apenas em função das características específicas das tecnologias de produto e processo nas atividades originais da empresa, como também em função da capacitação mercadológica acumulada por esta em suas atividades de origem. No que se refere à dimensão mercadológica das atividades realizadas, a disseminação de práticas baseadas em esforços de vendas e propaganda, paralelamente ao fortalecimento da marca comercial da empresa, facilita a entrada em novos negócios, não só devido à facilidade da empresa se adaptar a um padrão competitivo semelhante, como em razão da possibilidade dela vir a superar preferências pelos produtores já estabelecidos, atraindo consumidores no novo mercado.

A diversificação horizontal amplia a possibilidade de realização do potencial de acumulação da empresa, elevando sua flexibilidade operacional e diminuindo sua vulnerabilidade em relação às variações cíclicas da demanda em seu mercado original. É também possível elevar a capacidade de financiamento das diversas atividades por meio de práticas de “subsídios cruzados” entre elas, seja por meio de uma transferência de recursos que permita à empresa empreender políticas predatórias em ramos de atividade específicos no intuito de ganhar mercado, seja por meio da implementação de políticas de preço articuladas nos diversos ramos de atuação, as quais, adequando-se às condições de elasticidade-preço da demanda em cada mercado, poderiam permitir à empresa maximizar o seu ganho global.

15.3.2 DIVERSIFICAÇÃO (INTEGRAÇÃO) VERTICAL

No caso da integração vertical, a empresa assume o controle sobre diferentes estágios (ou etapas) associados à progressiva transformação de insumos em produtos finais. Na análise dessa alternativa, é importante considerar que, muitas vezes, a transformação de insumos em produtos, dada a complexidade do processo de produção, não ocorre de maneira linear, podendo diversos estágios convergir para uma etapa particular do processo (gerando ligações convergentes entre eles) ou, alternativamente, um determinado estágio oferecer insumos a serem utilizados sequencialmente em distintos processos (gerando ligações divergentes entre eles). É necessário considerar também os vínculos qualitativos que se estabelecem entre os diversos estágios, associados à maneira como as especificações referentes ao produto/processo em determinado estágio condicionam as atividades nos demais com os quais ele se articula. Quando essas especificações são muito rígidas, envolvendo uma baixa margem de tolerância, a empresa pode se ver obrigada a assumir ela própria a responsabilidade pela produção naquele estágio, sob pena de ter de adquirir no mercado um insumo que não atende satisfatoriamente às especificações de seu processo de produção.

Existem dois tipos básicos de integração vertical: a integração para trás (*upstream*), que corresponde à entrada em estágios anteriores do processo de produção; e a integração para frente (*downstream*) que, ao contrário, envolve a entrada em estágios posteriores.² Algumas diferenças importantes entre essas alternativas podem ser mencionadas. Em primeiro lugar, considerando o produto originariamente gerado pela empresa, percebe-se que a integração para trás não modifica a natureza desse produto, o que já não acontece no caso da integração para frente, que intensifica o processo de elaboração, aproximando-o do estágio associado à geração de um produto final. Em segundo lugar, percebe-se que o processo de integração para frente pode envolver a entrada em atividades não estritamente industriais, vinculadas em especial à distribuição-comercialização do produto final ou à prestação de serviços pós-venda. Finalmente, ao passo que na integração para trás corre uma elevação do valor agregado ao produto, mas não se altera o seu preço nem provavelmente a receita obtida ao longo do processo (supondo que a produção original permaneça inalterada), na integração para frente tanto o valor agregado como o preço e a receita obtidos são modificados.

A literatura sobre integração vertical identifica uma série de fatores que atuam como condicionantes do processo. Em primeiro lugar, destacam-se condicionantes de ordem técnica. Dentre estes, é possível mencionar: (1) desequilíbrios ou desbalanceamentos entre diferentes estágios do processo de produção, decorrentes de mudanças tecnológicas; (2) externalidades tecnológicas, relacionadas a situações em que a existência de padrões de interconexão e interdependência entre atividades do ponto de vista técnico (e/ou mercadológico) favorece processo de integração vertical;³ (3) a necessidade de aglutinação de novas competências quando, se integrando para frente, a empresa se defronta com um ponto de “divergência” das ligações entre estágios do processo produtivo, ou, alternativamente, se integrando para trás, defronta-se com um ponto de “convergência” dessas ligações; (4) a necessidade de equilibrar convenientemente uma cadeia de produção que comporta diferentes estágios, cada um deles possuindo uma escala econômica particular.⁴

Em segundo lugar, destacam-se condicionantes referentes à eficiência econômica. A análise desse tipo de alternativa de crescimento deveria ser confrontada com possíveis ganhos a serem obtidos pela redução de custos de produção e/ou transação. No intuito de sistematizar esses condicionantes, é possível mencionar os seguintes fatores: (1) a redução de custos de produção propiciada pela integração vertical, seja por meio da redução com gastos com estoques, transporte de materiais e com diversos custos indiretos, seja por meio da possibilidade de se utilizar determinados fatores comuns nos diferentes estágios do processo de produção da empresa integrada, incrementando os níveis de eficiência da empresa; (2) a possibilidade da obtenção de ganhos de eficiência, na medida em que, ao longo do processo de integração, a empresa expanda-se para atividades em que é possível obter expressivas economias de escala e escopo; (3) a possibilidade de redução de custos de transação, particularmente quando a integração estiver relacionada à expansão para atividades que envolvem ativos específicos ou nas quais outros entraves dificultam a aquisição de fatores no mercado; (4) a possibilidade de aumentar os níveis de segurança da empresa no aprovisionamento de insumos críticos, tanto em termos quantitativos como qualitativos. A eventual utilização de preços de transferência entre operações realizadas entre unidades de uma empresa integrada que opera em diferentes estágios da cadeia produtiva também podem constituir importantes estímulos para a consolidação desse tipo de expansão.

Em terceiro lugar, destacam-se condicionantes do processo de integração vertical relativas ao processo competitivo nas indústrias em que a empresa atua. Indústrias próximas àquelas tradicionalmente caracterizadas como oligopólio homogêneo apresentam maior grau de integração vertical. A lógica competitiva é que a exploração de mercados cativos internos à empresa e/ou de preços de transferência entre as diversas atividades da empresa integrada, o que pode representar uma importante vantagem competitiva diante de concorrentes com menor grau de integração. Um outro aspecto dessa lógica é a criação de uma proteção contra a concorrência potencial de novos produtores. Isto ocorre porque a integração vertical modifica qualitativa e quantitativamente as condições de entrada na indústria, reforçando eventuais barreiras e exigindo dos produtores potenciais um esforço expressivo para atingir a escala e o nível de integração dos produtores já existentes.

Na discussão da articulação entre a lógica competitiva e a importância do processo de integração vertical, deve-se considerar também as sinalizações oferecidas pela análise da evolução do “ciclo de vida” do produto, evitando-se os prejuízos de uma integração precoce e orientando-se o processo para uma etapa de maior maturidade da indústria, na qual a possibilidade de redução de custos assume maior importância. Por outro lado, é importante considerar também a possibilidade de a concorrência estimular processos de desverticalização de grandes empresas integradas – como tem ocorrido em algumas empresas de bens de consumo duráveis –, o que resultaria na intensificação de práticas de subcontratação e na consolidação de redes de empresas verticalmente estruturadas.⁵

A possibilidade de integração vertical geralmente se associa às estratégias de longo prazo da empresa. Além disso, a integração vertical se diferencia da diversificação horizontal na medida em que torna a empresa em questão mais sensível à instabilidade dos mercados em que atua, reduzindo sua flexibilidade ao elevar o nível de comprometimento do capital produtivo com o conjunto de atividades integradas. Na prática, porém, observa-se que as estratégias de diversificação horizontal e vertical muitas vezes se interpenetram e reforçam mutuamente. Assim, a diversificação horizontal pode preparar e facilitar a integração vertical, do mesmo modo que a integração vertical amplia a base tecnológica da empresa, possibilitando uma diversificação horizontal posterior.

15.3.3 DIVERSIFICAÇÃO CONCÊNTRICA

A diversificação concêntrica pode assumir diferentes formas de acordo com o grau de semelhança entre as competências requeridas para operar de forma eficaz as diversas unidades da empresa diversificada. Na diversificação concêntrica, o aspecto crucial refere-se à exploração do núcleo de competências essenciais da empresa – conceito a ser abordado em maior detalhe adiante – como fonte de vantagens competitivas que possibilita ou favorece a entrada em novas áreas de atuação. Desse modo, a empresa procura manter um padrão coerente de expansão para novos mercados, explorando e alargando suas competências originais.

Como resultado de movimentos de diversificação concêntrica, consolida-se um tipo de empresa diversificada que está presente em diversos mercados que se encontram relacionados entre si do ponto de vista técnico-produtivo e/ou do ponto de vista das capacitações gerenciais necessárias para operar aquelas unidades de maneira eficaz. Esse tipo de empresa baseia sua estratégia de diversificação no acúmulo de uma capacitação genérica que pode ser utilizada em diferentes mercados, o que lhe confere uma vantagem concorrencial em relação a outras empresas que não dispõem desse tipo de capacitação. Essa capacitação pode, inclusive, estar restrita ao plano gerencial. Nesses casos, supõe-se que a empresa pode auferir ganhos em virtude do intercâmbio de recursos produtivos, financeiros e gerenciais entre suas diversas unidades, o que resultaria em ganhos de eficiência para estas.

Na medida em que a diversificação concêntrica busca explorar o núcleo de competências essenciais da empresa como fonte de vantagens competitivas em novos mercados, ela tende a sobrepor e a se confundir com as alternativas de diversificação horizontal (principalmente) e integração vertical discutidas anteriormente. Desse modo, os mesmos condicionantes que eram válidos para aquelas direções exercem influência na determinação da viabilidade da diversificação concêntrica.

É possível identificar algumas diferenças entre o movimento mais geral de diversificação concêntrica e as demais alternativas anteriormente mencionadas em termos dos impactos advindos sobre as condições concretas da concorrência nas atividades para as quais a empresa está se expandindo. Assim, ao passo que a diversificação horizontal tem um impacto direto sobre as barreiras à mobilidade (ao fortalecer a posição competitiva dos produtores que atuam em vários mercados próximos) e a integração vertical sobre as barreiras à entrada (impondo uma maior escala de produção intraempresa), a diversificação concêntrica, por si só, não tende a reforçar diretamente tais barreiras, na medida em que, nesse caso, as empresas atuariam em vários mercados articulados de maneira relativamente tênue. Desse modo, a empresa poderia se defrontar com problemas para se posicionar dinamicamente diante da concorrência (efetiva e potencial) nos diferentes mercados, desde que não tenha ela própria uma capacitação econômico-financeira condizente com seu nível de diversificação, e não disponha de canais facilmente acessíveis para a transferência de recursos entre atividades, de acordo com os sinais de mercado.

É importante lembrar também que, provavelmente, o ritmo da incorporação de novos avanços tecnológicos não é o mesmo nos diferentes mercados em que a empresa atua. Assim, a necessidade de acompanhar o estado da arte nas atividades tecnologicamente mais dinâmicas pode acabar fazendo com que a empresa ajuste sua estratégia, com vistas a se precaver contra os riscos de um nível de diversificação excessivo e da consequente particularização dos recursos disponíveis.

15.3.4 DIVERSIFICAÇÃO EM CONGLOMERADO

A evolução de uma estratégia de diversificação concêntrica para uma estratégia de diversificação em conglomerado envolve uma progressiva redução dos níveis de sinergia entre as atividades da empresa, até atingir-se uma situação em que essas inter-relações são tão tênues que a empresa diversificada poderia ser visualizada como um conjunto de atividades não correlacionadas entre si. Nesse caso, apesar de a presença em diferentes atividades ampliar o horizonte de diversificação da empresa, o caráter extremamente díspar dessas atividades pode acarretar problemas em termos da consolidação de um nível de competência que lhe permita posicionar-se satisfatoriamente nos diferentes mercados em que atua.

Considerando fatores estritamente tecnológicos, existem três modificações para que a empresa diversifique-se de forma conglomerada. A primeira delas envolve o surgimento de novas oportunidades atrativas para a realização de investimentos, que podem não estar conectadas às atividades originais da empresa. A segunda alternativa decorre do impacto desestabilizador de uma inovação tecnológica mais radical sobre as atividades originais da empresa. Finalmente, a diversificação em conglomerado pode ser induzida por um nível extremamente específico de especialização da empresa, o que a impede de estabelecer relações de sinergia com outras atividades.

Diferente dos fatores estritamente tecnológicos, é possível caracterizar situações em que a diversificação em conglomerado é deliberadamente explorada como alternativa de crescimento. A opção por esse tipo de estratégia pode ser motivada pelo desenvolvimento de modernas tecnologias gerenciais, acrescido do fortalecimento das atividades de planejamento, que permite a identificação de níveis de sinergia entre atividades aparentemente desconectadas. A possibilidade de obtenção de um acesso privilegiado ao mercado de capitais também possibilita a realização de operações de fusões e aquisições que reforçam esse tipo de estratégia. Além disso, a implantação de modernas técnicas de planejamento financeiro em empresas diversificadas amplia as possibilidades de obtenção de ganhos por cada ação adquirida, reforçando cumulativamente a expansão para novos mercados. Observa-se também que as operações de fusão ou aquisição de empresas realizadas a partir de estratégias de diversificação em conglomerado são menos vulneráveis às proibições e obstáculos impostos por agências de regulação da concorrência do que as mesmas operações no caso de uma diversificação horizontal ou vertical.

A ênfase numa estratégia de diversificação conglomerada também pode acarretar alguns problemas para viabilização de um crescimento autossustentado da empresa. A entrada em atividades muito díspares entre si pode dar origem a uma estrutura organizacional confusa e a uma elevação dos níveis de risco do portfólio de negócios da empresa, o que pode acabar fazendo com que alguns proprietários da empresa – no caso de serem empresas baseadas em uma estrutura dispersa de propriedade – tentem se desfazer de suas ações no mercado. Além disso, o comprometimento de recursos disponíveis com uma grande variedade de produtos e mercados também pode não ser bem visto pelos acionistas, na medida em que resulte em aumento do risco e redução da flexibilidade operacional. Outro limitante é sua maior vulnerabilidade diante de uma reversão cíclica mais pronunciada, uma vez que a falta de articulação entre atividades pode tornar complicada uma reorganização de cunho defensivo destas diante de um contexto recessivo. Finalmente, é necessário ter cuidado com a ênfase atribuída à entrada em indústrias

com grande potencial tecnológico, pois se, no curto prazo, as perspectivas de ganho tendem a ser promissoras, é possível que a incapacidade de a empresa desenvolver a necessária capacitação exigida para o acompanhamento do ritmo de desenvolvimento tecnológico acabe por abortar o próprio processo de diversificação.

15.4 Diversificação e Crescimento da Empresa

A discussão do processo de diversificação como alternativa para viabilizar o crescimento da empresa fundamenta-se na análise da ação de um agente particular: a empresa “diversificada e diversificante”, que apresenta uma estrutura organizacional e um perfil de competências bastante complexos, o que reflete na sua presença em diferentes mercados. Essa discussão concebe a expansão para novos mercados como alternativa recorrente no horizonte decisório da empresa, viabilizando a realização do seu potencial de acumulação em seus mercados de origem e a expansão em novas áreas de atuação.

A diversificação proporciona uma série de benefícios que permitem acelerar o ritmo de acumulação e crescimento da empresa e que podem ser decorrentes de três grupos distintos de fatores. O primeiro deles está associado à busca de novas áreas de atuação como alternativa para acelerar o ritmo de crescimento da empresa. O segundo grupo de fatores envolve benefícios relacionados ao incremento da eficiência técnico-produtiva das empresas, envolvendo a exploração de sinergias e melhor utilização dos recursos disponíveis. Finalmente, um terceiro grupo de fatores envolve benefícios relacionados à ampliação da rentabilidade da empresa ao longo do tempo. O Quadro 15.1 apresenta uma sistematização e exemplos dos diversos tipos de ganhos advindos do processo de diversificação.

A diversificação é analisada nos vários modelos de Economia Industrial tendo como contraponto um objetivo mais geral: a compreensão da dinâmica de crescimento da empresa. A argumentação desenvolvida por Edith Penrose, em 1959, comprehende a empresa como entidade complexa orientada para o crescimento, e não como mero agente definidor de preços e de quantidades de equilíbrio, como visto no Capítulo 2. Enfatiza-se, nesse tipo de análise, que o processo de crescimento tem a capacidade de internalizar a tomada de decisões à empresa, ao reforçar a dimensão administrativa/deliberativa e ao diminuir a importância das forças de mercado na alocação dos recursos produtivos. O papel do mercado, como força abstrata que conduz necessariamente a uma situação de equilíbrio, é substituído pela ação da concorrência intercapitalista.

Segundo Penrose, a grande empresa industrial influencia diretamente suas possibilidades de diversificação para uma maior gama de mercados. Em primeiro lugar, essa empresa é intrinsecamente complexa do ponto de vista administrativo e organizacional, caracterizando-se pela presença de diversos níveis hierárquicos, pela separação entre propriedade e gerência, pela presença de múltiplos objetivos que norteiam suas decisões e pela sua atuação como unidade autônoma de planejamento, cujas atividades estão relacionadas e coordenadas. Em segundo lugar, a empresa é vista como um agrupamento de recursos produtivos tangíveis e intangíveis, que são capazes de gerar serviços responsáveis pela viabilização do processo de produção. Dentre esses serviços, particular importância é atribuída aos denominados serviços gerenciais (*managerial services*), que assumem um caráter idiossincrático em cada empresa, influenciando decisivamente seu processo de crescimento. Em terceiro lugar, é possível identificar condicionantes do processo de crescimento que dizem respeito a fatores internos à empresa, envolvendo a capacidade empresarial *vis-à-vis* os níveis de risco e incerteza atribuídos às expectativas de crescimento.

A disponibilidade de “serviços gerenciais” afeta decisivamente o processo de crescimento empresarial. A autora parte da hipótese de que os serviços gerenciais estão associados a um tipo de conhecimento que é acumulado de forma idiossincrática por cada empresa, por meio da experiência obtida com a operação dos negócios e com a implementação de planos de expansão. Ao mesmo tempo, é comum que, ao longo do processo de crescimento, uma parcela desses serviços permaneça ociosa, tornando-se rentáveis apenas quando existe uma possibilidade de expansão para novos mercados. A disponibilidade desses serviços gera incentivos à expansão para novos mercados que independem diretamente de estímulos ambientais. Além disso, ao longo do processo de crescimento, o estoque de “serviços gerenciais” modifica-se não apenas do ponto de vista quantitativo como qualitativo, devido a três fenômenos: (1) a aquisição de novos serviços, relacionados à expansão do quadro de pessoal envolvido com atividades de gestão, e a integração dos novos serviços ao estoque de serviços preexistente; (2) a existência de um certo lapso temporal para que os novos serviços adquiram a experiência necessária à identificação e exploração das possibilidades de expansão; (3) a tendência a que ocorram, ao longo do processo, mudanças na estrutura organizacional da empresa que afetam as possibilidades de exploração do potencial dos serviços gerenciais no processo de expansão.

A disponibilidade de serviços gerenciais limita as possibilidades de crescimento da empresa em cada momento no tempo. Em particular, ao se diversificar para novos mercados, a empresa defronta-se, de forma recorrente, com limites impostos pela necessidade de expandir seus serviços gerenciais, devido a problemas de coordenação e integração desses serviços àqueles existentes.

QUADRO 15.1 BENEFÍCIOS E EXEMPLOS ADVINDOS DO PROCESSO DE DIVERSIFICAÇÃO

- Aceleração do Ritmo de Crescimento da Empresa
 - Mudança de Direção: refere-se a situações em que a diversificação possibilita um reposicionamento competitivo da empresa, diante de constrangimentos do ambiente que afetam sua rentabilidade a longo prazo. Essa diversificação pode ser estimulada tanto devido à perspectiva de lucros declinantes em virtude de problemas inerentes ao ciclo de vida do produto nos mercados onde a empresa já opera, como em reação à perspectiva de obsolescência tecnológica para seus produtos e processos. A expansão para novos mercados pode também ser resultante de uma reação a mudanças no ambiente econômico e institucional que podem ameaçar a posição competitiva da empresa em seus mercados originais.
 - Exploração de Oportunidades Atrativas: refere-se à expansão para novos mercados que apresentam um potencial de expansão particularmente atrativo, possibilitando acelerar o ritmo de crescimento da empresa. Geralmente estes novos mercados estão associados a oportunidades tecnológicas e/ou mercadológicas promissoras. Como exemplos, é possível mencionar mercados emergentes nas áreas de biotecnologia e novos materiais, bem como as possibilidades abertas pela dinamização do comércio eletrônico e outras formas de *e-business*.
- Incremento da Eficiência Técnico-Produtiva
 - Eficiência na Alocação de Recursos: refere-se aos ganhos obtidos com uma utilização mais racional dos recursos possibilitada pela diversificação. Em termos produtivos, destacam-se o aproveitamento do excesso de capacidade e a exploração de economias de escala e escopo. Do ponto de vista mercadológico, destacam-se a geração de coprodutos, a exploração de vínculos comerciais com outras empresas e a utilização de canais de distribuição e do prestígio já conquistado no mercado. Finalmente, em termos gerenciais, destaca-se a utilização das competências gerenciais da empresa em outros ambientes.
 - Incremento de Sinergias: relaciona-se à exploração de quatro tipos básicos de sinergia entre as atividades: a sinergia comercial decorre da utilização comum de canais de distribuição e de quadros de administração de vendas; a sinergia operacional associa-se à utilização mais intensa de instalações e pessoal, bem como ao aproveitamento de vantagens de curvas de aprendizado comuns; a sinergia de investimentos resulta do uso conjunto das plantas, da repartição de estoques e da transferência de pesquisa e desenvolvimento de um produto para outro; finalmente, a sinergia de administração associa-se à experiência acumulada com problemas estratégicos, organizacionais e operacionais da mesma natureza.
- Estabilização e Dinamização das Vendas
 - Redução do Risco: relaciona-se à dificuldade para se estimar com precisão o comportamento da demanda ao longo do ciclo de negócios nas diversas atividades da empresa. Nesse caso, a diversificação atua para compatibilizar os ciclos de negócios nos diversos mercados, minimizando o risco e maximizando o retorno a longo prazo obtido.
 - Estabilização de Ganhos: refere-se à obtenção de menores flutuações quanto à rentabilidade do conjunto de atividades da empresa diversificada. Duas tendências principais podem ser associadas a este tipo de ganho: a diversificação para produtos de demanda estável, menos sensíveis a flutuações cíclicas e à diversificação para produtos com menor elasticidade-preço, que funcionariam como proteção contra os impactos desfavoráveis de variações cíclicas da demanda.
 - Adaptação às Necessidades de Consumidores: relaciona-se ao aprofundamento dos laços das empresas com seus consumidores por meio da diversificação, particularmente pelo lançamento de novos produtos explorando a similaridade destes em relação à área de comercialização da empresa. Em contrapartida, o contato prévio com consumidores potencialmente interessados em novos produtos facilita o aprofundamento do processo de diversificação.

A análise dos limites internos à empresa no processo de diversificação é desenvolvida por Penrose em três etapas, cada uma delas associada a um efeito distinto. Considerando-se um dado período de tempo – em analogia com a abordagem estática da microeconomia tradicional – é possível considerar a existência de um estoque fixo de serviços gerenciais, cuja utilização poderia estar associada à presença de deseconomias que afetam negativamente os custos unitários de produção. No entanto, considerar como fixo o estoque de serviços gerenciais implica desconhecer os efeitos de ganhos de experiência que poderiam contrabalançar os efeitos das deseconomias supracitadas. Um segundo efeito, relacionado a situações em que os incentivos proporcionados pela disponibilidade de serviços gerenciais tendem a desaparecer. Em particular, as seguintes situações podem conduzir a esse tipo de efeito: (1) quando a expansão para novos mercados envolve produtos totalmente novos para a empresa ou o estabelecimento de plantas ou subsidiárias em novas áreas geográficas; (2) quando existe um elevado grau de especialização nos recursos comprometidos com novas atividades, o que faz com que desapareçam as conexões com os serviços e a experiência disponíveis nas atividades originais da empresa; (3) quando a vantagem proporcionada pela existência de tais serviços resstringe-se à maior facilidade de entrada na nova atividade, devido a efeitos de reputação.

O terceiro tipo de efeito considerado diferencia-se da abordagem tradicional das deseconomias de escala, na medida em que diz respeito ao ritmo de expansão das atividades da empresa e não à sua dimensão absoluta. Admite-se, que a proporção dos serviços gerenciais requeridos para manter as operações da empresa em relação ao total de serviços produtivos pode se elevar. Este efeito é mais intenso quando o porte da empresa se torna excessivamente elevado em relação ao ritmo das respostas exigidas diante das pressões competitivas dos ambientes em que ela atua. Esse problema pode se agravar na medida em que

é improvável que a disponibilidade de serviços gerenciais possíveis de serem comprometidos com o processo de expansão – que funcionam como um estímulo interno à aceleração do crescimento empresarial – cresça a uma taxa mais elevada do que o crescimento dos serviços produtivos totais da empresa. Desse modo, tende a ocorrer um decréscimo da produtividade média dos serviços gerenciais mobilizados, enquanto eleva-se o custo unitário desses serviços para taxas de crescimento acima de um certo nível. Em consequência, observar-se-ia uma elevação dos gastos com serviços gerenciais por unidade monetária adicional obtida no processo de expansão para novos mercados. Esse efeito ficou conhecido na literatura de Economia Industrial como efeito Penrose.

A intensidade do efeito Penrose depende das características específicas da diversificação. A complexidade do processo de expansão, isto é, da variedade de novos produtos e mercados em que a empresa irá atuar aumenta a necessidade de comprometimento de serviços gerenciais com o processo de expansão. A proximidade entre as atividades contempladas no processo de expansão e as atividades originais da empresa reduz a necessidade de serviços gerenciais.

Outros autores procuram avançar no sentido do estabelecimento de conexões mais diretas entre o processo de diversificação e as possibilidades de crescimento da empresa nos diferentes mercados em que ela atua. Dentre estes, uma referência importante é a de Robin Marris. O autor pressupõe que a maximização do crescimento é o objetivo mais geral da empresa, vinculando-o diretamente ao processo de diversificação. O problema refere-se à possibilidade da empresa determinar a taxa de crescimento máximo que pode obter por meio da diversificação, sujeita a duas restrições básicas: uma restrição mercadológica – relacionada à possibilidade de compatibilizar-se a expansão da oferta com o crescimento da demanda nos novos mercados – e uma restrição financeira (suposta como exógena no modelo), relacionada ao montante de recursos e no nível de risco que a empresa está disposta a comprometer no processo.

Supõe-se que a expansão para novos mercados possibilita à empresa elevar sua taxa de crescimento. No entanto, o sucesso desse processo dependerá do volume de recursos mobilizados para lançamento de produtos nos novos mercados. Pode-se identificar um limite para o volume de recursos, uma vez que, para um montante predeterminado de gastos com o lançamento de produtos, a taxa de sucesso decrescerá quando o número de novos produtos aumentar. A razão implícita para o decréscimo é que o montante de gasto destinado a cada um deles irá diminuir.

Desse modo, é possível associar esse processo a uma taxa de crescimento da demanda (G_D), que é função positiva da taxa de diversificação (n) e da taxa de sucesso dos produtos lançados (h). A taxa de sucesso dos produtos lançados, por sua vez, estaria relacionada positivamente à própria taxa de diversificação e aos gastos em publicidade e P&D. No que se refere aos gastos em publicidade e P&D, estes estariam inversamente relacionados à margem de lucro da empresa (L), na medida em que, dadas as condições concretas da competição nos mercados em que ela atua, não tem como elevar autonomamente essa margem para financiar aqueles gastos.

Por outro lado, com o incremento da diversificação, supõe-se que o sucesso dos novos lançamentos cresce a taxas decrescentes, diante da limitação das competências e da capacidade mercadológica da empresa. Em consequência, para um dado volume de recursos destinado ao lançamento de novos produtos – associado a uma margem de lucro particular – a taxa de crescimento da demanda da empresa (associada à taxa de sucesso mencionada) aumentará menos que proporcionalmente em relação à taxa de diversificação, conforme demonstra a curva I_1 no Gráfico 15.1. Quando o montante de recursos disponíveis para o lançamento de novos produtos aumenta (em virtude de uma queda na margem de lucro), a curva se desloca para cima, permitindo a obtenção de maiores taxas de crescimento para cada nível de diversificação.

Existe, porém, uma contrapartida da elevação dos gastos com a promoção dos produtos: a redução dos recursos que podem ser direcionados para investimentos produtivos nos novos mercados. Nesse sentido, é importante considerar as interações existentes entre o montante de recursos destinados a esses investimentos, a taxa de crescimento da oferta da empresa e o seu nível de diversificação.

Como hipótese, supõe-se que a taxa de crescimento da oferta (G_S) é função positiva do lucro total (LT) disponível para financiar o investimento, da relação produto/capital (Y/K) e do nível de diversificação (n). O lucro total disponível para financiar a expansão da oferta está positivamente correlacionado à margem de lucro da empresa (L). Além disso, admite-se que esse lucro total está correlacionado no nível de diversificação da seguinte forma: quando a diversificação se intensifica, a produção e o lucro aumentam, atingem um máximo e começam a cair, em decorrência da presença de economias e deseconomias internas e da tendência à diminuição da relação produto/capital. Supõe-se, ademais, que o coeficiente de capital da empresa, como um todo, se eleva quando se intensifica o processo de diversificação, uma vez que o aumento menos que proporcional dos produtos de sucesso em relação aos novos produtos pressiona para cima aquele coeficiente. Em consequência, define-se uma taxa de crescimento da oferta que é função da margem de lucro e do nível de diversificação da empresa. A curva F_1 do Gráfico 15.1 reflete essa tendência, evidenciando que, para uma determinada margem de lucro, a taxa de crescimento da oferta da empresa tende a se reduzir quando se acentua o processo de diversificação, devido à pressão exercida sobre o coeficiente de capital. Por

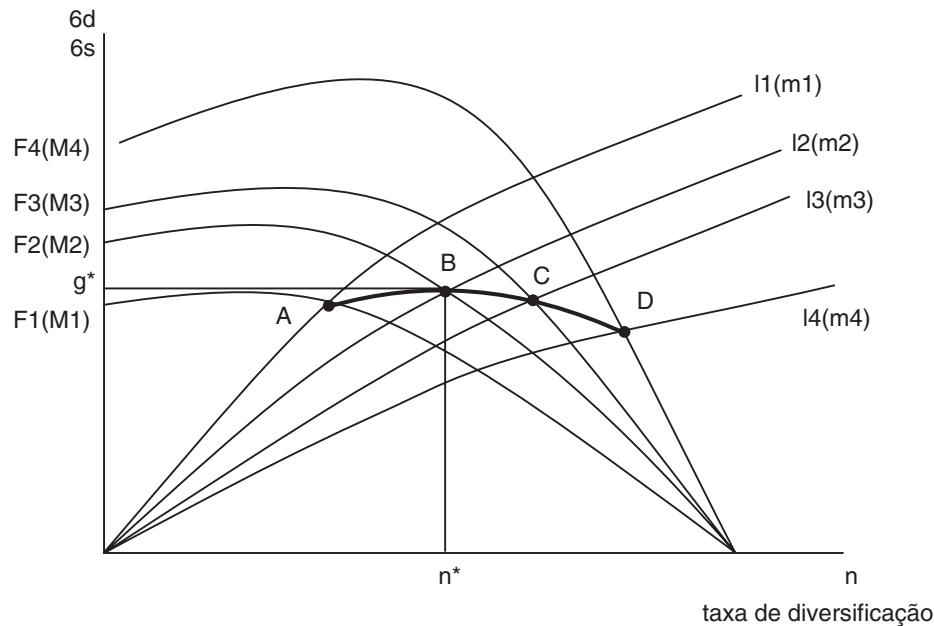


Gráfico 15.1 Taxa ótima de diversificação.

outro lado, a elevação da margem de lucro implica um aumento dos recursos disponíveis para investimento e, portanto, um deslocamento para cima da supracitada curva.

Considerando os dois processos em conjunto, percebe-se que existe, por um lado, uma interação entre a taxa de crescimento da demanda e a taxa de diversificação mediada pelos recursos destinados ao lançamento de novos produtos (o que afeta a perspectiva de crescimento da demanda nos respectivos mercados). Por outro lado, também existe uma interação entre a taxa de crescimento da oferta e a taxa de diversificação mediada pelo volume de recursos destinados a investimentos (o que afeta a perspectiva de crescimento da oferta naqueles mercados). Essas duas formas de aplicação de recursos concorrem entre si pela alocação de uma massa de lucros definida exogenamente, de tal forma que uma redução nos recursos destinados a investimento acarreta uma ampliação dos recursos disponíveis para o lançamento de novos produtos e vice-versa. Para cada repartição desses recursos, seria possível definir um ponto de equilíbrio que associa uma determinada taxa de crescimento da oferta e da demanda a determinada taxa de diversificação, conforme denotam os pontos A, B, C e D no Gráfico 15.1.

Considerando as repartições possíveis dos recursos disponíveis entre investimento e promoção de novos produtos, seria possível definir uma curva de “crescimento equilibrado” associada à união dos pontos de equilíbrio relacionados a cada repartição. De fato, considerando as funções $G_D = f(n, L)$ e $G_S = f(L, n)$, é possível estabelecer $G_D = G_S$ como condição para um “crescimento equilibrado”, o que gera um sistema com uma equação e duas incógnitas (L e n), que pode ser resolvido predeterminando-se o valor de uma das variáveis. Predeterminando-se o valor da margem de lucro (L), são definidos vários pares de curvas (G_D e G_S) em função da taxa de diversificação (n), a partir dos quais são obtidos pontos correspondentes à curva de crescimento equilibrado (para os quais $G_D = G_S$). Essa curva – correspondente no Gráfico 15.1 à união dos pontos A, B, C e D – define, para margens de lucro exógenas, as alternativas em termos de um grau de diversificação compatível com o crescimento dos mercados para os quais a empresa está se expandindo.

Este referencial fornece às empresas uma indicação geral de uma estratégia que compatibilize o risco crescente, associado à entrada em novos negócios, com as consequências advindas do emprego nessas atividades de seu capital produtivo.

15.5 Condicionantes Internos à Empresa no Processo de Diversificação

Evidências levantadas pela moderna literatura de Economia Industrial demonstram que o processo de diversificação raramente ocorre de maneira aleatória. Ao contrário, a expansão para novos mercados é decisivamente afetada pelo nível de “especialização” preexistente das empresas, bem como pelas sinalizações do ambiente competitivo em que estão inseridas. Esses fatores conferem uma lógica particular ao processo de diversificação, baseada em certa *coerência* quanto à definição de novas oportunidades de negócios a serem exploradas no processo de crescimento empresarial.

O nível de especialização de cada empresa está correlacionado a duas dimensões. A primeira delas diz respeito à maneira como se articulam as diversas atividades de uma empresa presente em diferentes mercados, compreendendo aspectos organizacionais da sua estrutura interna. Supõe-se, nesse sentido, que é possível identificar diferentes modelos organizacionais por meio dos quais a empresa diversificada pode se estruturar e que afetam sua capacidade de resposta diante de estímulos competitivos provenientes do meio externo.

A segunda dimensão é o nível de especialização de cada empresa expresso nos conceitos de base tecnológica e área de comercialização, também formulados por Penrose. As bases tecnológicas compreendem cada tipo de atividade produtiva que utiliza máquinas, processos, capacitações e matérias-primas complementares e estreitamente associados no processo de produção. Como visto no Capítulo 2, é possível não apenas que diferentes indústrias venham a ter a mesma base tecnológica (automóveis e caminhões, por exemplo), como também que uma indústria específica envolva bases tecnológicas distintas (como no caso da indústria farmacêutica, que tanto pode envolver processos químicos como biotecnológicos). O conceito de área de comercialização, por sua vez, refere-se a um grupo de clientes sobre os quais a empresa espera exercer influência por um mesmo programa de vendas. As características dessa área de comercialização estão particularmente associadas aos procedimentos mercadológicos adotados nos ambientes em que a empresa distribui seus produtos. Considerando esses conceitos, a coerência do processo de diversificação manifesta-se na existência de características comuns entre a base tecnológica e a área de comercialização das atividades originais da empresa e as das novas atividades incorporadas pelo processo de diversificação.

É possível delimitar, com base no nível de especialização expresso na base tecnológica e na área de comercialização da empresa, um determinado horizonte de diversificação, que compreende indicações quanto às direções em que a expansão para novos mercados se mostra mais factível. Diante desse horizonte, é comum que as empresas privilegiem uma linha de menor resistência no processo de expansão, baseada num menor afastamento em relação à sua área de especialização original. No entanto, ao mesmo tempo em que a expansão para novos negócios encontra-se restringida pelo horizonte de diversificação, essa expansão, uma vez ocorrida, alarga continuamente aquele horizonte. Desse modo, o processo de crescimento empresarial com base em indicações oferecidas pelo horizonte de diversificação assume um caráter nitidamente cumulativo, pois a diversificação envolve necessariamente o contato com novas bases tecnológicas e áreas de comercialização, o que amplia o leque de oportunidades a serem exploradas e reforça o movimento de expansão para novos mercados.

Não obstante o caráter cumulativo do horizonte de diversificação da empresa, existem alguns outros aspectos relacionados ao seu nível de especialização que também podem desempenhar um papel importante na criação de estímulos à expansão para novos mercados. Em particular, é possível identificar determinadas atividades funcionais integradas à estrutura organizacional da empresa, que funcionariam como polo irradiador de oportunidades, capazes de serem aproveitadas em um projeto de diversificação. As atividades de P&D, por exemplo, adquirem grande importância, principalmente quando os conhecimentos gerados podem ser aplicados em atividades distintas daquelas nas quais a empresa se encontra especializada. Do mesmo modo, os esforços de promoção e vendas permitem o fortalecimento de relações de clientela com eventuais consumidores e favorecem a consolidação da reputação da empresa no mercado, fatores também capazes de favorecer o aprofundamento de uma estratégia de diversificação.

Outro elemento que desempenha papel importante na definição dos condicionantes internos do processo de diversificação é a existência de serviços produtivos ociosos. Esses serviços envolvem, por um lado, indivisibilidades técnicas associadas à presença de economias de escala e, por outro, o surgimento de novos serviços produtivos no curso do processo normal de operação e expansão da empresa. Segundo Penrose, esses novos serviços funcionariam como economias de crescimento, distintas das economias de escala tradicionais, impulsionando a diversificação na medida em que fornecem às empresas vantagens em relação às demais para oferecer novos produtos no mercado. A obtenção dessas economias de crescimento se autorreforça, pois as novas unidades, uma vez estabelecidas, começam a crescer do mesmo modo que o resto da empresa, em resposta a economias similares.

A análise original de Penrose pode ser complementada por interpretações que concebem as empresas como organizações dotadas de competências específicas, que evoluem ao longo do tempo como resultado de processos internos de aprendizado e em função de mudanças adaptativas realizadas diante das alterações nas condições ambientais. Estas interpretações ressaltam a importância de processos cumulativos de aprendizado que alteram permanentemente as competências organizacionais e tecnológicas das empresas. As habilidades genéricas (*metaskills*) adquiridas pelas organizações ao longo desse processo resultam na consolidação de competências organizacionais que definem opções estratégicas mais factíveis para as empresas e delimitam as oportunidades de expansão para novos mercados possíveis de serem exploradas. Além disso, a consolidação de mecanismos de aprendizado e de rotinas organizacionais orientadoras do processo decisório exercem importante influência sobre as possibilidades de expansão da empresa para novos mercados. Em particular, a diversificação será facilitada se estiver orientada para indústrias onde se possa reproduzir, em parte, as rotinas já experimentadas pela empresa em seus mercados de origem, ou nas quais a experiência obtida por meio dos processos de aprendizado lhe permita usufruir algum tipo de vantagem

competitiva. A diversificação seria também facilitada quando orientada para setores em que a empresa possa aprofundar as estratégias mercadológicas e tecnológicas exploradas em seus mercados de origem.

Em contrapartida, cabe ressaltar dois impactos dinâmicos do processo de diversificação sobre o perfil de competências das empresas. Por um lado, a diversificação permite incrementar a capacitação técnico-produtiva dos agentes, a partir da ampliação das fontes potenciais de aprendizado tecnológico, bem como lhes permite explorar oportunidades tecnológicas e mercadológicas atrativas em mercados em que as competências já acumuladas representem algum tipo de vantagem competitiva diferencial. Por outro lado, a diversificação permite reduzir o risco implícito na desestabilização do nível de capacitação das empresas em virtude de mudanças tecnológicas mais radicais, facilitando a reconfiguração do seu perfil de competências e a reorientação do processo de crescimento em direções mais promissoras.

A análise de Prahalad e Hamel (1990) segue essas diretrizes e propõe o conceito de competências essenciais (*core competences*). O problema básico abordado pelos autores refere-se à identificação do núcleo de competências das empresas e de eventuais dilemas que costumam estar presentes na exploração dessas competências no processo de crescimento empresarial. O gerenciamento das competências essenciais da empresa envolve cinco dimensões básicas. A primeira delas compreende o desenvolvimento destas competências ao longo do tempo, envolvendo um processo cumulativo baseado em pequenos avanços em relação à situação inicial. A segunda dimensão baseia-se na difusão de competências no nível da empresa, correlacionando-se ao grau de articulação e formalização do conhecimento no plano intraempresarial. A terceira dimensão envolve a integração de competências para atingir determinados objetivos predeterminados. A quarta dimensão associa-se ao balanceamento entre o aprofundamento do aprendizado em campos correlatos aos das competências preexistentes e a extensão do esforço de capacitação para novos domínios. Finalmente, uma quinta dimensão envolve a renovação de competências, contornando-se a aversão natural a novas perspectivas devido ao fenômeno da inércia organizacional.

As alternativas de *diversificação* e *especialização* – esta última entendida como a busca da centralização das atividades da empresa em um número restrito de negócios (*core business*) que refletem suas competências essenciais – aparentemente constituiriam movimentos antagônicos no âmbito das estratégias de crescimento das empresas. Assim, enquanto a especialização procuraria focar as capacitações, concentrando o risco, a diversificação buscaria reduzir o risco por meio da expansão para novos negócios com maior ou menor sinergia em relação aos negócios originais da empresa. Na prática, porém, esse aparente antagonismo é muitas vezes substituído por uma complementaridade entre as duas estratégias, na medida em que a consolidação e fortalecimento de determinadas competências essenciais pode se converter em importante fator de estímulo à expansão para novos mercados. De fato, estas competências conferem à empresa vantagens competitivas diferenciais que podem viabilizar a expansão para novos mercados, em função das expectativas de evolução de suas atividades correntes e da rentabilidade potencial de novas atividades.

Um último condicionante interno às empresas do processo de diversificação refere-se especificamente ao formato organizacional e à estrutura de propriedade que as empresas adotam. A questão relativa à coerência do processo de diversificação aparece, de forma bastante nítida, no caso de empresas multidivisionais, sejam elas do tipo multiproduto ou verticalmente integradas.⁶ Essas empresas tendem a apresentar um padrão de diversificação que pode ser caracterizado como concêntrico, induzido por características tecnológicas e/ou mercadológicas das atividades previamente desenvolvidas. Nesse caso, o processo de diversificação visa possibilitar a exploração de economias de escopo e outras sinergias tecnológicas entre atividades, bem como os diversos canais de comercialização disponíveis para a empresa. Para esse tipo de empresa, a direção do processo está fortemente associada à presença de elementos de sinergia entre as suas atividades originais e novas oportunidades potencialmente interessantes. Essa sinergia envolve a identificação de níveis de complementaridade entre as diversas atividades realizadas, de tal maneira que a soma em termos da capacitação da empresa seja superior no nível de capacitação existente nas diversas atividades, se estas fossem tomadas isoladamente. Desse modo, o inter-relacionamento entre as várias atividades é tal que permite a redução do custo de operação e/ou a melhoria da qualidade do produto gerado, quando essas atividades são comparadas com os segmentos de uma empresa diversificada que operam independentemente.

A expansão para novos mercados de empresas que, *grosso modo*, podem ser caracterizadas como conglomerados gerenciais ou como companhias de investimento conduzem sua expansão de forma distinta da anteriormente mencionada. Em geral, seguem uma lógica financeira de distribuição dos riscos. Em particular, as companhias de investimento tendem a orientar seu processo de diversificação segundo a lógica da teoria do portfólio (ver Quadro 15.2).

A ideia básica dessa teoria é que existe um *trade-off* entre lucratividade e risco associado às atividades de uma empresa diversificada. Nesse sentido, duas estratégias alternativas podem ser mencionadas. A primeira delas privilegia a lucratividade e envolve a seleção de atividades altamente rentáveis, mas que estejam forte e negativamente correlacionadas entre si. A segunda estratégia privilegia o menor risco, estando baseada na seleção de atividades cuja correlação de lucratividade tenha uma variância próxima de zero. Utilizando a sua capacidade de mobilizar um grande volume de recursos líquidos, essas empresas podem se beneficiar tanto do efeito associado à redução do risco em seu portfólio de negócios, como obter vantagens resultantes da ampliação deste.

QUADRO 15.2 ANÁLISE DE PORTFÓLIO

A realização de uma análise de portfólio pressupõe que a lógica que coordena a entrada da empresa em novos negócios é essencialmente financeira, resultando naquilo que poderíamos denominar de estratégia de carteira de investimentos. No entanto, essa análise pode ser sofisticada para possibilitar a identificação de atividades que, combinadas, permitam à empresa elevar sua rentabilidade a longo prazo e minimizar os riscos associados à mobilização produtiva dos recursos disponíveis. Nesse sentido, é possível definir uma metodologia que amplia as possibilidades da empresa selecionar adequadamente atividades, viabilizando a obtenção dos objetivos implícitos no projeto de diversificação. Basicamente, essa metodologia contempla os seguintes aspectos:

- A determinação do valor potencial de um novo negócio, quando comparado com outras alternativas de mobilização de recursos. A determinação de um balanceamento entre produtos geradores de recursos líquidos (*positive cash flow*) e aqueles absorvedores desses recursos (*negative cash flow*).
- A determinação de um balanceamento entre o risco atribuído a cada uma das atividades incorporadas ao projeto de diversificação.
- A compatibilização dos diferentes estágios do ciclo de vida dos produtos associados às atividades da empresa diversificada, visando evitar impasses decorrentes do seu comprometimento com estratégias nos diferentes mercados muito disparecidos entre si.
- A representação do portfólio de atividades da empresa em um instrumental analítico (uma matriz bidimensional, por exemplo) que permita identificá-las e agrupá-las, considerando atributos relacionados à posição ocupada pela empresa, ao potencial de crescimento da atividade (em termos de rentabilidade e risco, por exemplo), bem como outros aspectos. O objetivo desse instrumental é possibilitar a identificação de uma combinação de atividades adequada à maximização do retorno a longo prazo da empresa. Como exemplo, é possível citar a “matriz” de atividades utilizada como referência nas análises do Boston Consulting Group (B.C.G.), que segmenta atividades tomando como base a participação relativa da empresa e o potencial de crescimento das vendas nos mercados das diversas atividades. A determinação de um conjunto de parâmetros para avaliar a possibilidade de incorporar novos negócios ao portfólio de atividades.

A diversificação pode combinar duas lógicas – coerência técnico-produtiva e financeira – como é o caso de empresas que fazem parte de grandes grupos econômicos. Nesse caso, não obstante a possibilidade das empresas individualmente procurarem explorar níveis de sinergia no processo de expansão para novos mercados, é provável que essa expansão se subordine à estratégia de crescimento mais geral do grupo a que pertencem. A título de ilustração, a Tabela 15.1 traz uma apresentação dos grandes grupos econômicos brasileiros.

TABELA 15.1 Maiores Grupos Privados Nacionais não financeiros em Patrimônio Líquido e Áreas de Atividade – 1997 – US\$ milhões

Grupos	Patrimônio líquido	Receita operacional líquida	Principais atividades
Vale do Rio Doce	8.506	4.872	Mineração, metalurgia, celulose
Votorantim	5.426	3.367	Celulose e papel, cimento, metalurgia
CSN	3.931	2.390	Siderurgia
Usiminas	3.814	3.123	Siderurgia
Copene	2.515	1.403	Petroquímica
C.R.Almeida	2.469	569	Construção, concessão de serviços públicos
Odebrecht	2.325	4.602	Construção, petroquímica
Andrade Gutierrez	2.317	1.288	Construção, concessão de serviços públicos
Aracruz	2.047	498	Celulose
Camargo Corrêa	2.034	662	Construção, cimento, concessão de serviços públicos
Org. Globo	2.004	4.174	Jornal, TV, telecomunicações
Antarctica	1.662	1.328	Bebidas
Nemofeffer	1.614	1.300	Petroquímica, papel e celulose
COSIPA	1.446	1.223	Siderurgia
Gerdau	1.443	1.837	Siderurgia
Klabin	1.353	1.069	Papel e celulose

Fonte: Gazeta Mercantil, Balanço Anual 1998.

15.6 Condicionantes Externos à Empresa no Processo de Diversificação

Vejamos como as características estruturais dos ambientes competitivos em que as empresas se inserem afetam a diversificação. Dois aspectos são particularmente importantes. O primeiro deles refere-se ao potencial de crescimento do mercado nas atividades originais da empresa, o qual, ao ser confrontado com o potencial de acumulação gerado pela empresa (decorrente de lucros retidos disponíveis para reinvestimento), pode estimular a expansão para novos mercados. O segundo aspecto envolve a maneira como elementos específicos das estruturas de mercado, e os padrões de competição associados predeterminam direções mais factíveis a serem exploradas na diversificação.

O comportamento da demanda nos mercados correntes da empresa estimula o processo de diversificação quando há: (1) tendência à sua retração, associada ao baixo dinamismo tecnológico da atividade, ao acirramento da competição, à reduzida elasticidade-renda ou a outros componentes da teoria do ciclo do produto; (2) intensificação de flutuações cíclicas dessa demanda, que poderiam apontar para o seu esgotamento a longo prazo; (3) crescimento relativamente lento em relação às expectativas de expansão da empresa. No primeiro caso (retração da demanda), a decisão de diversificar precede a identificação de oportunidades atraentes em novos negócios, com a diversificação operando não apenas como determinante do crescimento, mas também como exigência para a própria sobrevivência da empresa. Saliente-se que não necessariamente esse processo ocorre de maneira caótica, uma vez que mudanças duradouras da demanda costumam se manifestar apenas como tendência de longo prazo, o que permite uma certa adaptação da empresa no sentido de criar condições para que o processo seja bem-sucedido. No segundo caso (flutuações cíclicas da demanda), destaca-se a diversificação para atividades que, em conjunto com as originais, permitam uma relativa estabilização do ciclo de negócios da empresa. Finalmente, no terceiro caso (crescimento lento da demanda), a empresa não se encontra tão pressionada para diversificar como nos casos anteriores; no entanto, diante das suas expectativas de expansão, a abertura de novas frentes de atuação constitui uma alternativa interessante. Nesse caso, o principal estímulo do processo é a atratividade de novas oportunidades de negócios, particularmente em razão do maior dinamismo tecnológico e/ou mercadológico destas em relação às atividades originais da empresa.

A dinâmica competitiva não só estimula a expansão para novos mercados, como informa à empresa quais as direções mais factíveis para a diversificação, diante do seu nível de especialização. Desse modo, estabelece-se um vínculo entre o processo de diversificação e os padrões competitivos das diferentes estruturas de mercado, que predeterminam direções para o processo. Assim, cada padrão de concorrência setorial é responsável pela geração de estímulos – relacionados a fatores que atuam como fontes de vantagens competitivas em cada ambiente – que influenciam decisivamente o ritmo e a direção dos processos de diversificação. É importante lembrar, porém, que, como esse padrão de concorrência é essencialmente dinâmico, compreendendo fatores que podem se modificar ao longo do tempo, os estímulos aos processos de diversificação também podem se alterar.

Como exemplo, é possível mencionar transformações ocorridas no padrão de concorrência de diversas indústrias produtoras de *commodities* industriais – siderúrgica e de alumínio, por exemplo – que resultaram em mudanças nos estímulos gerados em termos do processo de diversificação de empresas nelas atuantes – evoluindo-se de uma ênfase na integração vertical “para trás” vinculada ao aprovisionamento de insumos no sentido de uma maior ênfase na diversificação horizontal vinculada ao esforço de enobrecimento e agregação de valor aos produtos gerados. Considerando os estímulos ao processo de diversificação provenientes do padrão de concorrência setorial, observa-se que a não diversificação enfraquece a posição competitiva da empresa, uma vez que os lucros obtidos por seus concorrentes em outros mercados poderiam ser deslocados para seu mercado de origem, ameaçando sua cota de participação e, no limite, eliminando-a do mercado.

Por fim, é interessante considerar uma situação alternativa em que a empresa já apresenta um nível preexistente de diversificação elevado, o que possibilita que sua estratégia de crescimento guarde uma relativa autonomia em relação ao padrão de competição das diferentes indústrias de que ela participa. Nesse caso, o processo de diversificação relaciona-se a uma política de crescimento que extrapola os diferentes mercados em que a empresa atua, assumindo uma dinâmica própria. A dimensão autônoma desse tipo de diversificação relaciona-se à necessidade da empresa planejar uma busca constante de atividades lucrativas em que possa atuar. Quando a diversificação converte-se num determinante crucial da política de crescimento da empresa, deixando de ser um evento pontual, a identificação e a análise de novas oportunidades convertem-se em atividade fundamental, sendo realizada por profissionais altamente especializados. Por outro lado, na medida em que o crescimento é o objetivo a ser alcançado, fortalece-se a possibilidade de entrada em indústrias promissoras do ponto de vista tecnológico ou mercadológico, mesmo que essas não se coadunem perfeitamente com a área de especialização e o padrão de competição das atividades originais da empresa.

15.7 Formas de Diversificação: Investimentos em Nova Capacidade e Operações de Fusões/Aquisições

Uma vez tomada a decisão de diversificar, duas alternativas se colocam para a empresa em termos da viabilização desse processo. A primeira delas envolve a criação de uma capacidade de produção totalmente nova, expressa em uma nova unidade produtiva. A esse tipo de estratégia, é comum correlacionar-se a noção de crescimento interno. A segunda alternativa compreende a aquisição ou a fusão com uma empresa já atuante no mercado objeto da diversificação. Essa empresa incorpora a nova unidade a sua estrutura organizacional, utilizando-a para incrementar seu nível de especialização e, consequentemente, para alargar o horizonte de diversificação que possibilita dar continuidade ao seu crescimento. As estratégias baseadas em aquisições ou fusões são geralmente associadas à noção de crescimento externo.

O crescimento externo se diferencia do interno na medida em que, enquanto o último resulta num aumento da capacidade produtiva não só da empresa, mas também da indústria em questão e de outras atividades a ela articuladas, o primeiro só altera a capacidade da própria empresa que adotou a estratégia, funcionando, no nível da indústria, como mera transferência de propriedade. O elemento fundamental que caracteriza o crescimento externo é, portanto, a aquisição de unidades já existentes que não deixam de funcionar; a eventual reorganização desses ativos pela empresa adquirente não altera essa característica, mesmo porque essa reorganização também é possível no caso de ativos criados por meio de crescimento interno.

As fusões e as aquisições relacionadas ao processo de crescimento externo constituem um importante determinante da intensidade e direção do processo de diversificação. Análises baseadas em levantamentos de informações sobre movimentos de fusões e aquisições de empresas geralmente demonstram que uma parcela significativa dessas operações pode ser classificada como diversificação. Além disso, levantamentos sistemáticos realizados sobre “maiores e melhores” empresas, como o da *Fortune 500*, demonstram que as novas linhas de negócios que emergem nessas empresas estão geralmente associadas a este tipo de operação, indicando que a maior parte das tentativas de diversificação ocorre por intermédio de processos de aquisição de outras empresas. O Quadro 15.3 aborda esse aspecto, tomando como referência a experiência recente da economia brasileira.

A grande vantagem das fusões e das aquisições como estratégia de expansão para novos mercados reside na possibilidade de reduzir ou eliminar os riscos tecnológicos e de mercado implícitos na diversificação. O esforço para a integração organizacional entre as duas empresas pode ser minimizado, na medida em que a empresa adquirida conserve sua autonomia, funcionando como “centro de lucros”, com os vínculos se restringindo à adequação de sua conduta aos objetivos mais gerais da empresa adquirente. No caso em que a empresa adquirida não apresenta uma rentabilidade satisfatória, a aquisição pode ser viabilizada se a empresa adquirente considera a possibilidade de correção desses problemas por meio, por exemplo, da introdução de práticas de gestão mais modernas. Além disso, do ponto de vista da empresa adquirente, esse tipo de operação possui uma série de vantagens, tais como: (1) diminuição de exigências em termos de desembolsos e de dificuldades diretivas e técnicas;

QUADRO 15.3 FUSÕES E AQUISIÇÕES NA ECONOMIA BRASILEIRA

A ocorrência de importantes mudanças no ambiente institucional interno e externo tem estimulado a aceleração do processo de fusões e aquisições na economia brasileira. O processo de globalização, na medida em que ampliou os mercados por meio de quedas das barreiras aos fluxos de bens, serviços e capitais, alterou também o ambiente institucional em que as empresas estavam acostumadas a operar, induzindo ao acirramento da concorrência. Nesse sentido, a evolução das fusões e aquisições deve ser entendida como resposta estratégica das corporações ao surgimento desse novo ambiente institucional e concorrencial proporcionado pelo processo de globalização. Além disso, o processo tem sido estimulado pela possibilidade que representa em termos da penetração em novos mercados, em curto período de tempo e da exploração de novas oportunidades de investimento, em função da privatização de empresas estatais e da desregulamentação dos mercados, bem como pela possibilidade de obtenção de sinergias de natureza tecnológica, financeira, mercadológica e organizacional. As principais consequências dessa evolução são o aparecimento de megacorporações, a concentração produtiva e a elevação das escalas mínimas de produção em diversos setores.

Dante das tendências mencionadas, a realização de fusões e aquisições vem assumindo crescente importância na economia brasileira. Entre 1990 e 1999, 53% das F&A realizadas no Brasil se concentraram em setores de infraestrutura, destacando-se a privatização e a reestruturação do setor de telecomunicações (ver Tabela 15.2). Por outro lado, extração mineral e petróleo tiveram uma participação bem menos pronunciada (5%), com destaque para a privatização da Companhia Vale do Rio Doce (CVRD). No início da década, o setor mais proeminente foi o de bens intermediários, consequência direta da privatização das empresas siderúrgicas e petroquímicas. Contudo, na segunda metade da década, o maior destaque são as F&A relacionadas à privatização da infraestrutura. Nesse período destacam-se também operações de F&A relacionadas ao setor financeiro. A realização de F&A acentuou em muito o processo de internacionalização patrimonial da economia brasileira. A participação estrangeira nas operações de F&A no Brasil alcançou 55% do valor total de transações no período, sendo Europa (33%) e EUA e Canadá (18%) os maiores compradores. Se considerarmos o número de transações, a internacionalização é de 64%, sendo as maiores participações as [de] EUA e Canadá (28%) e Europa (26%).

TABELA 15.2 Número e Valor das Transações de F&A no Brasil por Setor de Atividade da Empresa Adquirida, 1990-1999

	Valor (US\$ milhões)		Número de Transações	
	Valor	%	Número	%
Telecomunicações	36.925,3	31,5	79	6,9
Energia Elétrica, Gás e Água	25.746,8	21,9	57	5,0
Financeiro	14.561,9	12,4	210	18,3
Metalurgia e Siderurgia	6.837,6	5,8	85	7,4
Alimentar	4.254,6	3,6	88	7,7
Químico	4.109,1	3,5	74	6,4
Mineração	4.067,9	3,5	23	2,0
Comércio Varejista	3.040,5	2,6	61	5,3
Serviços de Transporte	2.764,1	2,4	46	4,0
Minerais Não Metálicos	2.481,5	2,1	27	2,4
Serviços Diversos	2.153,4	1,8	32	2,8
Petroquímico	1.468,0	1,3	21	1,8
Farmácia, Higiene e Limpeza	1.407,6	1,2	22	1,9
Papel e Celulose	1.373,6	1,2	23	2,0
Maquinaria	1.115,4	1,0	44	3,8
Comércio Atacadista	1.024,8	0,9	39	3,4
Automobilística e Autopeças	852,9	0,7	54	4,7
Produtos em Borracha e Plástico	569,2	0,5	18	1,6
Têxtil	561,0	0,5	15	1,3
Agricultura e Pesca	520,4	0,4	20	1,7
Eletroeletrônico	488,8	0,4	27	2,4
Editorial e Gráfico e Publicidade	200,9	0,2	33	2,9
Construção Civil	171,6	0,1	5	0,4
<i>Holdings</i>	157,0	0,1	1	0,1
Serviços de Engenharia	136,2	0,1	6	0,5
Móveis e Produtos em Madeira	113,3	0,1	4	0,3
Aeroespacial e Aviação	96,7	0,1	4	0,3
Informática	64,6	0,1	13	1,1
Outros Manufaturados	59,7	0,1	15	1,3
Fumo	50,0	0,0	2	0,2
Couro e Artigos de Couro	ND	ND	1	0,1
TOTAL	117.374,4	100,0	1.148	100,0

Fonte: Ferraz et alii (2000) a partir da Thomson Financial Securities Data.

(2) obtenção de uma posição favorável no mercado objeto da diversificação, reduzindo substancialmente a pressão do processo competitivo associada à entrada em uma nova indústria; (3) incorporação de um parque produtivo já dimensionado e de recursos já adaptados à utilização em que são empregados.

As evidências levantadas pela literatura sobre fusões e aquisições sugerem que a escassez de recursos técnico-produtivos internos para a expansão é um dos principais fatores propulsores dessas operações. Nesse sentido, se, por um lado, a diversificação parece ser explicada pelo excesso de recursos financeiros, por outro, as fusões e as aquisições, que são instrumentos fundamentais do processo de diversificação, muitas vezes constituem respostas a problemas de escassez de recursos técnico-produtivos por

parte das empresas. É importante considerar, porém, que a realização de aquisições para viabilizar a expansão para novos mercados pressupõe certos tipos de “qualidades empreendedoras” da empresa adquirente, bem como envolve custos específicos que devem ser avaliados estrategicamente por ela. Desse modo, uma empresa especializada dificilmente se aventuraria em um novo mercado em relação ao qual ela não possuísse um mínimo de capacitação potencialmente útil, mesmo que restrita à esfera gerencial. Uma exceção importante em relação a esta tendência seria representada por aquisições de empresas extremamente dinâmicas do ponto de vista tecnológico ou mercadológico.

Os custos relacionados ao processo de aquisição estão diretamente associados às avaliações sobre possíveis desdobramentos do processo em termos da competitividade das empresas, realizadas, respectivamente, pela empresa adquirente e pela adquirida. Apesar da influência de outros fatores (inclusive aqueles associados a facilidades fiscais), a tendência é que a avaliação realizada pela adquirente seja positivamente afetada pelo fato da produção da adquirida, de alguma forma, complementar suas próprias atividades. As evidências demonstram que, de fato, grande parte das fusões e aquisições é motivada pela possibilidade de acessar recursos complementares entre as empresas. Nesse sentido, diversas análises ressaltam a importância do acesso a ativos e competências complementares na realização de movimentos de fusões e aquisições, particularmente por meio da viabilização de processos de “fertilização cruzada” entre as competências dos agentes, que são fundamentais para o reforço da competitividade em um número cada vez maior de áreas. Segundo essas análises, observa-se uma tendência a que os campos de especialização de grandes empresas permaneçam relativamente estáveis ao longo do tempo, quando são considerados possíveis desdobramentos de processos de fusões e aquisições.

As operações de fusões e aquisições entre empresas realizadas com o propósito de viabilizar processos de diversificação têm impactos diretos sobre o poder de mercado das empresas.⁷ No nível setorial, a realização desse tipo de operação resulta em tendência à elevação do grau de concentração nas respectivas indústrias, não só devido aos possíveis impactos diretos em termos da integração de ativos, mas também em razão dos efeitos indiretos sobre as condições concretas da concorrência naqueles mercados, a qual tende a se acirrar, impondo uma revisão das estratégias das demais empresas, o que pode representar um estímulo adicional à continuidade daquelas operações. Além disso, a realização de operações de fusões e aquisições de caráter “diversificante” pode resultar em rearranjo das operações e dos ativos das empresas envolvidas nos respectivos mercados de atuação, com o intuito de explorar mais eficazmente o potencial daqueles mercados e o perfil de competências das empresas, incrementando sua competitividade e reforçando o seu poder de mercado. Por fim, é importante considerar possíveis impactos desse tipo de operação sobre as condições das barreiras à entrada nas indústrias associadas. Quanto a esses impactos, dois aspectos devem ser considerados. Por um lado, as operações de fusões e aquisições de caráter “diversificante” permitem que as empresas contornem eventuais barreiras à entrada nos mercados que são objeto de seu interesse, na medida em que passam a dispor de uma fatia do mercado previamente ocupada. Por outro lado, uma vez consumado esse tipo operação, as empresas envolvidas podem empreender um esforço calcado na recomposição de ativos e na adoção de diversas práticas concorrentiais (subsídios cruzados entre atividades, por exemplo), que resulta não só na desestabilização da concorrência naqueles mercados, mas também na elevação das barreiras à entrada a eles associadas, na medida em que impõe novos requisitos em termos do porte empresarial necessário para neles atuar.

Notas

1. O primeiro conceito refere-se, basicamente, às características dos processos de produção e às tecnologias empregadas, ao passo que o segundo refere-se às características dos processos de comercialização e vendas, bem como à diversidade de redes de comercialização e à base de clientes atingidos. A distinção entre esses conceitos é utilizada por Edith Penrose para definir os conceitos de “base tecnológica” e “área de comercialização” que delimitam as possibilidades de diversificação para uma empresa particular, conforme discutido na Seção 15.4.
2. É comum, na literatura de Economia Industrial, a caracterização dessas alternativas utilizando a nomenclatura francesa, que distingue, respectivamente, processos de integração “a montante” e “a jusante”.
3. Como exemplo, é possível citar o caso de indústrias em rede, geralmente associadas a setores de infraestrutura, nas quais se observa um padrão de interconexão e compatibilidade entre unidades produtivas presentes em diferentes estágios de determinada cadeia de produção e distribuição (caso dos setores de eletricidade e telefonia, por exemplo), que se constitui em requisito básico para a obtenção de níveis satisfatórios de eficiência para as empresas. Nesses casos, observa-se também a presença de externalidades de demanda que também favorecem a existência de um grau elevado de integração de atividades produtivas.
4. Este aspecto envolve, por exemplo, uma comparação entre o consumo cativo de determinado insumo e a escala mínima para produzi-lo de forma eficiente.
5. Ver, a esse respeito, a discussão desenvolvida no Capítulo 16 sobre os fatores que estimulam a consolidação desse tipo de arranjo e sobre as características de sua estrutura interna.
6. Ver, para a caracterização desse tipo de empresa, a análise desenvolvida no Capítulo 2.
7. É importante distinguir, nesse sentido, impactos relacionados a operações de fusões e aquisições entre firmas de uma mesma indústria, caso em que esses impactos são mais diretamente perceptíveis, de situações em que essas operações são realizadas no intuito de viabilizar processos de diversificação das empresas envolvidas para novos mercados, que requerem uma discussão mais cuidadosa.

Bibliografia

- Alchian A.; Woodward, S. (1988) "The firm is dead: long live to the firm". *Journal of Economic Literature*, v. XXVI, p. 65-79.
- Britto, J. N. P. (1993) "O processo de diversificação da firma: uma abordagem dinâmica exploratória". *Nova Economia*, v. 3, n. 1, p.195-224.
- Bonelli, R. (1998) "As estratégias dos grandes grupos industriais brasileiros nos anos 90", Texto para Discussão n. 569, IPEA.
- Chandler, A. (1990) *Scale and Scope*. Harvard University Press, Cambridge (MA).
- Chandler, A. D. (1992) "What is a firm? A historical perspective", *European Economic Review*, v. 36, p. 483-494, North-Holland.
- Coase, R. H. (1937) "The nature of the firm", *Economica*, v. 4, p. 386-405.
- Doz, Y. (1996) "Managing core competency for corporate renewal: towards a managerial theory of core competencies", in: Dosi, G.; Malerba, F. (orgs.) *Organisation and Strategy in the Evolution of Enterprise*, Elsevier Publishers.
- Ferraz, J. C.; Rocha, F; Lootty, M. (2000) "Fusões e Aquisições: estratégias de crescimento, investimento direto e inovação", mimeo, IE-UFRJ.
- George, K. D. (1972) "Diversification and aggregate concentration", *Journal of Industrial Economics*, v. 20; reimpresso em BASIL, S. Y. (org.) *Economics of Industrial Structure*. Londres: Penguin Modern Economic Readings, 1973.
- Granstrand, O.; Patel, P.; Pavitt, K. (1997) *Multi-technology corporations. why they have "distributed" rather than "distinctive core" competencies*. Mimeo, Brighton, Março.
- Marris, M. (1971) "The modern corporation and economic theory", in: Marris, R.; Wood, A. (org.) *The corporate economy: growth, competition and economic potential*, Londres, MacMillan.
- Nelson, R. (1991) "Why firms differ, and how does it matter?", *Strategic Management Journal*, v. 12, p. 61-74.
- Penrose, E. (1959) *The Theory of the Firm*. Oxford, Oxford University Press.
- Prahlad, C. K.; Hamel, G. (1990) "The core competence of the corporation", *Harvard Business Review*, maio-junho.
- Reed, R.; Luffman, G. (1986) "Diversification: the growing confusion", *Strategic Management Journal*, v. 7.
- Richardson, G. B. (1972) "The organization of industry", *Economic Journal*, v. 82, p. 883-896, set.
- Scherer, F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*. Boston, Houghton Mifflin Company.
- Siffert, N. F. (1998) "Governança corporativa: padrões internacionais e evidências empíricas no Brasil nos anos 90". *Revista do BNDES*, n. 9. Rio de Janeiro: BNDES, junho.
- Siqueira, T. V. (2000) "Os grandes grupos brasileiros: desempenho e estratégias na primeira metade dos anos 90". Rio de Janeiro, Revista do BNDES, v. 7, n. 13, p. 3-32, jun.
- Steindl, J. (1986) *Maturidade e estagnação no capitalismo americano*. Coleção Os Economistas, Nova Cultural, São Paulo.
- Teece, D.; Pisano, G. (1994) "The dynamic capabilities of firms: an introduction", *Industrial Corporate Change*, v. 1, n. 3.
- Teece, D.; Rumelt, R.; Dosi, G.; Winter, S. (1994) Understanding Corporate Coherence: Theory and Evidence. *Journal of Economic Behavior and Organization* v. 23, p. 1-30.
- Weston, J. F. (1970) "The nature and significance of conglomerate firms", *St. John's Law Review*, v. 44.; também impresso em YAMWY, B. S. (org.) "Economics of Industrial Structure", Londres, Penguin Books, 1973.
- Williamson, O. E. (1971) "Managerial discretion, organisation forms and the multi-division hypothesis", in: Marris, R.; Wood, A. (org.) *The corporate economy: growth, competition and economic potential*, Londres, MacMillan.
- Williamson, O. E. (1981) "The modern corporation: origins, evolution, attributes", *Journal of Economics Literature*, dez.
- Williamson, O. E. (1985) "The Economics Institutions of Capitalism: Firms, Markets, Relational Contracting", Nova York, Free Press.
- Williamson, O. E.; Winter, S. (orgs.) (1991) *The nature of the firm: origins, evolution and development*, Oxford, Oxford University Press.

Cooperação Interindustrial e Redes de Empresas

Jorge Britto

16.1 Introdução

A ocorrência de múltiplas formas de cooperação produtiva e tecnológica entre empresas é um tema que tem sido abordado de maneira recorrente pela literatura de Economia Industrial. Em consequência, observa-se uma crescente convergência entre visões de diferentes escolas de pensamento de que a análise dos fatores subjacentes a um melhor desempenho competitivo deve centrar-se não apenas na empresa individual, mas principalmente na investigação das relações entre as empresas e entre estas e as demais instituições. A amplitude e a complexidade das interdependências entre empresas e outras organizações ou instituições tem sido estudada, cada vez mais, por meio da utilização de um recorte analítico baseado no conceito genérico de rede. A utilização desse conceito como referencial analítico, de forma mais ou menos explícita, tem auxiliado a investigação de temas bastante diversos, tais como:

1. Alianças estratégicas entre empresas e outras formas de cooperação produtiva e tecnológica.
2. Programas de cooperação específicos, envolvendo agentes com competências em áreas distintas, que interagem entre si para viabilizar determinada inovação.
3. Processos de subcontratação e terceirização realizados por empresas especializadas em determinadas atividades, que dariam origem a redes estruturadas verticalmente no interior de cadeias produtivas.
4. Sistemas flexíveis de produção baseados em relações estáveis e cooperativas entre empresas atuantes em determinado ramo de atividades.
5. Distritos industriais baseados na aglomeração espacial de empresas e outras instituições que interagem entre si no âmbito de determinada região.
6. Sistemas nacionais e regionais de inovação baseados na especialização e na interação de diversos tipos de agentes envolvidos com a realização de atividades inovativas (empresas, universidades, outras instituições etc.).

A relevância deste conceito de estruturas em rede decorre da sua capacidade em captar a crescente sofisticação das relações interindustriais que caracteriza a dinâmica econômica contemporânea. A partir do início da década de 1990, diversas tendências relacionadas no padrão evolutivo das principais economias capitalistas reforçam a relevância dessa temática, tais como:

1. A consolidação de um paradigma organizacional baseado na experiência de empresas japonesas, incorporando novos princípios gerenciais que enfatizam a cooperação interindustrial nas articulações entre produtores e fornecedores.
2. A estruturação de sistemas produtivos que incorporam o conceito de especialização flexível como princípio organizador das atividades.

3. A intensificação da concorrência e a globalização dos mercados, que resulta em estímulos à montagem de alianças estratégicas com múltiplos formatos entre empresas.
4. A consolidação de um paradigma tecnológico baseado em novas tecnologias de informação e telecomunicação que facilitam a interação entre agentes.
5. A evolução no sentido de uma nova sistemática de realização de atividades inovativas, crescentemente baseadas na aglomeração de múltiplas competências e em projetos cooperativos de caráter interdisciplinar.
6. A mudança de enfoque da política industrial implementada em diversos países, no sentido de privilegiar-se o apoio a redes envolvendo diversas empresas, em contraposição ao apoio a empresas isoladas.

O interesse que o conceito de estruturas em rede vem despertando na literatura decorre, em boa medida, da sua maleabilidade. De fato, as estruturas em rede constituem um quadro de referência que pode ser aplicável à investigação de múltiplos fenômenos caracterizados pela densidade de relacionamentos cooperativos entre os agentes, o que reforça a interdependência entre suas respectivas competências e impõe a necessidade de algum tipo de coordenação coletiva das ações adotadas. Nesse sentido, observa-se também uma certa confusão semântica entre os conceitos de “empresas em rede”, “redes de empresas” e “indústrias em rede”. O primeiro deles associa-se a conformações intraorganizacionais que se estruturam como desdobramento evolutivo da empresa multidivisional, a partir do advento de novas tecnologias de informação-telecomunicação. As “indústrias em rede” estão geralmente associadas a setores de infraestrutura, baseando-se em padrão de interconexão e compatibilidade entre unidades produtivas, que se constitui em requisito básico para a operação eficaz destas. O conceito de “redes de empresas”, por sua vez, refere-se a arranjos interorganizacionais baseados em vínculos sistemáticos – muitas vezes de caráter cooperativo – entre empresas formalmente independentes, que dão origem a uma forma particular de coordenação das atividades econômicas.

Este capítulo aborda um tipo particular de estrutura em rede, as redes de empresas, procurando discutir algumas características e propriedades desses arranjos que condicionam a sua capacidade de resposta diante dos estímulos ambientais. O capítulo está dividido em quatro seções, além desta introdução. A primeira seção procura discutir a importância do conceito genérico de rede dentro da teoria econômica, ressaltando-se, em especial, a importância do conceito de redes de empresas como objeto específico de investigação. A segunda seção apresenta uma sistematização dos principais elementos estruturais constituintes das redes de empresas. A terceira seção procura construir um quadro analítico de referência no tocante aos processos internos e às propriedades dessas redes. A quarta seção apresenta uma discussão sobre as características de alguns modelos estilizados dessas redes, elaborada com base no quadro analítico anteriormente construído.

16.2 O Conceito de Rede na Ciência Econômica e a Noção de Redes de Empresas

A utilização do conceito de rede como artifício analítico na abordagem de problemas econômicos reflete não apenas a recuperação de temas tradicionalmente abordados pela economia política clássica – discutindo a especificidade da divisão social do trabalho entre as empresas – como também a incorporação de uma série de contribuições importantes da sociologia e da matemática, evidenciando uma abordagem nitidamente interdisciplinar. No âmbito das ciências exatas, o conceito de rede tem motivado o desenvolvimento de um instrumental sofisticado aplicável à caracterização e ao estudo da estrutura de sistemas complexos e dinâmicos. No caso das ciências sociais, a utilização deste tipo de recorte enfatiza a importância de se entender a estrutura do sistema de relações que conectam diferentes agentes, bem como os mecanismos de operação deste sistema, responsáveis pela sua reprodução, fortalecimento e eventual transformação ao longo do tempo. Análises baseadas no conceito de rede pressupõem que a configuração dos vínculos presentes e ausentes entre os pontos que conformam determinado sistema revelam estruturas específicas, que constituem um objeto relevante de investigação. Quando se transpõe esse conceito para o estudo de problemas atinentes à ciência econômica, uma primeira questão relevante diz respeito à identificação de situações nas quais este conceito poderia – ou deveria – ser utilizado.

Na utilização do conceito genérico de rede pela teoria econômica, é possível diferenciar duas abordagens distintas. A primeira delas ressalta o caráter instrumental do conceito de rede para a compreensão da dinâmica de comportamento dos diferentes mercados. Neste caso, o conceito é utilizado no tratamento de problemas de natureza alocativa recorrentemente enfrentados pela ciência econômica, estando relacionado à noção de externalidades em rede como princípio orientador da análise. A presença de externalidades em rede em determinados mercados reflete a existência de efeitos diretos e indiretos da interdependência entre as decisões de agentes que neles atuam. Esta perspectiva de análise opta por um recorte nitidamente microeconômico, procurando entender como determinada rede de relações afeta as decisões tomadas pelos agentes econômicos fundamentais (produtores e consumidores) em mercados particulares. Usualmente, a literatura sobre o fenômeno distingue os seguintes tipos de externalidades em rede:

1. Externalidades técnicas relacionadas a situações em que a interdependência entre os agentes do ponto de vista técnico resulta em modificações nas características das respectivas funções de produção.
2. Externalidades pecuniárias que se traduzem em uma mudança nos preços relativos dos fatores e em modificações das estruturas de custo das empresas.
3. Externalidades tecnológicas associadas a efeitos do tipo *spill-over* que resultam em mudanças no ritmo de adoção e difusão e inovações em determinado mercado.
4. Externalidades de demanda presentes em situações em que a demanda de bens oferecidos por unidade é afetada por modificações na demanda de outras unidades ou em que a demanda de um consumidor individual é influenciada pela demanda agregada do mesmo bem.

A análise tradicional das externalidades em rede procura modelar e discutir o fenômeno relativo ao surgimento de rendimentos crescentes no interior de mercados nos quais algumas das características mencionadas estão presentes. Este fenômeno mostra-se particularmente importante no caso de um elenco variado de indústrias, caracterizadas por um grau elevado de integração e interdependência entre unidades produtivas, as denominadas indústrias em rede. Supõe-se, nesse sentido, que as mencionadas externalidades funcionam como fator de fortalecimento das interdependências entre unidades produtivas inseridas nessas indústrias, possibilitando a caracterização dos mercados respectivos como uma rede de agentes interdependentes. Como principais características dessas estruturas em rede, é possível mencionar:

1. A presença de um grau elevado de compatibilidade e complementaridade técnica entre os agentes e as atividades por eles realizadas.
2. A existência de um grau elevado de integração de atividades produtivas no nível da rede, devido à presença de externalidades técnicas, pecuniárias e de demanda.
3. A geração de externalidades tecnológicas e outros tipos de ganhos relacionados ao progresso técnico, devido à variedade de empresas inseridas nesses arranjos e à complementaridade entre as respectivas competências.
4. A consolidação de uma infraestrutura particular que conforma tais sistemas, a qual implica um certo grau de irreversibilidade quanto a investimentos realizados por agentes que a eles se integram.

Nesta perspectiva, a rede de empresas pode ser referenciada a um conjunto organizado de unidades de produção parcialmente separáveis que operam com rendimentos crescentes que podem ser atribuídos tanto a economias de escala como a uma função global de custos “subaditivos” que refletem a presença de externalidades significativas de natureza técnica, pecuniária e tecnológica, assim como efeitos relacionados a importantes externalidades de demanda. Esta definição salienta a importância dos conceitos de interdependência e de compatibilidade, no que se refere às esferas da produção e da demanda nos mercados em que se inserem as estruturas em rede. Considerando esses aspectos, a discussão sobre como operam estas redes é referenciada à maneira como surgem estes diversos tipos de externalidades nos mercados respectivos.

Em contraste com análises que ressaltam o papel das externalidades em rede sobre a dinâmica alocativa de diferentes mercados, é possível caracterizar um outro tipo de abordagem que discute o conceito de rede menos a partir dos possíveis efeitos gerados sobre o comportamento de consumidores e produtores, e mais do ponto de vista da constituição de um tipo particular de instituição, com a capacidade de coordenar a realização de atividades econômicas. Nesse caso, a ênfase recai na caracterização das estruturas em rede como um *objeto específico* de investigação. Estas estruturas estariam associadas a determinados elementos básicos constituintes, bem como a mecanismos de operação particulares, responsáveis pela geração de estímulos endógenos indutores de processos adaptativos face à evolução do ambiente.

Comparando-se este enfoque com a perspectiva anteriormente mencionada, que privilegia a discussão dos efeitos de externalidades em rede sobre os mecanismos de ajustamento dos diferentes mercados, duas diferenças básicas podem ser destacadas. Por um lado, a ênfase da análise recai nos processos de estruturação e transformação destas redes a partir de estímulos internos e externos, e não apenas no impacto que a formação destas estruturas acarreta sobre a dinâmica alocativa dos diferentes mercados. Por outro lado, considerando estas redes como objeto específico de investigação, os processos alocativos que ocorrem em seu interior passam a ser concebidos como uma faceta particular dos mecanismos de operação destas estruturas, tornando-se necessária a discussão de outras dimensões associadas a estes mecanismos.

Na literatura de Economia Industrial, a identificação das redes de empresas como objeto específico de investigação vem assumindo crescente importância. A análise dessas redes é desenvolvida a partir de uma crítica à divisão artificial entre a empresa e o ambiente externo inserem que está inserida. Pressupõe, portanto, que este ambiente se encontra institucionalmente estruturado em função da densidade de vínculos produtivos e tecnológicos estabelecidos entre as empresas e outras instituições. O conceito de redes de empresas baseia-se, assim, em uma perspectiva de análise que ressalta a dimensão social das relações entre empresas e seus possíveis desdobramentos sobre a conformação institucional do ambiente econômico e sobre o padrão de

conduta dos agentes. No plano metodológico, as redes de empresas são, por excelência, um objeto enfocado por análises que privilegiam um recorte mesoeconômico da dinâmica industrial, as quais ressaltam o papel desempenhado por subsistemas estruturados na modulação daquela dinâmica. Desse modo, estas redes caracterizam-se pela existência de uma autonomia relativa em relação às forças externas, bem como a presença de um certo grau de auto-organização e de uma capacidade endógena de transformação, que lhes confere um caráter essencialmente dinâmico.

É possível avançar no sentido de algumas implicações da utilização do conceito de redes de empresas. Em particular, três procedimentos gerais de operacionalização de estudos empíricos baseados nesse conceito podem ser mencionados. Em primeiro lugar, é importante ter claro qual a estrutura da rede a ser considerada na investigação, o que envolve uma série de questões importantes, tais como: a definição de critérios para o agrupamento dos seus elementos constituintes e das ligações a eles associadas; a demarcação dos limites da estrutura; a caracterização das forças endógenas que emergem da consolidação da rede etc. Em segundo lugar, observa-se que, quando se evoluí do plano das empresas individuais para o plano mesoeconômico dos arranjos interorganizacionais, a discussão deve contemplar não apenas as estratégias individualmente definidas pelas empresas, como também diversas práticas socialmente construídas que permitem a continuidade e o aprofundamento da cooperação entre as partes envolvidas no arranjo, proporcionando algum tipo de benefício econômico para estas. Em terceiro lugar, é importante considerar na análise possíveis desdobramentos da consolidação daqueles arranjos em termos da geração de vantagens competitivas diferenciais para as empresas neles integradas.

16.3 Redes de Empresas: Elementos Estruturais

A utilização do conceito de rede como artifício analítico na compreensão de múltiplos fenômenos pode ser correlacionada a alguns elementos morfológicos que são comuns a este tipo de estrutura. Especificamente, quatro elementos morfológicos genéricos – nós, posições, ligações e fluxos – podem ser ressaltados como partes constituintes das estruturas em rede. No caso específico das redes de empresas estes elementos básicos assumem características particulares. O Quadro 16.1 procura sintetizar estas características, associando a cada um dos elementos morfológicos genéricos constituintes das estruturas em rede a expressão destes no âmbito específico das redes de empresas.

Em primeiro lugar, é possível definir um conjunto de agentes, objetos ou eventos em relação aos quais a rede estará definida. Na caracterização morfológica de uma rede, este conjunto associa-se ao conceito de pontos focais ou nós que compõem a estrutura. Na caracterização dos nós que constituem as unidades básicas das redes de empresas, duas perspectivas distintas de análise podem ser ressaltadas. A primeira identifica as empresas inseridas nestes arranjos como unidades básicas a serem investigadas. Nessa perspectiva, estas redes são concebidas como o produto das estratégias adotadas pelos agentes nelas inseridos, que induzem o estabelecimento de relacionamentos sistemáticos entre eles. Partindo-se das empresas como nós fundamentais das redes, torna-se possível captar a conformação da estrutura a partir da análise das estratégias de relacionamentos dessas empresas, as quais se refletem na formação de alianças estratégicas com outros agentes (ver Quadro 16.2). Às análises que elegem as empresas como unidades básicas dos arranjos estruturados na forma de redes, é possível contrapor um outro tipo de enfoque que caracteriza determinadas atividades como pontos focais daqueles arranjos. Em relação à perspectiva anteriormente mencionada, uma diferença básica associa-se à unidade básica considerada no levantamento de informações empíricas. Enquanto no caso anterior esta unidade estaria associada a determinado agente e seus relacionamentos externos, neste último caso esta unidade estaria referenciada a uma determinada atividade produtiva ou a uma determinada indústria. Assim, particular relevância costuma ser atribuída aos fatores que explicam a aproximação-integração de diferentes atividades produtivas no interior de uma estrutura em rede.

O detalhamento morfológico das estruturas em rede pressupõe também a identificação das posições que definem como os diferentes pontos se localizam no interior da estrutura. Estas posições estão associadas a uma determinada divisão de trabalho

QUADRO 16.1 ELEMENTOS ESTRUTURAIS DAS REDES DE EMPRESAS

Elementos Morfológicos Gerais das Redes	Elementos Constitutivos das Redes de Empresas
Nós	Empresas ou Atividades
Posições	Estrutura de Divisão de Trabalho
Ligações	Relacionamentos entre Empresas (aspectos qualitativos)
Fluxos	Fluxos de Bens (tangíveis) e de Informações (intangíveis)

QUADRO 16.2 COOPERAÇÃO ENTRE EMPRESAS E A FORMAÇÃO DE ALIANÇAS ESTRATÉGICAS

A consolidação de estruturas em rede conectando diferentes empresas muitas vezes surge a partir da formação de “alianças estratégicas” pontuais entre elas. Estas alianças envolvem acordos formais e informais entre empresas que permitem um intercâmbio de informações e uma aglutinação de competências, associando-se à estruturação de arranjos cooperativos – em geral de caráter pré-competitivo – que permitem aos agentes explorar oportunidades tecnológicas e mercadológicas promissoras. O conceito de “alianças estratégicas” é suficientemente amplo para englobar múltiplos formatos institucionais, tanto do ponto de vista dos arranjos contratuais associados como do ponto de vista da estrutura societária subjacente ao arranjo.

A identificação dos motivos para formação dessas alianças tem sido abordada por uma vasta literatura. Utilizando informações coletadas em bancos de dados sobre cooperação tecnológica entre empresas, essas análises demonstram que os motivos relacionados a critérios estáticos de eficiência – em particular, a minimização dos custos associados a determinadas funções produtivas – não explicam o grande crescimento no número de “alianças estratégicas” entre empresas observado a partir da década de 1980. Ao contrário, as informações coletadas corroboram a tese de que estas alianças têm sido montadas a partir de um “posicionamento estratégico” dos agentes, que contempla a exploração de oportunidades e a adequação às tendências definidas no âmbito de um novo “paradigma tecnológico”, fortemente baseado nas denominadas “tecnologias de informação”. Além disso, a montagem dessas alianças tem sido correlacionada à crescente complexidade do processo de P&D, que requer cada vez mais a aglutinação de competências e a intensificação do intercâmbio de informações entre agentes envolvidos num esforço tecnológico e/ou mercadológico comum.

Alguns autores atribuem particular importância à análise de “alianças estratégicas” do ponto de vista da lógica de integração de competências no âmbito destes. Observa-se, nesse sentido, que a montagem dessas alianças tende a privilegiar, numa primeira etapa, arranjos pré-competitivos que facilitam a introdução de inovações no mercado e, numa etapa posterior, a montagem de relacionamentos estáveis entre empresas, capazes de impulsionar o aumento da eficiência operacional. Basicamente, estas alianças compreendem um “balanceamento” entre princípios de cooperação e competição, que pode assumir formas institucionais distintas. Dentre essas formas, três poderiam ser destacadas: (1) alianças baseadas numa *“integração conjunta”* de atividades, baseadas numa evolução no sentido de um estágio mais avançado da cadeia de produção e comercialização de bens, que não pode ser atingido em condições favoráveis pelos membros participantes do arranjo tomados isoladamente; (2) alianças baseadas numa *“configuração aditiva”*, que articulam duas ou mais empresas de uma indústria particular para viabilizar um aumento de escala, um alargamento do mercado “interno” às empresas e um amortecimento da concorrência entre os membros participantes; (3) alianças baseadas numa *“configuração complementar”*, que integram duas ou mais empresas para permitir uma aglutinação de ativos e competências complementares controlados por cada um de seus membros, o que amplia a competitividade destes.

que conecta os diferentes agentes visando atingir determinados objetivos. A consolidação desta divisão de trabalho é uma consequência natural da diversidade de atividades necessárias à produção de determinado bem, envolvendo a integração de capacidades operacionais e competências organizacionais dos agentes, bem como a compatibilização-integração das tecnologias incorporadas nos diferentes estágios das cadeias produtivas.

É possível também associar as estruturas em rede a determinadas ligações entre seus nós constituintes. Em função da estrutura destas ligações, é possível distinguir estruturas dispersas – nas quais o número de ligações entre pontos é bastante limitado – de estruturas saturadas – nas quais cada ponto está ligado a praticamente todos os demais pontos que conformam a rede. A identificação da configuração das ligações entre nós que conformam a rede é também particularmente importante para a caracterização desse tipo de estrutura.

Nesse sentido, alguns aspectos devem ser levados em conta. Em primeiro lugar, é possível caracterizar uma determinada densidade da rede. Genericamente, o conceito de densidade pode ser associado à relação existente entre o número efetivo de ligações observados na estrutura e o número máximo de ligações que poderiam ocorrer no interior do arranjo em questão. Outro conceito importante refere-se à definição de uma determinada medida que expresse o grau de “centralização” da estrutura. Nesse ponto, dois aspectos costumam ser considerados. O primeiro deles refere-se ao número de ligações que pode ser associado a um ponto particular. Supõe-se que estruturas nas quais determinados pontos concentram um grande número de ligações são mais centralizadas do que outras em que esta característica não pode ser captada. O segundo aspecto refere-se ao número de pontos que constituem passagem necessária entre as ligações estabelecidas entre dois pontos quaisquer da estrutura. Na medida em que seja possível identificar um grande número destes pontos de passagem, a estrutura como um todo poderia ser associada a um maior grau de centralização.

No caso das redes de empresas, a caracterização destas ligações deve contemplar um detalhamento dos relacionamentos organizacionais, produtivos e tecnológicos entre os membros da rede, inclusive no que se refere a aspectos qualitativos destes. Basicamente, estes relacionamentos podem ser referenciados quanto a dois aspectos-chave: a forma e o conteúdo. Quanto à forma dos relacionamentos, um aspecto crucial refere-se ao grau de formalização do arcabouço contratual que regula as relações entre agentes. A funcionalidade desse arcabouço contratual pode ser associada a três propriedades. Em primeiro lugar, este

arcabouço deve definir um conjunto de mecanismos de coordenação com vistas a atingir determinados objetivos pelas partes envolvidas. Em segundo lugar, ele deve contemplar mecanismos de prevenção contra a adoção de posturas oportunistas pelos agentes que estabelecem a relação. Finalmente, em terceiro lugar, este arcabouço deve incluir mecanismos de incitação à adoção de um comportamento eficiente pelas partes envolvidas.

A caracterização morfológica das redes de empresas requer também a identificação do conteúdo de seus relacionamentos internos. Considerando que esses relacionamentos estão articulados a um determinado esquema de divisão de trabalho, é possível identificar três tipos de ligações qualitativamente distintos, em função de um nível crescente de complexidade. Em primeiro lugar, existem ligações sistemáticas entre agentes que se restringem ao plano estritamente mercadológico, não envolvendo o estabelecimento de diretrizes comuns relacionadas a procedimentos produtivos nem a compatibilização-integração das tecnologias empregadas. Em segundo lugar, é possível caracterizar ligações que envolvem a integração de etapas sequencialmente articuladas ao longo de determinada cadeia produtiva. Nesse caso, a compatibilização de uma série de procedimentos técnico-produtivos se faz necessária para elevar o nível de eficiência proporcionado pela estruturação da rede. Finalmente, o terceiro tipo de ligação – qualitativamente mais sofisticado – que envolve a integração de conhecimentos e competências retidos pelos agentes para viabilizar a obtenção de inovações tecnológicas. Nesse caso, as ligações entre agentes extrapolam a mera compatibilização de procedimentos produtivos, envolvendo também a realização de um esforço tecnológico conjunto e coordenado.

Avançando no sentido da caracterização morfológica das estruturas em rede, observa-se que a mera descrição das ligações entre nós é insuficiente, tornando-se necessário identificar a natureza específica dos fluxos que circulam pelos canais de ligação entre aqueles nós. No caso específico das redes de empresas, a análise desses fluxos é mais complicada em função do caráter complexo dessas redes.¹ Para operacionalizar a análise, é possível identificar diferentes fluxos internos que estão presentes nas redes de empresas. Em primeiro lugar, destacam-se fluxos tangíveis baseados em transações recorrentes estabelecidas entre os agentes, pelas quais são transferidos insumos e produtos. Esses fluxos compreendem operações de compra e venda realizadas entre os agentes integrados à rede. Três aspectos diferenciam qualitativamente esses fluxos de transação daqueles externos à rede: (1) o caráter sistemático das transações realizadas, devido à presença de incentivos específicos à continuidade e ao aprofundamento das articulações entre agentes; (2) a realização de algum tipo de adaptação nos procedimentos produtivos realizados devido à integração da empresa à rede; (3) o reforço da especificidade dos ativos envolvidos na transação, como reflexo de adaptações mútuas realizadas nos procedimentos operacionais com a continuidade e o aprofundamento das articulações entre empresas.

Simultaneamente aos fluxos tangíveis, existem os fluxos intangíveis: os fluxos informacionais que conectam os diversos agentes integrados às redes. Do ponto de vista metodológico, a investigação desses fluxos é mais problemática devido à sua natureza intangível, o que dificulta o processo de quantificação dos estímulos que são emitidos e recebidos pelos agentes. Além disso, não existe – como no caso dos fluxos baseados em bens – um arcabouço contratual que regule a transmissão e a recepção desses fluxos. Por fim, deve-se considerar que o conteúdo das informações transmitidas pode variar bastante em termos de seu grau de codificação. Uma parcela importante dessas informações apresenta, inclusive, um caráter tácito, estando baseadas em padrões cognitivos idiossincráticos retidos pelos agentes responsáveis pela sua transmissão e recepção.

A diferenciação proposta entre os vários elementos morfológicos das estruturas em rede – nós, posições, ligações e fluxos – constitui um exercício de simplificação. De fato, a utilização do conceito de redes de empresas como instrumental analítico requer não apenas a identificação daqueles elementos no contexto abordado, como também das interconexões que se estabelecem entre eles, o que requer um esforço de sistematização de dupla direção. Por um lado, é importante realizar uma análise que parta das características dos elementos básicos da rede – determinados nós compostos por empresas e atividades – para, a partir daí, expandir-se o foco no sentido das posições por eles ocupadas em determinado esquema de divisão de trabalho, especificando-se as características das ligações entre eles estabelecidas e dos fluxos associados a essas ligações. Por outro lado, é importante também realizar um percurso analítico em sentido inverso, verificando-se como a necessidade de coordenar e agilizar os fluxos intrarrede afeta as ligações e o posicionamento dos pontos focais da estrutura.

É importante ressaltar também alguns problemas rotineiramente presentes na análise das redes de empresas. O primeiro deles decorre do fato de que essas estruturas são, na verdade, construções abstratas elaboradas com o intuito de reforçar o poder explicativo de um determinado tipo de análise. Em particular, não se deve esperar que os agentes econômicos integrados a essas redes tenham maior clareza sobre as características morfológicas dessas estruturas. Um segundo problema refere-se à dificuldade prática para se definir os limites dessas redes. Tradicionalmente, as análises sobre o fenômeno procuram contornar esse problema por meio da definição de limites vinculados a um determinado ramo industrial, tecnologia, produto, país ou região, para o qual confluem as ligações intrarrede. No entanto, a utilização de um ou outro critério para delimitação daquelas redes não elimina o caráter arbitrário do recorte. Um terceiro aspecto refere-se especificamente ao processo de transformação destas estruturas ao longo do tempo. De fato, as evidências demonstram que essas redes são essencialmente heterogêneas no

que se refere à sua velocidade de transformação, a qual é afetada pela capacidade dos agentes ajustarem seu comportamento em função dos estímulos do processo competitivo.

16.4 Redes de Empresas: Dimensões Relevantes de Operação e Propriedades Internas

Considerando os elementos estruturais discutidos anteriormente, analisaremos a seguir como as redes de empresas operam em situações concretas. Em particular, procuraremos discutir como os elementos estruturais mencionados se combinam para proporcionar uma coordenação eficiente das transações e dos processos de produção e inovação no ambiente intrarrede. Nessa discussão, três impactos distintos associados à consolidação destes arranjos são considerados. Em primeiro lugar, identificam-se impactos diretos associados à esfera técnico-produtiva, que dizem respeito ao aumento da eficiência operacional decorrente da exploração de economias técnicas e à redução de custos de produção e de transação em virtude da consolidação da rede. Em segundo lugar, é possível identificar impactos indiretos associados a instâncias de coordenação das decisões produtivas e tecnológicas dos agentes inseridos na rede, que permitem um melhor enfrentamento da incerteza subjacente à concorrência intercapitalista. Finalmente, em terceiro lugar, identificam-se impactos dinâmicos associados à estruturação dessas redes, que dizem respeito à criação, à circulação e à difusão de informações e ao aprofundamento de mecanismos específicos de aprendizado no interior da rede, o que reforça a capacitação tecnológica e alavanca o potencial inovativo dos agentes integrados ao arranjo. Esses impactos refletem-se em três dimensões distintas – sistematizadas na Figura 16.1 – que procuram representar a natureza específica dos fenômenos que ocorrem no interior das redes de empresas, que são discutidas mais detalhadamente nesta seção do capítulo.

16.4.1 COOPERAÇÃO TÉCNICO-PRODUTIVA EM REDES DE EMPRESAS

Esta dimensão associa-se à sistemática de divisão do trabalho e ao padrão de especialização de funções produtivas entre os diversos agentes inseridos na rede, a partir das quais se conformam sistemas técnico-produtivos com características específicas, que proporcionam ganhos de eficiência para os participantes da rede. A consolidação de sistemas técnico-produtivos estruturados na forma de redes implica a necessidade de aperfeiçoamento da logística de coordenação dos fluxos produtivos no interior

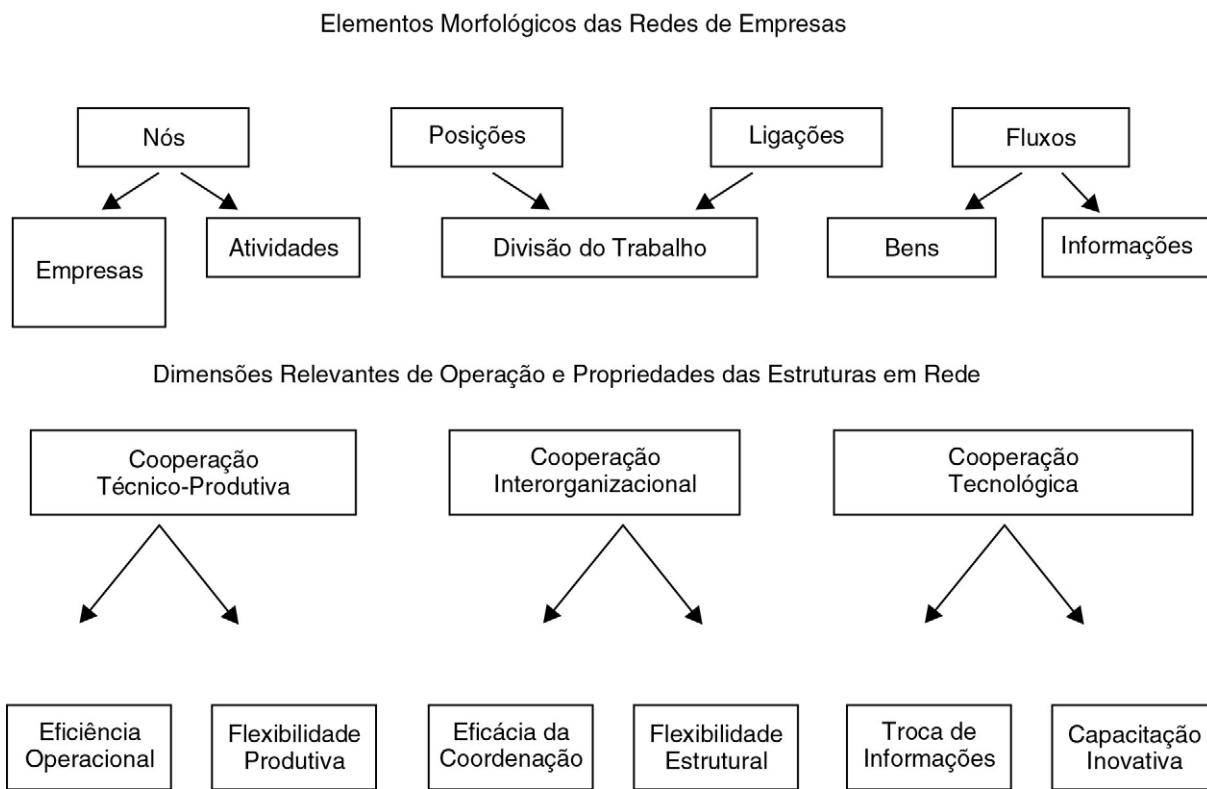


Figura 16.1 Referencial Analítico

desses arranjos. Quanto mais complexa for essa logística, maior será a necessidade de uma coordenação coletiva eficaz das ações dos agentes, tanto do ponto de vista quantitativo como qualitativo. Em especial, é importante compatibilizar os níveis de desempenho técnico-produtivo entre os diversos agentes integrados à rede, contemplando desde ajustamentos nas tecnologias relacionadas ao produto físico gerado e às características dos processos de produção, até procedimentos gerais de formação de recursos humanos, de padrões de controle de qualidade e de normalização técnica.

A obtenção de uma maior eficiência técnico-produtiva, por meio da consolidação de redes de empresas, encontra-se condicionada por dois fatores principais. Em primeiro lugar, observa-se que a obtenção desse tipo de ganho depende das características dos produtos gerados e dos processos empregados no nível da rede. De fato, as possibilidades de obtenção de economias de escala ou escopo no nível da rede – por meio de um processo de especialização produtiva de seus membros – são decisivamente afetadas pelas características tecnológicas dos produtos gerados. Existem também determinados requisitos em termos de compatibilidade técnica entre componentes e subsistemas – que variam de acordo com a complexidade tecnológica do produto – que influenciam a distribuição de tarefas entre os membros da rede. Em segundo lugar, a obtenção de ganhos de eficiência depende dos mecanismos utilizados para coordenar os fluxos produtivos no interior dessas redes. Supõe-se que, quanto mais complexa for a estrutura da rede em termos de sua logística interna, mais importante será a coordenação coletiva dos procedimentos operacionais em seu interior, tanto do ponto de vista quantitativo como qualitativo. A otimização da logística intrarrede por meio da incorporação e difusão de inovações organizacionais – via sistemas do tipo *just-in-time*, por exemplo – deve também estar contemplada neste nível de análise.

Considerando os impactos mencionados, é possível destacar dois tipos de ganhos associados às práticas produtivas adotadas no nível da rede: o aumento da eficiência operacional e o incremento da flexibilidade produtiva. O aumento da eficiência operacional envolve a geração de ganhos técnico-econômicos associados a economias de escala e escopo, assim como outros impactos relacionados ao aumento da qualidade e produtividade dos processos produtivos em seu interior. A obtenção desses ganhos requer a estruturação de um esquema de divisão do trabalho suficientemente azeitado para proporcionar ganhos de especialização, assim como a incorporação de práticas produtivas que garantam uma compatibilização dos níveis de eficiência entre os membros da rede.

O incremento da flexibilidade produtiva está relacionado à capacidade da rede realizar ajustes na sua logística interna face à evolução do mercado e às pressões competitivas mais gerais. Esse tipo de ganho diz respeito à velocidade de resposta da produção gerada pela rede diante de uma eventual mudança nas condições do mercado. Em ambientes dinâmicos, a logística operacional da rede deve ser suficientemente ágil, para viabilizar os ajustes requeridos diante de mudanças circunstanciais nas condições do mercado. Para responder eficazmente aos sinais do mercado, a logística intrarrede deve estruturar-se para evitar a acumulação de estoques indesejados entre as diferentes etapas do processo de produção e a impedir a geração de desbalanceamentos produtivos relevantes entre as diversas atividades realizadas.

16.4.2 COORDENAÇÃO INTERORGANIZACIONAL NAS REDES DE EMPRESAS

Os ganhos competitivos proporcionados pela consolidação de redes de empresas extrapolam uma dimensão estritamente técnico-produtiva, envolvendo também a capacidade de enfrentar, de forma coordenada, a instabilidade ambiental. Esse aspecto diz respeito à estrutura de poder e à conformação hierárquico-funcional da rede, ressaltando os mecanismos internos de resolução de conflitos e a especificidade da concorrência existente entre os membros da rede. Basicamente, essa dimensão diz respeito ao tamanho relativo dos agentes participantes da rede, bem como ao grau de centralização das relações internas que a conformam. Ela contempla também as sistemáticas contratuais que regem as interações entre agentes, as quais envolvem tanto mecanismos de incentivo que estimulam a interação – promovendo a repartição dos ganhos de produtividade e das quase rendas geradas por meio da rede – como formas particulares de resolução dos conflitos. Nesse sentido, é possível associar cada tipo de rede a um regime de transações específico, no qual se destacam uma determinada base contratual – incorporando mecanismos específicos de incentivo e controle – que regula as interações entre seus membros, bem como um determinado nível de confiança mútua que pode ser encontrado em seus relacionamentos internos.

A diversidade institucional inerente às redes de empresas dificulta a caracterização do regime de transações que lhes é subjacente. Apesar disso, há indícios de que as características desses regimes são decisivamente afetadas pelo grau de hierarquização interna da rede, o qual, por sua vez, é influenciado pelo tamanho relativo de seus membros e pela maior ou menor centralização de seus fluxos internos. Além disso, outros aspectos correlacionados ao regime de transações subjacente às redes de empresas podem ser mencionados: (1) relações típicas entre empresas que podem ser observadas em cada tipo de rede, a partir das quais se pode ter uma noção mais precisa sobre as respectivas estruturas de governança; (2) as principais formas de coordenação das decisões no nível da rede, ressaltando-se o objeto dessa coordenação – que pode envolver múltiplas dimensões – e a base contratual na

qual ela se apoia; (3) características das ações coletivas adotadas pelos membros da rede, que permitem a obtenção de ganhos de eficiência, facilitando a repartição de tarefas e permitindo um enfrentamento coordenado da turbulência ambiental; (4) níveis de confiança mútua presentes nos relacionamentos internos às redes.

Considerando os aspectos mencionados, é possível destacar dois tipos de ganhos associados às práticas cooperativas adotadas no nível da rede: o fortalecimento da eficácia do processo de coordenação interorganizacional e a realização de ajustes na estrutura morfológica da rede em função de estímulos ambientais. A eficácia do processo de coordenação interorganizacional é decisivamente afetada por características específicas da estrutura interna da rede. Em especial, esta coordenação é afetada pelo grau de centralização dos fluxos de autoridade internos à rede, envolvendo a presença (ou não) de empresas ou outras instituições com a função de coordenar ou compatibilizar as decisões operacionais e estratégicas adotadas pelos membros da rede. Essa propriedade está também associada à natureza específica das ações coletivas adotadas pelos membros da rede. Dois elementos de caracterização dessas ações coletivas podem ser mencionados: o seu caráter bilateral ou multilateral, associado ao número de partes envolvidas, e o seu caráter horizontal ou vertical, associado ao padrão de especialização dos agentes envolvidos. Com base nesses critérios, quatro tipos básicos de ações coletivas podem ser identificados no nível da rede: (1) ações coletivas bilaterais de caráter horizontal, envolvendo pares de empresas que realizam o mesmo tipo de atividade; (2) ações coletivas bilaterais de caráter vertical, envolvendo ligações específicas entre duas empresas na cadeia produtiva; (3) ações coletivas multilaterais de caráter horizontal, quando um conjunto de empresas similares decide se articular para realizar uma tarefa qualquer; (4) ações coletivas multilaterais de caráter vertical, envolvendo diversas empresas localizadas em diferentes pontos da cadeia produtiva que se articulam para atingir um objetivo comum.

Ainda no que se refere ao processo de coordenação intrarrede, é possível considerar outro tipo de ganho, associado à capacidade da rede ajustar sua conformação morfológica em função de estímulos ambientais. Essa flexibilidade interorganizacional compreende mudanças nas funções desempenhadas pelos membros da rede, por meio de modificações na própria estrutura do arranjo, devido ao processo de entrada e saída de agentes. De fato, em ambientes voláteis, as redes de empresas terão maiores probabilidades de gerar ganhos competitivos se operarem como estruturas reversíveis, capazes de promover ajustes na sua estrutura interna em função de estímulos ambientais. Nessas condições, seria interessante que a rede operasse a partir de um núcleo central de agentes e competências, em torno do qual gravitariam outros agentes periféricos que seriam absorvidos ou excluídos da rede em função de estímulos ambientais.

16.4.3 COOPERAÇÃO TECNOLÓGICA EM REDES DE EMPRESAS

Do ponto de vista dinâmico, o reforço da competitividade por meio da consolidação das redes de empresas envolve o fortalecimento do potencial inovativo de seus membros constituintes. Nesse sentido, uma das principais características das redes de empresas refere-se à criação e à circulação de conhecimentos e informações, envolvendo a consolidação de um processo de aprendizado coletivo que amplia o potencial inovativo da rede. Esse aprendizado coletivo é resultante de um intercâmbio de informações e competências, envolvendo a incorporação do aprendizado individual de cada agente a um *pool* social de conhecimentos (comerciais, gerenciais, mercadológicos, tecnológicos etc.) gerados a partir da rede.

É possível identificar quatro formas de aprendizado coletivo que são específicas ao ambiente intrarrede. A primeira delas envolve a criação de conhecimentos tecnológicos intencionalmente desenvolvidos em cooperação, contemplando a realização de atividades conjuntas de Pesquisa e Desenvolvimento entre os componentes das redes, a partir da consolidação de uma certa divisão de trabalho que orienta o esforço tecnológico realizado. Nesse caso, as redes estruturam-se a partir da montagem de projetos particulares, nos quais interagem agentes dotados de competências complementares envolvidos com diferentes etapas do ciclo de P&D-produção. Essa característica remete à análise desses arranjos no sentido da discussão dos seguintes aspectos: (1) a identificação dos principais objetivos que orientam a realização dos esforços conjuntos de P&D entre os membros da rede; (2) a descrição do arcabouço institucional que serve de base para a sua realização, incluindo a identificação do papel dos agentes e dos mecanismos de incentivo mobilizados; (3) a avaliação do volume de recursos efetivamente comprometidos com a realização destes esforços.

Um segundo tipo de aprendizado intrarrede relaciona-se à circulação de conhecimentos tecnológicos no interior desse tipo de arranjo. Esse tipo de aprendizado assume um caráter informal, na medida em que não requer necessariamente o comprometimento de recursos com atividades de P&D. Nesse caso, a cooperação tecnológica atua no sentido de permitir uma aceleração do processo inovativo, por meio de um intercâmbio de informações que retroalimenta o esforço tecnológico dos agentes. A análise das redes de empresas deve incluir a discussão dos mecanismos mobilizados para disseminar informações tecnológicas relevantes em seu interior, contemplando os seguintes aspectos: (1) a descrição dos sistemas de informação tecnológica que se encontram disponíveis para os membros da rede; (2) a avaliação dos mecanismos pelos quais o conhecimento codificado

circula no interior da rede; (3) a avaliação dos instrumentos mobilizados para viabilizar a circulação de conhecimentos tácitos entre os membros da rede.

Uma terceira forma de aprendizado intrarrede está relacionada ao incremento coordenado das competências dos agentes em seu interior. Como exemplo, é possível mencionar a realização de investimentos na qualificação de recursos humanos, por meio da coordenação dos esforços de treinamento entre os membros da rede. Outro exemplo importante refere-se à realização de esforços visando à redução do hiato de competências técnicas entre os membros da rede, por meio da disseminação de padrões comuns de tecnologia industrial básica. Para compreender esse aspecto, a análise das redes de empresas deve contemplar a discussão dos seguintes tópicos: (1) a identificação das instituições especificamente montadas com a intenção de facilitar o incremento das competências dos membros da rede; (2) a avaliação do processo de evolução dessas competências ao longo do tempo, visando a identificar como evoluem os diferenciais existentes entre os diversos membros da rede.

Uma última forma de aprendizado intrarrede refere-se à sua conversão em estrutura propulsora da difusão de novas tecnologias. Nesse sentido, a rede funciona como um mercado organizado que favorece a difusão de novas tecnologias, quando comparado a mercados comuns baseados em transações particularizadas. Essa perspectiva pressupõe que novas tecnologias se difundem mais rapidamente entre agentes que estabelecem relações sistemáticas entre si. Nesse ponto, a análise das redes de empresas deve incluir um melhor detalhamento dos fluxos tecnológicos internos à rede, bem como a discussão dos mecanismos internos de transferência tecnológica e de outros aspectos que podem acelerar o processo de difusão em seu interior.

Duas propriedades genéricas podem ser associadas ao processo de cooperação tecnológica que ocorre no nível da rede: a capacidade de seus membros identificarem e processarem informações relevantes e o fortalecimento de capacitações inovativas, por meio da aglutinação de competências e qualificações complementares. O fortalecimento da capacidade de processamento de informações refere-se à complexidade e à densidade dos fluxos informacionais intrarrede, bem como à “socialização” de conhecimentos entre os agentes integrados à rede. Essa propriedade assume particular importância em ambientes complexos e voláteis, que estão permanentemente emitindo sinais que necessitam ser capturados, interpretados e respondidos pelos agentes econômicos.

Além do processamento de informações, é importante considerar como a rede é capaz de fortalecer as capacitações inovativas de seus membros. Tal propriedade assume particular importância em ambientes dinâmicos, nos quais o reforço da competitividade requer a introdução continuada de inovações no mercado. Para que essa propriedade se torne operativa, é necessário que ocorra uma aglutinação das competências dos membros da rede, por meio de projetos conjuntos orientados à integração de competências. Em algumas situações, as inovações tecnológicas obtidas são o resultado de um processo consciente e intencional por parte dos agentes, que conduz à montagem de projetos cooperativos de P&D. Em outras situações, é possível que inovações tecnológicas venham a ser obtidas, naturalmente, como subprodutos dos relacionamentos produtivos e tecnológicos que interligam os diferentes agentes integrados à rede.

16.5 Redes de Empresas na Prática: Uma Tentativa de Sistematização

Nos últimos anos, uma vasta literatura vem se dedicando à construção de tipologias sobre redes de empresas. Em geral, essas tipologias estão baseadas na caracterização de determinados modelos estilizados, por meio dos quais se procura formar a diversidade institucional das redes. O Quadro 16.3 apresenta, a título de ilustração, três dessas tipologias. Apesar dos problemas relacionados à construção de tais tipologias,² estas são importantes como um quadro de referência capaz de captar a diversidade institucional desses arranjos. Com esse intuito, são discutidos, a seguir, três formas estilizadas de redes de empresas recorrentemente mencionadas na literatura: (1) redes de subcontratação, nas quais se destaca a presença de uma empresa principal responsável pela coordenação dos fluxos internos à rede; (2) redes baseadas na aglomeração espacial de agentes em distritos industriais; (3) redes tecnológicas, montadas com o intuito de permitir um intercâmbio de competências a partir do qual seria possível viabilizar a introdução de inovações no mercado.

16.5.1 REDES DE SUBCONTRATAÇÃO

Um primeiro tipo de rede descrito na literatura comprehende redes verticais baseadas em estratégias de desintegração e subcontratação de grandes grupos industriais. Essas redes caracterizam-se pela presença de relações de cooperação entre fornecedor-cliente e produtor-usuário incorporando, dentre outras tendências, o alargamento da duração dos acordos entre empresas, uma nova repartição de tarefas entre o contratante e as empresas subcontratadas, uma intensificação da cooperação técnica e do intercâmbio de informações entre os agentes e uma redução do número de subcontratados diretos em favor de práticas de

QUADRO 16.3 TIPOLOGIAS DE REDES DE EMPRESAS**1. Garofoli (1993)**

1. *Sistemas de Produção em Grande Escala (Redes Verticais)*: aglomeração espacial de unidades com presença de vínculos fortemente hierarquizados, que confluem no sentido de grandes empresas especializadas na montagem de componentes;
2. *Sistemas de Pequenas Empresas (Distritos Industriais)*: pequenas empresas concentradas do ponto de vista espacial cujos inter-relacionamentos não se prendem a vínculos hierárquicos, e sim a práticas de cooperação bidirecionais;
3. *Produção Descentralizada (com Presença de Empresa Dominante)*: presença de unidades dispersas do ponto de vista espacial que, no entanto, mantêm sólidos vínculos de dependência hierárquica em relação à empresa responsável pela montagem de componentes;
4. *Acordos Cooperativos Baseados em Alianças Estratégicas*: colaboração entre agentes dispersos do ponto de vista espacial, que estabelecem, entre si, práticas cooperativas não hierarquizadas baseadas no intercâmbio de informações e na reciprocidade de ações.

2. Markunsen (1994)

1. *Distritos Marshallianos Tradicionais*: redes baseadas na especialização funcional de agentes atuantes em determinada região, nas quais se destaca a presença de pequenas e médias empresas;
2. *Distritos do Tipo Centro Radial*: redes localizadas espacialmente em determinada região, cuja estrutura se articula em torno de uma ou várias grandes empresas atuantes em determinado setor;
3. *Plataformas Industriais Satélites*: redes ancoradas na presença de subdivisões ou sucursais de empresas multinacionais atraídas por uma dotação particular de recursos locais;
4. *Distritos Suportados pelo Estado*: redes que são estruturadas a partir da ação de alguma agência ou empresa estatal que funciona como âncora do desenvolvimento econômico regional.

3. Langlois e Robertson (1995)

1. *Distrito Marshalliano*: estrutura com grau de integração da propriedade e grau de coordenação baixos. Tendência à especialização de recursos, via especialização horizontal e vertical de PMEs autônomas.
2. *Distritos do Tipo "Terceira Itália"*: estrutura com baixo grau de integração de propriedade e elevado grau de coordenação, caracterizada pela forte especialização horizontal-vertical de pequenas empresas. Competição entre agentes restringe-se a campos que geram competências distintivas (*design*, por exemplo). Presença de cooperação na provisão de infraestrutura e serviços.
3. *Distritos Inovativos do Tipo Venture Capital* (como o Silicon Valley): presença de empresas de base tecnológica como núcleo central da rede. Coordenação promovida pela mediação de *venture capital*, com tendência à paulatina consolidação de propriedade. Crescimento a partir de capacitação tecnológica preexistente.
4. *Redes Japonesas (Kaisha Networks)*: redes com núcleo bem definido, composto por empresa-principal que promove coordenação. Ênfase em contratos “relacionais” de longo prazo, com estímulo à confiança e redução de custos de transação. Otimização da logística de produção de sistemas do tipo *just-in-time*. Possibilidade de interpenetração de propriedade e de conexões financeiras entre agentes.

cooperação mais interativas com os subcontratados principais. Geralmente, essas redes surgem a partir do processo de desverticalização de grandes empresas produtoras de bens de consumo duráveis, induzido por fatores tecnológicos ou estímulos competitivos. Nesse caso, a integração de componentes costuma estar baseada em arquitetura modular do produto, que aumenta a variedade de modelos possíveis de serem obtidos, elevando a flexibilidade associada ao conjunto de atividades produtivas integradas à rede. O principal objetivo dessas ligações produtivas é a geração de produtos diferenciados, a partir da integração de componentes e subsistemas, utilizando-se plataformas similares como meio para elevar a variedade do leque de produtos obtidos. O exemplo da evolução recente das relações entre montadoras e fornecedoras na indústria automobilística, retratado no Quadro 16.4, ilustra algumas das principais características desse tipo de rede.

Os ganhos técnico-produtivos obtidos usualmente dizem respeito à geração simultânea de economias de escala e escopo ao longo da cadeia produtiva. As economias de escala estão presentes na especialização das empresas (montadoras e fornecedoras), ao passo que as economias de escopo decorrem da possibilidade de obtenção de diferentes especificações finais do produto a partir da mesma linha de produção, por meio da arquitetura modular de componentes. Além disso, o reforço da flexibilidade produtiva como fator de geração de ganhos competitivos assume grande importância no âmbito dessas redes. O incremento dessa flexibilidade está geralmente vinculado à otimização da logística interna da rede entre montadoras e fornecedoras, pela disseminação de técnicas organizacionais do tipo *just-in-time*.

A estruturação de relações de subcontratação baseadas na cooperação interindustrial requer a montagem de uma estrutura de autoridade que seja funcional para a coordenação dessas relações. A montagem de uma estrutura de poder sustentável no âmbito dessas redes compreende um processo pelo qual esta estrutura vai sendo legitimada como arranjo interorganizacional adequado ao enfrentamento da incerteza inerente ao ambiente econômico. Existem quatro aspectos que merecem

QUADRO 16.4 REDES VERTICais DE SUBCONTRATAÇÃO: A EXPERIÊNCIA DA INDÚSTRIA AUTOMOBILÍSTICA

A indústria automobilística mundial vem passando nas últimas décadas por um processo de acirramento da competição que tem reforçado a busca de uma maior flexibilidade produtiva, particularmente pela disseminação de novas formas de organização e gerenciamento da produção, que se baseiam nos conceitos de *just-in-time*, produção por encomenda, minimização do *lead time*; qualidade perfeita e controle de qualidade total, operários multiqualificados e desverticalização de atividades produtivas. Essas transformações têm modificado, de forma significativa, as relações entre empresas montadoras e seus fornecedores. Destaca-se, nesse sentido, a tendência, por parte das montadoras, de transferência (ou terceirização) para fornecedores da produção de componentes e sistemas maiores e mais complexos, bem como uma maior participação de fornecedores no design e construção, execução e montagem de sistemas automotivos. Consolida-se, desse modo, um processo de estruturação de redes de subcontratação que contemplam uma diferenciação hierárquica de fornecedores. Essa tendência remonta à década de 1980, refletindo a experiência bem-sucedida de empresas japonesas, que se difundiu paulatinamente para demais empresas do setor.

A presença de uma estrutura piramidal de fornecedores era originariamente mais nítida no caso da indústria automobilística japonesa, fazendo com que as montadoras operassem com um número de fornecedores diretos menor que empresas americanas e europeias. A Tabela 16.1 ilustra esse aspecto, apresentando a distribuição de fornecedores da Toyota por nível de importância. A dinamização das redes de subcontratação no interior da estrutura industrial japonesa remonta aos impactos recessivos da crise do petróleo em meados da década de 1970, que determinou a formulação de estratégias direcionadas para a melhoria da qualidade e o aumento da eficiência no suprimento de insumos. Para incrementar a qualificação técnica de fornecedores, foi estimulada, por grandes empresas, a formação de kyoryokukai, associações formais orientadas à disseminação de práticas cooperativas entre fornecedores que conduzissem à consolidação de normas técnicas adotadas pelos membros da rede. Essas normas operavam como instrumento para transferência e socialização de inovações organizacionais – *just-in-time*, programas de qualidade total, kan-ban etc. – e para intensificação do intercâmbio de informações no âmbito da rede. Para garantir uma maior eficiência no processo de subcontratação, as empresas japonesas procuraram reduzir o número de fornecedores com os quais interagiam diretamente, ao mesmo tempo em que intensificaram o intercâmbio de informações com eles. Além disso, as redes de subcontratação japonesas reforçaram a sua estruturação em vários “níveis” hierárquicos, nos quais se diferenciam, sucessivamente, fornecedores de primeira ordem, segunda ordem e assim sucessivamente.

TABELA 16.1 Estrutura de Fornecedores por Nível – Toyota

Parte	Primeiro nível	Segundo Nível	Terceiro Nível	Total
Partes do motor	25	912	4.960	5.897
Componentes elétricos/eletrônicos	1	34	352	387
Transmissão, câmbio, direção	31	609	7.354	7.994
Freios e suspensão	18	792	6.204	7.014
Partes dos freios e suspensão	18	92	5.936	6.880
Chassis e partes	3	27	85	115
Carroceria e prensagem	41	1.213	8.221	9.475
Outros	31	924	8.591	9.546
Total	168	5.437	41.703	47.308

Fonte: MITI, in: Fruin e Nishiguchi (1993: 226)

Mais recentemente, em função do acirramento da concorrência na indústria, observou-se uma tendência à diminuição do número de fornecedores diretos de empresas automobilísticas não japonesas. A Tabela 16.2 demonstra que, no período compreendido entre 1986 e 1992, algumas empresas montadoras europeias reduziram, consideravelmente, sua base de fornecedores. Alguns exemplos são particularmente expressivos, como os da Peugeot-Citroën (que reduz sua base de fornecedores de 1.310 para 790 empresas), Renault (redução de 1.240 para 810) e GM europeia (redução de 2.000 para 1.190). O mesmo tipo de fenômeno pode ser observado em empresas instaladas em países em desenvolvimento. No caso brasileiro, por exemplo, o número de fornecedores diretos da Fiat reduziu drasticamente, na última década, de 500 em 1989 para aproximadamente 220 no final dos anos 1990, sendo intenção da empresa reduzir esse número para 100 fornecedores diretos.

TABELA 16.2 Evolução da Base de Fornecedores de Empresas Montadoras Europeias

Empresa	1986	1988	1990	1992
Fiat	860	1.000	900	810
Ford	1.290	1.190	1.100	1.000
GM	2.000	2.000	1.810	1.190
Renault	1.240	1.050	950	810
Volkswagen	1.760	1.480	1.290	1.190
Peugeot-Citroën	1.310	950	860	790

Fonte: Lamming (1992:242)

ser ressaltados quanto a esse processo: (1) a criação de códigos de conduta consensualmente aceitos que integrem as ações dos agentes; (2) a coordenação das relações de troca e dos fluxos de informações que se estabelecem entre os agentes; (3) a montagem de uma estrutura de comando subjacente à rede, com determinado grau de centralização, que seja validada pelos seus membros; (4) a incorporação de práticas de gestão que garantam a eficácia do arranjo interorganizacional ao longo do tempo.

Muitas vezes, é possível observar uma nítida segmentação das relações de subcontratação nessas redes, que diferencia uma subcontratação de primeiro nível, baseada em relações mais cooperativas, de uma subcontratação de segundo nível, fundamentada em relações puramente mercantis. No primeiro caso, observa-se que o agente subcontratado desempenha uma função de concepção, adequando-se à especificação de resultados e às normas funcionais definidas pelo contratante, mas gozando de autonomia para desenvolver o produto de acordo com suas competências. Nesse tipo de relação, os agentes contratados desempenham uma função completa, beneficiando-se de um contrato plurianual que amplia as possibilidades de interação e aprendizado mútuo com o contratante. A informação técnica que circula na rede é extremamente densa e sofisticada, movendo-se em um duplo sentido que permite a compatibilização dos patamares de eficiência técnico-produtiva e a integração das competências tecnológicas e organizacionais.

Em contraste, no segundo nível do mercado de subcontratação predominam relações particularizadas entre os agentes, que se aproximam das relações mercantis tradicionais, contemplando um baixo grau de interatividade e um fluxo informacional relativamente restrito. Nesse segmento, é possível à empresa contratante substituir com facilidade as empresas subcontratadas, minimizando os custos de troca de fornecedores e favorecendo a flexibilidade operativa. O subcontratado de segundo nível realiza um conjunto de tarefas restritas, geralmente associadas à obtenção de uma peça ou componente que atenda aos requerimentos técnicos impostos pelo contratante. Esse tipo de empresa, em geral, não participa do processo de concepção do produto, possuindo um nível restrito de competência.

A hierarquização de fornecedores, por meio da criação de diversos segmentos de subcontratação, amplia as barreiras à entrada externas à rede e cria barreiras à entrada internas, relacionadas às dificuldades que um fornecedor enfrenta para ascender do segmento de segundo nível para o segmento de primeiro nível do mercado de subcontratação. Essas barreiras estão menos associadas a fatores estritamente técnicos e mais relacionadas a uma capacitação tecnológica prévia dos fornecedores, que facilite a maior interação com as empresas contratantes. Outro aspecto importante refere-se à localização espacial das empresas que estabelecem entre si relações de subcontratação de caráter cooperativo. Observa-se que a integração de competências no âmbito dessas redes e a consolidação de relações interativas em seu interior requerem uma certa proximidade espacial entre os agentes, estimulando a organização de subsistemas locais.

Em termos da organização e gestão das empresas, é interessante distinguir os impactos sobre subcontratados e contratantes. Para as empresas subcontratadas, a consolidação de relações interindustriais cooperativas requer uma intensificação do processo de capacitação produtiva e tecnológica, imprescindível ao aprofundamento das interações com os contratantes. Em especial, é necessário que os fornecedores de componentes adequem seus procedimentos produtivos a patamares mínimos de eficiência, incorporando modernas técnicas organizacionais que incrementam a eficiência técnico-produtiva de seus membros. Dentre estas, destacam-se modernas técnicas de controle e gerenciamento da qualidade ao longo das várias etapas dos processos produtivos, bem como o incremento contínuo da produtividade. A difusão dessas técnicas insere-se na consolidação de um sistema de *just-in-time* externo que integre os diversos componentes da rede de subcontratação, otimizando a configuração do sistema técnico-produtivo por ela mobilizado.

Para as empresas contratantes, a consolidação de relações cooperativas também resulta em importantes impactos. Em linhas gerais, observa-se uma paulatina reconfiguração organizacional dessas empresas, que tendem a se especializar em funções mais nobres, associadas a um maior conteúdo intelectual e relacional. Dentre essas atividades, destacam-se: pesquisa e desenvolvimento; concepção; montagem e controle final; disseminação de normas relativas a procedimentos operacionais e à sistemática de controle e garantia da qualidade; compatibilização técnica e tecnológica dos insumos adquiridos; venda e financiamento etc. Além disso, a redistribuição de tarefas e funções resulta na combinação de três efeitos, que alteram a configuração interna das empresas contratantes: (1) o desenvolvimento e/ou dinamização de determinados serviços e funções internos (compras, controle de qualidade, P&D); (2) a intensificação dos fluxos informacionais, por meio da integração de sistemas computacionais e da incorporação de novas tecnologias de telecomunicação; (3) o posicionamento dessas empresas como gestores das relações de subcontratação realizadas no âmbito da rede.

No que se refere ao processo de cooperação tecnológica, a consolidação desse tipo de rede geralmente envolve uma sofisticação dos fluxos de informação que interligam empresas montadoras e seus fornecedores. Alguns aspectos desse processo podem ser enfatizados. Em primeiro lugar, destaca-se o intercâmbio de informações sobre o desempenho e a qualidade dos componentes entre empresas montadoras e seus fornecedores. Esse intercâmbio facilita o aperfeiçoamento dos produtos existentes e a geração de novos modelos com base em ajustes na arquitetura modular do produto. A intensificação dos fluxos de informações entre montadoras e fornecedores é resultado também do codesenvolvimento de novos componentes. Nesse caso, os fluxos de informação assumem um caráter bidirecional, utilizando uma infraestrutura sofisticada e estando associados ao desenvolvimento de códigos específicos de comunicação. O conteúdo desses fluxos envolve não apenas a transmissão de ordens entre montadoras e fornecedores, como também os seguintes aspectos: (1) desenvolvimento de especificações de componentes e subsistemas (via intercâmbio de arquivos de CAD,³ por exemplo); (2) negociações e ajustamentos relacionados à logística de entrega de componentes, visando a sua otimização; (3) monitoramento da qualidade dos componentes produzidos; (4) solução de problemas técnicos relacionados ao processo de suprimento no menor espaço de tempo; (5) intercâmbio geral de informações técnicas.

A consolidação de um processo de codesenvolvimento de novos componentes e subsistemas entre montadoras e empresas fornecedoras requer a integração de capacitações inovadoras no interior dessas redes. A integração dessas capacitações envolve inovações programadas baseadas em projetos de desenvolvimento de produtos que procuram integrar novos componentes e subsistemas no interior de determinada arquitetura modular. Para viabilizar a geração dessas inovações, é comum a divisão de tarefas e atribuições entre os agentes participantes, em especial empresas montadoras e seus fornecedores de primeiro nível. Ao mesmo tempo em que o *design* de alguns componentes e subsistemas é transferido para os fornecedores mais capacitados, as empresas montadoras devem ser capazes de avaliar adequadamente a capacitação de seus fornecedores, para verificar até que ponto esta transferência é possível, visando garantir que os componentes desenvolvidos atendam aos requisitos de desempenho e de qualidade exigidos.

16.5.2 DISTRITOS E AGLOMERAÇÕES INDUSTRIAS

Um segundo tipo de rede estaria baseado na obtenção de ganhos de eficiência em virtude da aglomeração espacial de atividades complementares do ponto de vista tecnológico e/ou mercadológico, por meio da formação de distritos ou aglomerações industriais. O conceito de distritos industriais é elaborado a partir de desdobramentos da análise originariamente formulada por Alfred Marshall, estando relacionado aos diversos ganhos proporcionados pela especialização produtiva das empresas e pela sofisticação da divisão do trabalho proporcionada pela aglomeração espacial de empresas atuantes em um mesmo ramo de atividade, ou em atividades relacionadas. Além disso, a noção de distritos industriais tem sido correlacionada não apenas à presença de empresas, mas também a um conjunto de outras instituições que conformam um ambiente local capaz de reforçar a competitividade das empresas atuantes nesses distritos.⁴ Na medida em que a existência de uma aglomeração espacial de indústrias é um pré-requisito para a existência desses distritos, evidências relativas a esse processo de aglomeração podem constituir um ponto de partida para a caracterização desses distritos. Nesse sentido, o Quadro 16.5 apresenta algumas evidências relativas à existência de aglomerações industriais na economia brasileira, o que possibilita um mapeamento preliminar daqueles distritos. O Quadro 16.6, por sua vez, discute algumas características específicas dos distritos industriais associados a setores de elevado conteúdo tecnológico.

O conceito de distritos industriais está associado à presença de sistemas de produção locais fortemente integrados, marcados pelas seguintes características: (1) a ocorrência de uma especialização produtiva no nível local, com base em processo histórico de acúmulo de competências; (2) a relevância da produção concentrada em determinadas localidades em relação à produção total de determinada indústria no âmbito nacional; (3) a existência de uma intensa divisão de trabalho no nível local,

QUADRO 16.5 AGLOMERAÇÕES INDUSTRIALIS: EVIDÊNCIAS DA EXPERIÊNCIA BRASILEIRA

O conceito de aglomerações (ou *clusters*) industriais refere-se à emergência de uma concentração geográfica e setorial de empresas, a partir da qual são geradas externalidades produtivas e tecnológicas indutoras de um maior nível de eficiência e competitividade. Ao se apoiarem mutuamente, as empresas integradas a esses arranjos conferem vantagens competitivas no nível industrial para uma região particular, permitindo explorar diversas economias de aglomeração. Apesar da cooperação produtiva e/ou tecnológica não estar necessariamente presente nessas aglomerações, supõe-se que a estruturação destas estimula um processo de interação local que viabiliza o aumento da eficiência produtiva, criando um ambiente propício à elevação da competitividade dos agentes integrados ao arranjo. Além disso, a intensificação das articulações e interações entre empresas presentes nessas aglomerações pode ter impactos importantes em termos da geração de efeitos de aprendizado e da dinamização do processo inovativo em escala local ou regional.

No caso brasileiro, a importância dessas aglomerações tem sido reforçada em função de aspectos específicos. Em particular, o processo de reestruturação produtiva do setor industrial ocorrido na década de 1990 gerou importantes desdobramentos sobre as articulações entre agentes no interior das cadeias produtivas e sobre o padrão de localização espacial das atividades industriais. As crescentes pressões pela busca de maiores níveis de eficiência na utilização de fatores produtivos estimularam a localização de atividades produtivas em regiões onde a disponibilidade de fatores – mão de obra e recursos naturais, em especial – fosse mais favorável, tanto do ponto de vista quantitativo como qualitativo. É possível mencionar também um processo de desconcentração espacial da indústria, com o consequente surgimento de novas áreas industriais, que remonta à década de 1970, mas que vem adquirindo uma nova dinâmica no período mais recente. Além disso, destacam-se estímulos de instrumentos de política econômica, definidos no plano federal, estadual e municipal, à relocalização espacial de indústrias. Por fim, um fator adicional diz respeito aos impactos do processo de integração inter-regional, em particular aquele associado ao advento da criação do Mercosul, em termos da dinâmica de localização espacial da indústria.

Algumas análises recentes têm contribuído para um mapeamento mais detalhado da presença de aglomerações industriais na economia brasileira. Sugerimos a consulta aos dados levantados por Sabóia (2000) sobre as aglomerações (microrregiões) que concentravam mais de 15% do emprego industrial naquele tipo de atividade. A partir dos dados apresentados, algumas tendências podem ser captadas:

- A formação de aglomerações (ou *clusters*) industriais é mais nítida em determinados setores do que em outros. De fato, um número mais elevado dessas aglomerações pode ser encontrado em alguns setores, como têxtil, de calçados, madeira/mobiliário e metalúrgico.
- Em alguns setores claramente ocorre uma concentração das aglomerações em determinadas regiões do país, como no caso dos setores de calçados (no Rio Grande do Sul, São Paulo e Ceará), material de transporte (São Paulo), extrativo mineral (Minas Gerais), madeira/mobiliário (Paraná e Santa Catarina), metalúrgico (São Paulo e Minas Gerais) e Químico (São Paulo).
- Em alguns setores (como têxtil, calçados, madeira/mobiliário e metalúrgico) é comum a presença de aglomerações industriais que concentram mais de 60% do emprego industrial. A princípio, é possível caracterizar essas atividades como indústrias com tendência à *clusterização*. O fato de essas aglomerações estarem localizadas predominantemente em setores tradicionais deve, porém, ser visto com cuidado, pois a menor importância de *clusters* industriais em setores de maior conteúdo tecnológico pode, no caso brasileiro, refletir, em parte, a inexistência de uma base industrial mais sólida nesses setores, com capacidade de geração de efeitos dinamizadores sobre as regiões onde as principais empresas estão localizadas.

em termos intra e interindustriais, responsável pela redução dos custos de transação nas operações realizadas entre empresas locais; (4) a presença de uma pluralidade de protagonistas no nível local, associada à existência de um grande número de agentes locais satisfatoriamente capacitados e à ausência de uma empresa dominante claramente identificável para o conjunto de atores participantes da rede; (5) a ampliação cumulativa da base de conhecimentos dos agentes locais por meio do processo de especialização, facilitando a introdução de novas tecnologias e viabilizando o aumento dos níveis de produtividade; (6) a existência de um sistema eficiente de transmissão de informações no nível local, garantindo uma circulação rápida e eficiente de informações sobre escoadouros para a produção, tecnologias alternativas, disponibilidade de insumos-componentes e novas técnicas de marketing e comercialização; (7) a existência de um elevado nível de qualificação dos recursos humanos no nível local, como resultado de um processo histórico de sedimentação de habilidades e conhecimentos; (8) a generalização de relações diretas entre os agentes no nível local, o que facilita a difusão extensiva de inovações tecnológicas e organizacionais, favorecendo o aumento da eficiência dos sistemas locais de produção.

A literatura especializada sobre o tema associa este tipo de rede a um conjunto de empresas e instituições espacialmente concentradas que estabelecem entre si relações verticais – compreendendo diferentes estágios de determinada cadeia produtiva – e horizontais – envolvendo o intercâmbio de fatores, competências e informações entre agentes genericamente similares. Em termos da sua conformação interna, esses *clusters* geralmente abrangem empresas interdependentes (incluindo fornecedores especializados), agentes produtores do conhecimento (universidades, institutos de pesquisa, empresas de consultoria etc.), instituições-ponte (consórcios, incubadoras etc.) e consumidores, que se articulam entre si por meio de uma cadeia produtiva espacial e setorialmente localizada. Ao se apoiarem mutuamente, os agentes integrados a esses arranjos conferem vantagens competitivas no nível industrial para uma região particular, permitindo explorar diversas economias de aglomeração e outros

QUADRO 16.6 DISTRITOS INDUSTRIAS EM SETORES DE ALTA TECNOLOGIA: O CASO DOS PARQUES TECNOLÓGICOS

O conceito de parques tecnológicos refere-se à montagem de redes relativamente semelhantes aos distritos industriais que, entretanto, direcionam-se especificamente para setores de alta tecnologia. Os parques tecnológicos, por sua vez, compreendem a integração numa mesma região de componentes, *softwares* e serviços necessários à obtenção de produtos com elevado conteúdo tecnológico – como computadores, produtos biotecnológicos etc. Esse tipo de configuração, em geral, localiza-se próximo a regiões que dispõem de uma infraestrutura científico-tecnológica consolidada em universidades e institutos de pesquisa, aproveitando os avanços científicos gerados por essas instituições e absorvendo uma parcela da mão de obra especializada formada a partir destas. Basicamente, esse tipo de rede garante o entendimento comum dos problemas tecnológicos relevantes, bem como permite a intensificação de processos de aprendizado entre seus membros participantes. A necessidade de integrar vários componentes e as diversas etapas do processo de P&D no âmbito desses arranjos estimula a aglomeração espacial das empresas integradas a esse tipo de rede. Em termos da conformação institucional dos parques tecnológicos, observa-se uma confluência dos fluxos de informações e das transações realizadas no sentido de empresas de base tecnológica. Muitas vezes, essas empresas são originárias do próprio meio universitário, integrando profissionais de perfil nitidamente acadêmico que decidem redirecionar sua qualificação no sentido de aplicações industriais. A estruturação de empresas de base tecnológica dinâmica – geradoras de um efeito transbordamento (*spin-off*) – não requer maiores economias internas do ponto de vista técnico, na medida em que as tecnologias emergentes para as quais estão direcionadas não requerem grandes unidades de produção para viabilizar o processo industrial. Muitas vezes, inclusive, essas empresas podem surgir a partir de instalações comuns relativas à infraestrutura e a equipamentos repartidos no interior daqueles parques, as denominadas incubadoras de empresas. Nesse sentido, a integração de empresas a esses parques possibilita um acesso privilegiado a fornecedores altamente especializados de componentes e serviços.

Usualmente, em determinado país, apenas um número relativamente limitado de regiões dispõe das economias de localização e da massa crítica requeridas para a montagem desses arranjos, o que estimula o processo de concentração espacial de indústrias de alta tecnologia. Nesse sentido, a concentração espacial dessas indústrias por meio da montagem de parques tecnológicos favorece os seguintes processos: (1) o estabelecimento de canais adequados de comunicação entre empresas, universidades e prestadores de serviços tecnológicos especializados; (2) a possibilidade de um monitoramento e controle mais eficaz das ações realizadas por outras empresas atuantes no mesmo tipo de atividade; (3) a consolidação de princípios de confiança nos relacionamentos produtivos e tecnológicos que se estabelecem entre os agentes, engendrados por relações de longo prazo que permitem a consolidação de um quadro de referência sociocultural comum; (4) o envolvimento efetivo de universidades e institutos de pesquisa no esforço tecnológico, por meio da consolidação de uma visão regional do processo de desenvolvimento tecnológico; (5) o envolvimento, no nível local, do governo e instituições semipúblicas com objetivos similares, tais como agências de fomento tecnológico e instituições-ponte dedicadas especificamente à transferência de tecnologias entre a universidade e a indústria.

Apesar da tendência à confluência dos fluxos internos à rede no sentido de empresas de base tecnológica presentes nesse tipo de configuração, a morfologia institucional dos “parques tecnológicos” varia consideravelmente de caso para caso, podendo, inclusive, existir regiões onde convivem arranjos interorganizacionais distintos no mesmo espaço territorial. Em alguns casos, empresas de porte relativamente elevado organizam um sistema de fornecedores de componentes concentrado espacialmente no nível local, como ocorre no caso dos polos tecnológicos de Vale do Silício e de Southern Califórnia. Em outros casos, esses polos tecnológicos baseiam-se essencialmente na aglomeração regional de pequenas empresas, que estabelecem entre si laços cooperativos relacionados como Swiss Jura e Danish Jutland, na Europa. A morfologia institucional desses arranjos pode incluir também instituições financeiras (especializadas na concessão de recursos na forma de *venture capital*), serviços empresariais (de consultoria legal, econômica ou tecnológica), além de fornecedores e clientes de indústrias de alta tecnologia. Há indícios também de que a morfologia institucional dos parques tecnológicos varia consideravelmente de indústria para indústria. No caso específico da indústria de computadores, por exemplo, esses parques tendem a privilegiar relacionamentos interempresas, direcionados ao desenvolvimento e suprimento de componentes e softwares. Já no caso de parques dedicados à geração de produtos biotecnológicos, assumem maior importância os vínculos que articulam empresas de base tecnológica e universidades ou institutos de pesquisa, além da articulação daquelas empresas com empresas especializadas na provisão de *venture capital* e com outros parceiros dotados de maior capacitação econômico-financeira (responsáveis, muitas vezes, pela comercialização dos produtos gerados por empresas de base tecnológica no mercado mais amplo).

tipos de externalidades indutoras de um maior nível de eficiência econômica. Apesar da cooperação produtiva e/ou tecnológica não ser um requisito necessário para a consolidação desse tipo de rede, supõe-se que a sua estruturação estimula um processo de interação local que viabiliza o aumento da eficiência produtiva, criando um ambiente propício à elevação da competitividade dos agentes integrados à rede. Análises sobre essas redes também ressaltam os impactos das articulações entre agentes em termos da geração de efeitos de aprendizado e da dinamização do processo inovativo em escala local ou regional.

A “divisão de trabalho” no interior dessas redes está baseada na geração de economias de especialização na produção de insumos, partes e componentes, visando reduzir os custos de produção e facilitar as adaptações requeridas por uma demanda volátil. Os ganhos técnico-produtivos obtidos estão associados à redução de custos e ao reforço da flexibilidade produtiva, decorrente da possibilidade de ajustar-se à logística de operação da rede em função do comportamento sazonal do mercado.

A provisão de serviços técnicos especializados no nível local também tenderia a se desenvolver como consequência dos vínculos cooperativos estabelecidos entre os membros da rede.

As análises que abordam a constituição de distritos industriais ressaltam três propriedades básicas dessas redes. Em primeiro lugar, destaca-se o papel crucial desempenhado pela presença de economias externas específicas ao espaço territorial onde interagem os agentes, relacionadas à possibilidade de se reproduzirem e difundirem localmente conhecimentos técnicos e qualificações profissionais especializadas que conferem vantagens competitivas para os participantes da rede. Em segundo lugar, destaca-se a presença de um balanceamento permanente de princípios de cooperação e competição entre as empresas participantes, o que resulta na consolidação de regras e normas de conduta que permitem uma redução substancial dos custos de transação com os quais se defrontam os agentes. Em terceiro lugar, destaca-se a existência de um balanço entre regras de interação estritamente mercantis e regras de regulação social estabelecidas no nível local, que se refletem na criação de instituições especificamente dedicadas à resolução de problemas de falhas de mercado e ao reforço de valores baseados em princípios de solidariedade social.

Outro aspecto recorrentemente mencionado como fator de fortalecimento da competitividade de empresas inseridas nessas redes refere-se à realização de um elenco de ações conjuntas e coordenadas pelos agentes que resultam em ampliação da eficiência coletiva proporcionada pela sua consolidação, aspecto ressaltado na análise de Schmitz, 1994, sobre o tema. Essa eficiência coletiva é geralmente associada a um processo dinâmico que permite a redução dos custos de transação e o aumento das possibilidades de diferenciação de produto ao longo do tempo, em virtude do intercâmbio de informações e do fortalecimento de laços cooperativos entre os agentes. Para o incremento dessa eficiência, contribuem também diversos estímulos endógenos que elevam as possibilidades de ganhos competitivos para os membros desse tipo de rede. Destaca-se, em primeiro lugar, a difusão de inovações tecnológicas e organizacionais no nível local. Em segundo lugar, destaca-se a montagem de um sistema eficiente para circulação e difusão de informações relevantes no âmbito dessas redes, principalmente aquelas relativas a inovações tecnológicas e à disponibilidade de insumos fundamentais. Em terceiro lugar, destaca-se a capacidade dos sistemas locais de produção desenvolverem uma capacitação comercial e mercadológica que facilite a antecipação das tendências de comportamento do mercado, viabilizando a rápida introdução de novos produtos em função dessas tendências.

Em termos da sua morfologia, essas redes caracterizam-se pela presença de um certo nível de dispersão dos agentes, envolvendo um baixo nível de hierarquização interna e estando baseadas na especialização funcional de empresas independentes. Desse modo, essas estruturas podem ser caracterizadas como redes policêntricas, nas quais os atores cooperam ou competem entre si de forma voluntária, por meio de um conjunto de relações verticais e horizontais. A coordenação dos fluxos internos às redes envolve diversas alternativas, que refletem a sua diversidade institucional. Em alguns distritos, observa-se claramente uma coordenação promovida por agentes externos, responsáveis pela distribuição dos produtos gerados. Outra alternativa envolve uma coordenação promovida por empresas montadoras localizadas nos estágios finais da cadeia produtiva, que, em função do seu nível de capacitação, também atuam na definição do design dos produtos. É possível também identificar situações em que alguns fornecedores especializados em insumos, equipamentos ou etapas críticas do processo de produção também acabam assumindo um papel importante na coordenação dos fluxos internos à rede.

O processo de coordenação intrarrede caracteriza-se também pela implementação de ações coletivas indutoras de aumentos dos níveis de eficiência. As análises sobre distritos industriais costumam atribuir particular importância a dois tipos principais de ações coletivas. O primeiro deles compreende a montagem de centros prestadores de serviços técnicos especializados nestes distritos. Estes centros funcionam como núcleos geradores de externalidades no nível da rede, que podem ser apropriadas pelos seus membros para gerar um maior nível de eficiência. A segunda alternativa compreende a montagem de associações empresariais em locais que funcionam como núcleo de intercâmbio de informações e competências, operando também como grupos de pressão dos interesses de seus membros no plano político-institucional mais amplo.

A intensidade do intercâmbio de informações entre os agentes integrados a essas redes reflete as suas condições institucionais específicas. É comum que os fluxos informacionais assumam um caráter unidirecional, originando-se dos agentes de distribuição e das empresas montadoras na direção da vasta teia de fornecedores integrados ao arranjo. Esses fluxos envolvem a predefinição do *design* e de outros atributos que devem ser atendidos pelos fornecedores. Devido à simplicidade das informações transmitidas e ao caráter não sistemático do processo de transmissão, ele raramente requer a criação de uma infraestrutura particular ou a definição de protocolos específicos que facilitem a comunicação. Isto não significa que, em determinadas redes, não possa haver um intercâmbio horizontal de informações mais intenso, devido à presença de associações empresariais e centros de prestação de serviços técnicos especializados. Além disso, em alguns distritos destaca-se a presença de fornecedores especializados que desempenham um papel importante na disseminação de informações relativas às práticas operacionais que se fazem necessárias para ampliação dos níveis de eficiência das diversas empresas integradas à rede.

A consolidação dessas redes é também responsável pela intensificação de mecanismos de aprendizado por interação em escala local que reforçam a eficiência produtiva de seus membros. Esse aprendizado por interação é afetado pela intensidade do intercâmbio de informações entre agentes e pela infraestrutura institucional subjacente. Em primeiro lugar, este tipo de aprendizado induz o fortalecimento das competências dos fornecedores, facilitando o aperfeiçoamento do *design* de produtos e componentes produzidos localmente. Em segundo lugar, esse aprendizado facilita a difusão de padrões técnicos mais sofisticados, reduzindo assimetrias entre fornecedores quanto no nível de eficiência produtiva. Outro aspecto importante refere-se à difusão de padrões de controle de qualidade mais rígidos por meio dessas interações, assim como de técnicas organizacionais mais modernas que aumentam a produtividade ao longo da cadeia de suprimento. Finalmente, a interação entre agentes desempenha um papel importante na coordenação de ações coletivas dos membros da rede, reforçando os compromissos mútuos estabelecidos entre eles.

A intensidade dos esforços inovativos realizados no interior dessas redes varia em função do grau de sofisticação tecnológica dos produtos gerados. No caso de produtos pouco sofisticados, os esforços inovativos assumem um caráter não sistemático, envolvendo inovações incrementais baseadas em mecanismos de aprendizado que emergem como subproduto das práticas produtivas adotadas. Os próprios esforços inovativos de caráter formal – consubstanciados em gastos de P&D – adquirem especificidades nesse tipo de rede. Geralmente, esses esforços estão baseados numa centralização das atividades de *design* em determinadas empresas – montadoras, fornecedores especializados ou agentes distribuidores – envolvendo melhorias incrementais nos produtos gerados e variações dos componentes que devem ser atendidas pelos fornecedores. A capacidade de realizar aperfeiçoamentos constantes no *design* dos produtos gerados, inclusive visando à conquista de novos mercados, é um aspecto que distingue os distritos industriais mais dinâmicos daquelas que envolvem uma mera parcelização de tarefas visando à redução de custos. É possível também identificar nichos mais dinâmicos do mercado atendido por algumas dessas redes, que trabalham com produtos mais sofisticados do ponto de vista tecnológico. Nesses segmentos de mercado, as atividades de *design* requerem um nível de capacitação mais elevado, sendo centralizadas em empresas montadoras que coordenam uma teia de fornecedores com base na definição de rígidos padrões de qualidade a serem atendidos.

16.5.3 REDES TECNOLÓGICAS

Estas redes estão associadas a inter-relacionamentos cooperativos entre empresas e agentes inseridos na infraestrutura científico-tecnológica, integrando múltiplas competências e viabilizando a exploração de oportunidades tecnológicas promissoras. É possível correlacionar o conceito de rede tecnológica⁵ à situação em que a configuração organizacional envolve a articulação de agentes autônomos e interdependentes, objetivando a introdução de uma inovação no mercado e/ou a sua difusão extensiva pelo tecido industrial, cuja interação resulta na consolidação de mecanismos de “coordenação coletiva” das decisões tomadas. Na medida em que a estruturação desse tipo de configuração associa-se a um objetivo concreto – a viabilização do processo inovativo – a sua morfologia institucional adquire um caráter essencialmente transitório, modificando-se paulatinamente em função dos requisitos impostos pela dinâmica inerente ao processo inovativo. Devido à multiplicidade de atores envolvidos, a consolidação delas muitas vezes não é um fenômeno “natural”, destacando-se a importância de programas tecnológicos que integram os agentes e suas respectivas ações, os quais podem ser montados a partir de agências governamentais. A princípio, a montagem dessas redes não requer uma aglomeração espacial de unidades nem se associa diretamente a indústrias particulares, e sim a determinadas tecnologias cuja base de conhecimento adquire um caráter nitidamente multidisciplinar.

A estruturação de redes tecnológicas evidencia que determinadas articulações interorganizacionais são particularmente importantes para impulsionar o processo inovativo e alavancar a criação de novos mercados. Em particular, a formação dessas redes se mostra funcional nas seguintes circunstâncias: (1) existindo forte interdependência e complementaridade entre as competências dos agentes; (2) no caso de contratos cujos resultados não podem ser identificados e repartidos *ex-ante* (contratos incompletos); (3) no caso de atividades cujo esforço de P&D assume um caráter essencialmente multidisciplinar, integrando profissionais de diferentes áreas do conhecimento científico e tecnológico; (4) no caso de inovações cujos direitos de propriedade não estejam claramente estabelecidos; (5) no caso de conhecimentos de caráter tácito que não são facilmente transferíveis entre os agentes, demandando algum tipo de requalificação e cooperação; (6) em contextos sujeitos à elevada incerteza mercadológica e tecnológica, o que amplia os riscos e custos inerentes ao esforço inovativo. As redes tecnológicas podem também ser percebidas como um arranjo organizacional particularmente funcional no que se refere à adequação ao caráter sistêmico do processo inovativo. Dois aspectos principais caracterizam essas redes: (1) o estabelecimento de conexões entre os agentes diretamente responsáveis pela realização de atividades de P&D e aqueles envolvidos com a produção industrial de bens que incorporam as tecnologias geradas; (2) o estabelecimento de uma divisão de trabalho nos diferentes estágios

do ciclo de P&D, caracterizado pela presença de mecanismos de retroalimentação (*feedback loops*) entre as etapas realizadas. Considerando estes aspectos, dois objetivos principais podem ser associados à consolidação deste tipo de rede. Em primeiro lugar, a integração de competências permite reduzir o tempo do desenvolvimento de novas tecnologias. Em segundo lugar, essas redes desempenham um importante papel para o fortalecimento de ligações entre agentes, induzindo progressivamente a consolidação de uma cadeia produtiva mais bem estruturada.

Os relacionamentos estabelecidos entre as empresas no interior dessas redes estão geralmente associados à divisão de tarefas no ciclo P&D-produção. A estrutura interna dessas redes baseia-se em definição relativamente fluida de tarefas e responsabilidades, de acordo com os requerimentos associados à integração de conhecimentos nos diferentes estágios do processo de P&D. Em termos da conformação institucional, observa-se a presença nessas redes de empresas de base tecnológica que geram efeitos de transbordamento (*spin-offs*) indutores do processo de capacitação no âmbito destes arranjos. Muitas vezes, essas empresas são originárias do próprio meio universitário, integrando profissionais de perfil nitidamente acadêmico que decidem redirecionar sua qualificação no sentido de aplicações industriais. Além disso, é comum a presença, nesse tipo de arranjo, de instituições-ponte entre a ciência e a indústria.

Observa-se também uma tendência à realização de ajustes na estrutura dessas redes, em função da evolução do ciclo de vida das tecnologias e produtos gerados, o que confere um caráter mais fluido a elas. Na fase de consolidação dessas redes, a realização de projetos de P&D coordenados por empresas de base tecnológica é particularmente relevante. Nessa fase, destaca-se também a coordenação baseada em programas cooperativos de caráter público ou semipúblico. Ao longo dos diferentes estágios do ciclo de vida das tecnologias geradas, o grau de centralização da estrutura tende a aumentar, com a rede evoluindo de um formato em que predominam instituições dedicadas a atividades de pesquisa para um formato no qual assumem maior proeminência atividades diretamente produtivas. Em consequência, é comum que ocorra uma progressiva formalização dos mecanismos de coordenação e da base contratual que orienta as relações entre empresas, a qual se combina com uma maior centralização da estrutura da rede. Como reflexo desse processo, é possível mencionar a inter penetração dos direitos de propriedade entre empresas de base tecnológica e grandes empresas localizadas nos setores finais e usuários das tecnologias geradas, funcionando como mecanismo de incentivo que favorece a sustentação de determinadas formas de cooperação.

Os fluxos de informação que conectam os diferentes agentes integrados a esse tipo de rede são usualmente bastante complexos, estando associados a mecanismos de retroalimentação (*feed-back loops*) entre diferentes estágios do processo de P&D. É possível associar esses fluxos a uma transmissão de estímulos semelhante àquela retratada no modelo de Kline-Rosenberg, conforme será visto no Capítulo 19. Em geral, esses fluxos estruturam-se a partir de relacionamentos interpessoais entre indivíduos ou grupos envolvidos no processo de pesquisa, a partir dos quais é possível viabilizar a transmissão de conhecimentos tácitos. Desse modo, a rede desempenha um importante papel como estrutura facilitadora da codificação de conhecimentos associados a diferentes quadros cognitivos e a diferentes campos técnico-científicos. Essa codificação permite a transferência de conhecimentos gerados em ambientes – universidades, institutos de pesquisa, empresas etc. – nos quais as ações dos agentes baseiam-se em quadros de referência distintos.

Considerando essas características, a capacidade de processamento de informações dessa rede contempla, em geral, três aspectos. O primeiro deles refere-se à identificação de informações científico-tecnológicas relevantes, a partir das quais oportunidades atrativas de investimento podem ser vislumbradas. A montagem dessas redes também facilita o acesso de seus membros a informações relevantes para a solução de problemas enfrentados no processo de desenvolvimento. O segundo aspecto refere-se à criação de mecanismos próprios de codificação do conhecimento no âmbito da rede, que tornam mais fácil o intercâmbio de informações entre as esferas científica e industrial. O terceiro aspecto refere-se à possibilidade de integração de diferentes conhecimentos, visando impulsionar o processo inovativo. Por fim, a importância dos conhecimentos tácitos na viabilização do esforço inovativo implica uma necessidade de contatos diretos entre agentes inseridos em grupos de pesquisa. O uso de sistemas dedicados com grande capacidade de processamento de informações pode ser importante em etapas específicas do trabalho de P&D, mas não é necessariamente um requisito para o intercâmbio de informações entre agentes envolvidos em esforços tecnológicos conjuntos.

Um dos aspectos-chave dessas redes refere-se à criação de conhecimentos intencionalmente desenvolvidos a partir da cooperação entre agentes. A criação desses conhecimentos envolve a realização de um esforço conjunto de P&D entre os membros da rede. Nesse sentido, o aprendizado interativo entre agentes inseridos nestes arranjos associa-se à redução dos custos e do tempo do processo de P&D. O aprofundamento desse aprendizado tem também duas outras importantes implicações. Em primeiro lugar, ele facilita a definição das “condições de apropiabilidade” relacionadas às inovações geradas. Em segundo lugar, essa interação viabiliza a definição de códigos de linguagem e comunicação entre agentes que operam a partir de quadros de referência distintos, facilitando o intercâmbio de informações e a integração de competências.

Notas

1. De fato, as redes de firmas se confundem e se sobrepõem a diversos outros tipos de estruturas, tais como: (1) sistemas técnico-produtivos baseados em relações interindustriais recorrentes entre empresas; (2) redes de comunicação estruturadas para facilitar um intercâmbio sistemático de informações entre agentes; (3) redes sociais baseadas em relacionamentos interpessoais sistemáticos de caráter informal entre agentes.
2. Muitas vezes essas tipologias apresentam um caráter essencialmente descritivo, restringindo seu poder explicativo à especificidade dos casos retratados, sendo problemática, a partir destas descrições, a identificação de elementos comuns aos vários tipos de estruturas, que permitam realizar comparações entre elas.
3. *Computer Aided Design*.
4. Para uma sistematização de análises com base nesse tipo de recorte ver, por exemplo, Nadvi e Schmitz (1994), Schmitz e Musyck (1994), Pyke e Sengenber (1992), Pyke (1994), bem como o número especial do periódico World Development de setembro de 1999.
5. Callon *et al* (1992), da escola da sociologia da inovação, criaram o conceito alternativo de rede técnico-econômica.

Bibliografia

- Antonelli, C. (1995) "Economie des réseaux: variété et complémentarité", in: Rallet, A.; Torre, A. (org.) "Économie Industrielle et Économie Spatiale", *Economica*.
- Arcangeli, F.; Belussi, F.; Grunin, V. (1995) *Towards the 'penelope' firm: retractile and reversible networks* Periódico apresentado no Third ASEAT Conference "Managing New technologies into the 21th Century", Manchester, p. 6-8 setembro.
- Axelsson, B.; Easton, G. (orgs.) (1993) *Industrial Networks: a new view of reality*, Routledge, Londres.
- Bandt, J. (1989) "Aproche meso-économique de la dynamique industrielle", *Revue d'Economie Industrielle*, n. 49, 3º trimestre.
- Bidault, F. (1993) "Apprentissage et réseaux", *Economies et Sociétés – Série Dynamique technologique et organisation*, v. 1, n. 5, p. 79-101 maio.
- Britto, J. (1999) "Características estruturais e modus operandi das redes de firmas em condições de diversidade tecnológica", Dissertação de Doutorado, IE-UFRJ.
- Callon, M.; Laredo, P.; Rabeharison, V.; Gonard, T.; Loray, T. (1992) "The management and evaluation of technological programs and the dynamics of techno-economic networks: the case of the AFME", *Research Policy*, v. 21, p. 215-236.
- Cooke, P.; Morgan, K. (1998) *The associational economy: firms, regions and innovation*. Oxford: Oxford University.
- Curien, N. (2000) "Économie des réseaux", Repères, La Decouverte, Paris.
- Dussage, P.; Garette, B. (1991) "Alliances stratégiques – mode d'emploi", *Revue Française de Gestion*, set.-out., p. 4-18.
- Economides, N. (1996) "The economics of networks", *International Journal of Industrial Organization*, v. 14, n. 2, março.
- Freeman, C. (1991) "Networks of innovators: a synthesis of research issues", *Research Policy*, n. 20, p. 499-514.
- Fruin, W. M.; Nishiguchi, T. (1993) "Supplying the Toyota Production System: intercorporate organizational evolution and supplier subsystems", in: Kogut, B. (org.) *Country Competitiveness: Technology and the Organizing of Work*, Oxford University Press.
- Garofoli, G. (1993) "Economic Development, Organization of Production and Territory", *Revue d'Economie Industrielle*, n. 64, 2º trimestre.
- Grabher, G. (org.) (1993) "The embedded firm: on the socioeconomics of industrial networks", Routledge, Londres e Nova York.
- Grandori, A.; SODA, G. (1995) "Inter-firm Networks: Antecedents, Mechanisms and Forms", *Organization Studies*, v. 16, n. 2, p. 183-214.
- Granovetter, M. (1985) "Economic action and social structure: the problem of embeddedness", *American Journal of Sociology*, v. 91, p. 3.
- Hagedoorn, J.; Schakenraad, J. (1993) "Technology partnering and corporate strategies", in Hughes, K. (org.) "European Competitiveness", Cambridge University Press.
- Harrison, M.; Storper, B. (1991) "Flexibility, hierarchy and regional development: the changing structure of industrial production systems and their forms of governance in the 1990s", *Research Policy*, n. 20, p. 407-422.
- Johansson, B.; Karlsson, C.; Westin, L. (orgs.) (1994) "Patterns of a Network Economy", Springer-Verlag.
- Kline, S. J. (1986) "Innovation is not a linear process", *Research Management*, julho-agosto, p. 36-45.
- Langlois, R.; Robertson, P. (1995) "Firms, Markets and Economic Change – a dynamic theory of business institutions", Routledge, Londres e Nova York.
- Markusen, A. (1995) "Sticky places in slippery space: a typology of industrial districts", *Economic Geography*, p. 293-313.
- Piore, M.; Sabel, C. (1984) *The Second Industrial Divide*, Nova York, Basic Books.
- Porter, M. (1998) "Clusters and new economics of competition", *Harvard Business Review*, nov.-dez., p. 77-90.
- Prochnick, V. (1996) *Redes de firmas em setores intensivos em tecnologia no Brasil*. Dissertação de Doutorado. Coppe-UFRJ, dezembro.
- Pyke, P. (1994) "Small firms, technical services and inter-firm cooperation", Research Series n. 99, International Institute for Labour Studies, ILO, Genebra.
- Pyke, P.; Sengenberger, W. (orgs.) (1992) *Industrial districts and local economic regeneration*, International Institute for Labour Studies, Genebra.
- Sabóia, J. (2000) *Desconcentração Industrial no Brasil nos Anos 90 um enfoque regional*, Mimeo.
- Saxenian, A. (1994) *Regional Advantage: culture and competition in Silicon Valley and Route 128*. Cambridge, Harvard University.
- Schmitz, H. (1997) "Collective efficiency and increasing returns", *IDS Working Paper 50*, março.
- Schmitz, H.; Musyck, B. (1995) "Industrial districts in Europe: policy lessons for developing countries", *World Development*, v. 23, n. 1, p. 9-28, 1995.
- Schmitz, H.; Nadavi, K. (1994) "Industrial clusters in less developed countries: review of experiences and research agenda", IDS Discussion Paper, University of Sussex, janeiro.
- Thompson, G.; Frances, J.; Levacic, R.; Mitchell, J. (1991) *Markets, Hierarchies and Networks*, Sage Publications, Londres.
- Yoshino, M.; Rangan, U. (1995) *Strategic Alliances: an entrepreneurial approach to globalization*, Harvard Business School Press, Boston.

A Empresa Transnacional

Reinaldo Gonçalves

17.1 Introdução

Os desenvolvimentos teóricos recentes a respeito da empresa transnacional (ET) ainda são ignorados ou minimizados. Há os que ignoram esses desenvolvimentos, ao mesmo tempo em que se entusiasmam com os avanços ocorridos em áreas como teoria dos jogos, equilíbrio geral, crescimento endógeno etc.¹ A minimização, por seu turno, ocorre em decorrência do viés formalista que contagiou a Ciência Econômica e que associa formalização (geralmente, a matemática) com conhecimento científico.²

A importância da ET é evidente. No contexto do atual processo de globalização, a ET é o principal *locus* de acumulação e de poder econômico a partir do seu controle sobre ativos específicos (capital, tecnologia e capacidades gerencial, organizacional e mercadológica).³ No capitalismo contemporâneo, a ET encontra-se cada vez mais identificada com a categoria de grupo econômico do que com a de empresa.⁴ O grupo econômico (ou a grande empresa) tem sofrido transformações importantes (ver Quadro 17.1 para alguns fatos estilizados). Na dimensão organizacional, passou-se da estrutura piramidal para a de rede, o foco deslocou-se do interno para o externo, e o alcance transferiu-se do mercado doméstico para o mercado global.

A expressão empresa transnacional tem tido uso mais corrente que empresa multinacional desde os anos 1970 do século passado. Essa distinção surgiu em decorrência do debate sobre a criação de empresas multinacionais no âmbito de esquemas regionais de integração econômica envolvendo países em desenvolvimento. Nesse sentido, a expressão “multinacional” estaria reservada a empresas formadas por associações entre empresas de países em desenvolvimento (inclusive, com forte presença de associações e parcerias entre empresas estatais) com atuação regional, ao passo que a expressão “transnacional” estaria referenciada às grandes empresas originárias dos países desenvolvidos com atuação em escala global.

Este capítulo trata da natureza da empresa transnacional. Não há dúvida de que houve importantes desenvolvimentos teóricos, no âmbito da microeconomia, a partir do final dos anos 1970. No entanto, procuramos transcender a dimensão estritamente microeconômica da ET, ainda que esta dimensão esteja no centro da nossa análise.

A ET é a empresa de grande porte que controla ativos em pelo menos dois países. A ET responde pela quase totalidade dos fluxos de investimento externo direto (IED). A teoria desenvolvida neste capítulo assenta-se na relação entre a empresa, o território e o sistema econômico. A ET é entendida como o principal agente dos processos de internacionalização da produção, centralização e concentração do capital, e destruição criadora. Ocorre que a ET também é o resultado da interação desses processos. A ET é sujeito e objeto dos processos de internacionalização da produção, centralização e concentração do capital, e destruição criadora.

QUADRO 17.1 DA EMPRESA LOCAL À EMPRESA GLOBAL: QUE DIFERENÇA UM SÉCULO FAZ

Características	Século XX	Século XXI
Organização	Piramidal	Em rede
Foco	Interno	Externo
Estilo	Rígido	Flexível
Fonte de força	Estabilidade	Mudança
Estrutura	Autossuficiente	Interdependente
Recursos	Átomos – ativos físicos	Bits – informação
Operações	Integração vertical	Integração virtual
Produtos	Produção em massa	Customização em massa
Alcance	Doméstico	Global
Resultados financeiros	Trimestrais	Em tempo real
Controle de estoque	Mensal	Por hora
Estratégia	Do topo à base	Da base ao topo
Liderança	Dogmática	Por inspiração
Trabalhadores	Empregados	Empregados e terceirizados
Expectativa no emprego	Segurança	Crescimento pessoal
Motivação	Para competir	Para construir
Melhorias	Incrementais	Revolucionárias
Qualidade	A melhor possível	Sem compromisso

Fonte: Revista Business Week, apud Dinheiro, 13 de setembro de 2000, p. 53.

Na dimensão microeconômica, a ET é o agente de realização do processo de internacionalização da produção. Esse processo ocorre sempre que residentes têm acesso a bens e serviços com origem em não residentes (isto é, mercadorias originalmente produzidas por não residentes). E ele se expressa de três formas: exportação, licenciamento de ativos e investimento externo direto. O território é o território usado (base da produção e das trocas materiais), que impõe às empresas um conjunto de fatores que influenciam ou determinam sua localização espacial.⁵ O sistema é o capitalismo, com sua macrodinâmica de acumulação em escala global.⁶

O argumento central deste capítulo é que a natureza da ET é tão complexa, e os seus determinantes operariam em condições tão diversas e seriam tão heterogêneos, que não parece ser possível a construção de uma teoria geral. Todavia, isto não nos impede de tentar desenvolver uma visão abrangente e uma moldura analítica para a compreensão do fenômeno da ET.

O conceito de *abrangência* significaria a adoção de um enfoque que incorpore variáveis associadas à própria ET e aos territórios (espaços nacionais), bem como fatores relacionados ao *modus operandi* do sistema capitalista. Procuramos ir além de interpretações que restringem a análise a planos relativamente elevados de abstração, concentrando-se unicamente em elementos sistêmicos. Ao mesmo tempo, procuramos superar as deficiências de interpretações que restringem o escopo de análise, concentrando-se em elementos microeconômicos. Assim, a nossa análise abrangente significaria um esforço de integração de elementos de natureza micro e macroeconômica e, mais precisamente, a integração entre empresa, território e sistema econômico.

Na Seção 17.2 discutimos uma interpretação microeconômica, segundo a qual fatores que operam no âmbito das empresas determinariam a natureza da ET e o fluxo de IED. Na Seção 17.3 analisamos os fatores locacionais específicos, isto é, as variáveis relativas ao território (espaços nacionais), que determinariam as formas do processo de internacionalização da produção. Na Seção 17.4 desenvolvemos uma interpretação sistêmica da internacionalização da produção e da natureza da ET a partir de processos próprios à dinâmica de acumulação capitalista. Na Seção 17.5 apresentamos questões relativas à macrodinâmica capitalista. A última seção apresenta uma síntese da visão abrangente a respeito da natureza da ET.

17.2 Empresa Transnacional: Interpretação Microeconômica

Há uma ampla literatura que analisa os fundamentos microeconômicos da ET, bem como resenhas úteis desta literatura.⁷ A primeira contribuição importante, utilizando o enfoque da organização industrial para o estudo dos determinantes do IED, foi dada por Penrose em 1956.⁸ O argumento básico de Penrose é que o IED é uma consequência do processo de crescimento da empresa. Ela argumenta que as empresas têm uma forte tendência para se expandir, e expansão significa diversificação da produção, assim como penetração em novos mercados internos e externos.

Penrose também argumenta que as empresas que são bem-sucedidas têm alguns recursos internos que lhes permitem aproveitar as oportunidades de expansão. Ela assinala que capacidade gerencial, conhecimento tecnológico e inovações constituem os elementos que criam, em grande parte, as condições para expansão da empresa. Assim, coloca-se indireta e pioneiramente a ideia de que uma empresa, a fim de realizar IED, deve possuir alguma vantagem especial.

17.2.1 FATORES ESPECÍFICOS À PROPRIEDADE

A ideia de que uma empresa deve possuir algumas vantagens para poder penetrar em mercados externos por meio do IED foi, na realidade, a contribuição básica de Hymer em 1960.⁹ Segundo este autor, empresas engajam-se em operações externas porque têm algumas vantagens que as empresas do país receptor do investimento não têm. Dessa forma, os custos associados com a operação internacional – devido à falta de conhecimento do novo meio ambiente, discriminação eventual e outros tipos de riscos e incertezas – devem ser compensados por lucros extras derivados dessas vantagens.

Dentre essas vantagens, cabe destacar o uso de tecnologia de produção mais eficiente, de um sistema de distribuição mais eficaz ou, então a fabricação de um produto diferenciado. Talvez o caso mais famoso de vantagem específica à propriedade seja o segredo da fórmula industrial da Coca-Cola (ver Quadro 17.2).

Hymer apresenta outra razão importante para o IED: empresas podem investir no exterior a fim de se antecipar à competição. Ele também assinala como outro objetivo a redução de riscos por meio da diversificação, em termos do número de mercados onde a empresa tem atividades. No entanto, essa explicação não é tão importante quanto aquelas baseadas nos argumentos de vantagens específicas e de investimento defensivo.

A tese básica de Hymer (das vantagens específicas) também foi utilizada por vários outros autores (por exemplo, Balassa, Kindleberger, Caves, Vernon) para explicar diferentes aspectos do comportamento das empresas transnacionais.

Neste ponto, tendo em mente o fato de que o IED é somente uma das formas de se explorar vantagens monopolísticas em mercados externos, deveríamos colocar a principal questão teórica relacionada ao IED: Por que empresas que investem no exterior, devido ao fato de possuírem uma certa vantagem monopolística, não exploram esta mesma vantagem por meio da exportação ou do licenciamento? Esta é, na realidade, a questão-chave da moderna teoria da internacionalização da produção.

17.2.1.1 IED e licenciamento

O primeiro fator importante que leva empresas a preferir o IED, no lugar de concessão de licenças, refere-se ao problema da valorização do ativo intangível (por exemplo, tecnologia). Assim, na medida em que existem expectativas diferentes no que se refere ao retorno obtido por meio do uso deste ativo e às imperfeições no mercado de ativos intangíveis, o concessionário (ou empresa licenciadora) e o concessionário (empresa licenciada) podem avaliar diferentemente os benefícios futuros. A dificuldade

QUADRO 17.2 FATOR ESPECÍFICO À PROPRIEDADE: A FÓRMULA DA COCA-COLA

“Desvendar os segredos nucleares do laboratório americano de pesquisa atômica em Los Alamos soa mais fácil do que uma pesquisa na biblioteca. Já a Coca-Cola há mais de cem anos mantém sob sigilo a fórmula do refrigerante mais popular do planeta. Qual o segredo?

A composição da Coca-Cola é guardada a sete chaves desde 1886, quando John Pemberton produziu o primeiro lote da bebida. A empresa quase nada diz sobre o precioso líquido. A única cópia da fórmula fica trancada num cofre do Sun Trust Bank em Atlanta, informa o porta-voz Trey Paris. Em que agência, ele não diz. Escrita à mão, essa receita só pode ser fornecida a alguém depois do voto favorável do conselho de administração da Coca-Cola, revela Frederick Allen no livro *A fórmula secreta*.

Não significa que ninguém sabe o segredo. Segundo Paris, “um punhado” de funcionários da Coca-Cola sabe a fórmula de cabeça. Quem são, ele não revela. A lenda, de acordo com um executivo da Coca-Cola, é que três pessoas conhecem o segredo, entre elas o presidente da empresa, Douglas Daft (que se recusou a falar). Algumas outras saberiam parte da fórmula. Os poucos felizardos nunca viajam juntos, para garantir que a fonte da Coca-Cola nunca seque.

E por que a ciência não revela pistas? A Coca-Cola contém 17 ou 18 ingredientes. Aí entram o tradicional corante caramelô e açúcar, bem como um mix de óleos conhecido por 7X (segundo boatos, uma mistura de laranja, limão, canela e outros). Destilar substâncias naturais como essas é complicado, já que há milhares de elementos em sua composição, diz Tom Murray, especialista em química analítica da Chemir/Polytech Laboratories. Uma empreitada dessas sairia por cerca de US\$ 100 mil. E ainda assim, Murray não teria como descobrir o volume exato de cada ingrediente.

Uma coisa ninguém deve achar no xarope: cocaína. Embora a fórmula original de Pemberton incluísse uma dose mínima da droga, a Coca-Cola de hoje não contém cocaína. Desde quando? É outro segredo.”

Fonte: Reed Tucker, *Fortune Américas*, 22 de agosto de 2000, p. 6.

em se chegar a um acordo sobre o valor real do ativo afeta a decisão da empresa entre usar ela própria o ativo por meio do IED ou, então, conceder uma licença para uso deste ativo por uma empresa local. Em outras palavras, a decisão é entre apropriação direta dos benefícios da vantagem monopolística, via IED, ou apropriação indireta por meio da concessão de licença a outras empresas.

Ocorre que uma inovação (por exemplo, produto novo) pode estar relacionada a um pacote de conhecimento. Nesse sentido, não é suficiente ser capaz de produzir o novo produto, mas é necessário também saber como servir o mercado. Assim, o sucesso de um novo produto está intimamente associado aos seus aspectos mercadológicos, em termos não somente de publicidade, mas também de capacidade organizacional específica ao mercado deste novo produto. Isso também é verdade no caso, por exemplo, de marcas. O retorno relativo do pacote inteiro pode ser superior à soma dos retornos relacionados a cada um dos seus componentes. Assim, se existe alguma diferença nos retornos esperados pelo concessionário e concessionário, o primeiro no que se refere ao pacote como um todo, e o último somente quanto aos componentes que pretende comprar, é provável que a negociação encontre sérios obstáculos. Portanto, surge a preferência pelo IED.

Vale também mencionar que a transferência de tecnologia, por meio da concessão de licenças, envolve um certo risco em termos de controle. Assim, se algum tipo de conhecimento é transferido para uma dada empresa torna-se necessário proteger o caráter sigiloso e inovador da tecnologia. Espera-se que, com o objetivo de evitar a difusão do conhecimento e, portanto, o desaparecimento de uma vantagem monopolística, as empresas prefiram investir no exterior em vez de conceder licença.

Além disso, existe o risco de que o concessionário, usando a tecnologia, penetre em outros mercados, que de alguma forma eram ocupados pelo concessionário. Por exemplo, um concessionário pode usar os novos métodos de produção (licenciados) e reduzir seus custos e, portanto, obter uma vantagem competitiva, em termos de outros mercados externos nos quais o concessionário ou qualquer de suas subsidiárias esteja operando. Consequentemente, isto prejudicaria a posição de mercado do concessionário. A existência de fortes restrições às exportações em muitos contratos de transferência de tecnologia é a melhor prova deste argumento.

Outro fator importante, que induz empresas a investirem em vez de concederem licenças, origina-se das políticas governamentais orientadas para contrabalançar o poder de barganha da ET no processo de negociação dos contratos de transferência de ativos intangíveis. É possível que uma melhor supervisão do governo sobre esses contratos, com o objetivo de limitar o alcance e reduzir o número de restrições impostas pela ET nesses contratos, tenda a afetar as decisões dessas empresas na direção do IED ou, pelo menos, desestimular a concessão de patentes.

Vale a pena assinalar que empresas podem ter alta propensão a evitar riscos, seja porque têm gerência e diretoria conservadoras, seja porque a empresa é de tamanho pequeno e, consequentemente, não tende a ter investimentos significativos no exterior, que requerem um certo volume mínimo de capital. Esses aspectos indicam que empresas pequenas e menos agressivas tendem a ter uma propensão a investir no exterior inferior às de empresas grandes e dinâmicas.

O risco de não recuperação de um ativo específico, após a vigência de um contrato de patente (salvo se associado a uma marca) tende a desestimular a relação contratual.

O IED toma como garantido o tamanho do mercado e sua estabilidade. Quando o mercado do ponto de vista do concessionário não é suficientemente grande e estável, é mais sensato conceder licenças a produtores domésticos, com o objetivo não somente de ter um retorno sobre um ativo intangível que é um bem público dentro da empresa, mas também testar o mercado para investimento futuro. Assim, é evidente a importância do território e sua influência sobre o processo de internacionalização da produção e sobre a natureza da ET.

17.3 Território: Fatores Locacionais

Os desenvolvimentos recentes mais significativos a respeito da natureza da ET têm sido feitos no sentido de integrar a teoria do comércio e do IED num único corpo teórico.¹⁰ A ideia central desses desenvolvimentos teóricos é que comércio, investimento e licenciamento são formas da internacionalização da produção, de modo que os principais determinantes de cada um desses movimentos (por exemplo, tecnologia e dotação de fatores) tendam também a ser os elementos determinantes dos outros movimentos.

17.3.1 IED E EXPORTAÇÃO

O primeiro determinante encontrado na literatura sobre IED, que tem suas raízes na teoria tradicional do comércio internacional, está relacionado às restrições ao movimento de fatores. O diferencial de dotação de fatores é a variável-chave. A existência

de amplas fontes de matéria-prima (*resource seeking*), assim como reduzidos custos de mão de obra (*efficiency seeking*), justificariam o IED em setores intensivos em recursos naturais e trabalho. A capacitação tecnológica também pode ser vista como uma variável de estoque (de fator de produção). Nesse caso, a ET pode ter uma estratégia de busca de ativos específicos como o conhecimento tecnológico e a capacidade mercadológica (*asset seeking*), inclusive, por meio de acordos que envolvam fluxos de IED.¹¹ Além disso, elevados custos de transporte também estimulariam o IED.

A imobilidade internacional dos recursos naturais e os custos de transporte foram determinantes bastante significativos do IED no passado e, mesmo atualmente, encontramos importantes fluxos de IED em subsidiárias com expressiva atividade de exportação, inclusive orientadas para mercados regionais, como o Nafta (Acordo de Livre Comércio Norte-americano) e o Mercosul (Mercado Comum do Cone Sul). No entanto, o diferencial na dotação de fatores não parece ser responsável por uma parcela significativa do total do IED na indústria de transformação. No passado recente, é provável que a estratégia de busca de ativos esteja crescendo em importância tendo em vista a aceleração do progresso técnico.

A política comercial também tem sido mencionada na literatura há algum tempo (*market seeking*). No caso dos países em desenvolvimento, quando há medidas orientadas para a substituição de importações (tarifas, quotas etc.), o comportamento das empresas que exportam para esses países é influenciado por tais medidas. O processo de substituição de importações tende, de modo geral, a mudar a forma da internacionalização (do comércio para o IED). Experiência mais recente mostra, também, que o IED japonês na Europa e nos EUA foi, em certa medida, determinado pelo neoprotecionismo de meados dos anos 1970 até o final dos anos 1980.¹²

A ET que opera em oligopólio diferenciado, por meio do seu conhecimento único sobre a comercialização e adaptação do produto, é capaz de se apropriar de uma quase-renda. Isso é verdadeiro para bens de consumo, bem como para bens de produção em que a existência de serviços de manutenção é importante. A empresa localizada no próprio mercado tem maior possibilidade de prever corretamente os resultados da introdução de novos produtos do que uma empresa localizada no exterior. Isto significa que se todos os fatores pró e contra forem tomados em consideração, teríamos um estímulo ao IED.

A estrutura de mercado internacional pode, também, ser determinante. Na medida em que existe uma interdependência de empresas em estruturas oligopolísticas, uma reação por parte das outras empresas é esperada quando um dos membros do oligopólio penetra num mercado particular. Se supomos que a presença da ET está associada a estruturas oligopolísticas de mercado, devemos esperar que alguns investimentos originem-se deste padrão de comportamento – reação oligopolística – criado pela interdependência.¹³

A necessidade de se controlar os sistemas de distribuição em um dado território também surge como um impulso para se investir no exterior. Em alguns tipos de produto e mercado, o problema da disponibilidade do produto em tempo exato e quantidades corretas é muito importante para o desempenho da empresa. Outro aspecto é a necessidade de se ter controle sobre o mercado, de modo que uma previsão mais acurada da demanda seja feita, e, consequentemente, isso induziria ao IED. Vale também mencionar que atitudes do tipo “compre o produto nacional” existente em alguns mercados, de forma que consumidores tenderiam a dar maior valor aos produtos nacionais do que aos estrangeiros, poderiam também ser determinantes do IED.

Vale mencionar, ainda, o modelo do ciclo do produto. Esse modelo foi desenvolvido por Vernon e outros autores, na década de 1960, para explicar a relação entre IED e exportação de novos produtos desenvolvidos nos Estados Unidos.¹⁴ De forma resumida, o argumento é que os produtos passam por três fases distintas. A primeira é a do produto novo. Nesta fase, a produção está localizada em países desenvolvidos, onde está concentrada grande parte das atividades de pesquisa e desenvolvimento tecnológico. De modo geral, nesta fase são pouco expressivos os fluxos de investimento externo para localização de plantas em outros mercados e, portanto, a internacionalização da produção expressa-se, sobretudo, por meio das exportações da empresa inovadora a partir do seu país de origem.

A segunda fase é a do produto em maturação, caracterizada por um produto mais padronizado. Nesta fase, as economias de escala começam a ser importantes e surge uma preocupação maior quanto aos custos de produção. Começam, então, a surgir ameaças à posição estabelecida da empresa inovadora em decorrência da importância crescente dos produtos substitutos e dos diferenciais de custo. A decisão entre exportar ou fabricar o produto no exterior começa a ser determinada pela concorrência com outras empresas e por fatores locacionais (por exemplo, políticas de substituição de importações). O resultado tende a ser o maior crescimento dos fluxos de IED comparativamente aos de exportação.

A terceira fase é a de completa padronização do produto. Nesta fase, os custos de produção são a base da concorrência em decorrência do avanço dos produtos substitutos. O principal resultado é a realocação internacional da produção tomando como base os preços relativos dos fatores de produção. Naturalmente, os fatores locacionais (por exemplo, tamanho do mercado e política comercial) têm influência crescente na forma da internacionalização da produção. Isto significa que, nesta última fase, o país de origem da empresa inovadora pode se tornar um importador líquido do produto (importações provenientes, inclusive, de subsidiárias da empresa-matriz localizadas em países com custos mais baixos).

O modelo do ciclo do produto não explica a parcela do IED que não é substituidora de importações, assim como não trata dos investimentos que objetivam a fabricação de produtos antigos e completamente padronizados. Finalmente, o modelo desconsidera o fato de que alguns novos produtos são planejados de tal forma que possam penetrar em países diferentes com distintos padrões de consumo.

17.3.2 IMPERFEIÇÕES DE MERCADO E INTERNALIZAÇÃO

Na ótica microeconômica, o fenômeno da ET é visto como uma resposta às imperfeições de mercado. As imperfeições nos mercados, tanto de bens quanto de fatores e, especialmente, as imperfeições relativas a produtos intermediários e conhecimento levam ao desenvolvimento da ET por meio do processo de internalização das vantagens específicas à propriedade. Dessa forma, a empresa prefere, “ela própria fazer” (exportar ou realizar IED) do que “fazer outra empresa fazer” (licenciamento).¹⁵

Há tipos diferentes de imperfeições de mercado que criam incentivos à internalização. A internalização ocorre sempre que os benefícios compensem os custos. Há benefícios da criação de mercados futuros dentro da empresa, o que garante o fornecimento permanente de fluxos de conhecimento, isto é, conhecimento não é uma mercadoria como, por exemplo, o petróleo que tem um mercado futuro. A possibilidade de a empresa ter um sistema discriminatório de preços de produtos intermediários e de conhecimento permite definir estruturas diferenciadas de custo em diferentes mercados. No caso de produtos intensivos em conhecimento, a internalização permite evitar os custos da barganha bilateral em decorrência das diferentes expectativas quanto ao valor do próprio conhecimento. A internalização reduz a incerteza do comprador, que controla melhor as atividades intraempresa do que as operações realizadas via mercado. Por fim, há o benefício de se minimizar o efeito da política governamental por meio do mecanismo de preços de transferência (subfaturamento e superfaturamento dos produtos importados e exportados). Também chama-se a atenção para os ganhos associados à capacidade para melhorar o controle de qualidade.¹⁶

Os principais custos estariam relacionados aos custos de administração e comunicação causados pela internalização, a questão da discriminação política contra empresas estrangeiras e os custos associados ao mercado externo (operações internacionais com custos marginais crescentes).

Embora haja esforços no sentido de apresentar o enfoque da internalização como uma teoria geral do IED e da ET, seu impacto em termos da consolidação de um paradigma tem sido um tanto limitado. Assim, Caves não dá nenhuma ênfase especial ao conceito de internalização, ao passo que os trabalhos relativos a este enfoque são mencionados como extensões e variantes do enfoque da organização industrial ou dos ativos intangíveis.¹⁷ Kojima, por exemplo, argumenta que os trabalhos recentes relacionados à teoria da internalização tendem a negligenciar os aspectos macroeconômicos do IED, tanto no país investidor, quanto no país receptor do investimento.¹⁸

Há a percepção de que internalização é um “conceito à procura de uma ‘teoria’, visto que permanecem dúvidas importantes referentes à capacidade da síntese emergente de explicar e prever o comportamento da ET”.¹⁹ Um dos principais contribuintes da literatura na área argumentou que “a procura por uma teoria geral da ET tem levado ao ‘esticamento’ de conceitos parciais ou a uma taxonomia cada vez mais complicada. Os desafios à nova ortodoxia são enfrentados pela redefinição de conceitos centrais ou inventários de classificação cada vez maiores”.²⁰

A principal crítica que pode ser feita a esse enfoque eclético do fenômeno da internacionalização da produção que enfatiza as imperfeições de mercado e tenta integrar comércio, investimento e licenciamento de tecnologia numa única estrutura teórica é que, na realidade, o que se tem alcançado é uma taxonomia dos determinantes comuns dos movimentos internacionais de fatores de produção e produtos. De fato, necessitamos ir mais longe e argumentar que não é possível a elaboração de uma teoria geral da ET ou da internacionalização da produção, da mesma forma que não é possível construir uma teoria geral, quer para comércio, quer para IED.²¹

17.4 Sistema Econômico

A nossa proposta é apresentar não uma teoria geral da internacionalização da produção ou da ET (não acreditamos na possibilidade de tal empreendimento), mas sim uma estrutura analítica abrangente, que integre elementos de natureza sistêmica com elementos específicos à propriedade (empresas) e fatores locacionais específicos (países). Isto é, desenvolvemos uma visão que integre elementos macro e microeconômicos.

17.4.1 CONCENTRAÇÃO E CENTRALIZAÇÃO DO CAPITAL

De Marx aprendemos que o capitalismo é essencialmente um processo de acumulação de capital, que não existe capitalismo sem relações econômicas internacionais, e que as condições da produção capitalista e suas crises levam a uma concentração e centralização do capital.²² A expansão da produção capitalista “exige uma concentração simultânea de capital, pois as condições de produção requerem o emprego de capital numa escala maior. A produção capitalista também exige sua centralização, isto é, a absorção dos pequenos capitalistas pelos grandes e sua privação do capital”.²³

Segundo Marx, o aumento da produtividade do trabalho é um elemento intrínseco ao modo de produção capitalista. Este aumento na produtividade é alcançado por meio de uma crescente divisão do trabalho e crescente cooperação dos trabalhadores assalariados no processo produtivo. E ambos dependem da escala de produção. Assim, o desenvolvimento da produtividade do trabalho exige enormes, e cada vez maiores, quantidades de capital. Com o desenvolvimento do sistema capitalista e “com a crescente massa de riqueza que funciona como capital, a acumulação aumenta a concentração daquela riqueza nas mãos de capitalistas individuais, e, em consequência, amplia a base da produção em larga escala e os métodos específicos de produção capitalista”.²⁴ Neste sentido, no que se refere ao desenvolvimento da produção capitalista “uma crescente concentração de capital é, por conseguinte, uma das exigências materiais assim como um dos seus resultados”.²⁵

Dados a competição intercapitalista, o desenvolvimento do capital financeiro e o aumento das escalas mínimas de produção, defrontamo-nos com um processo de centralização de capital. Este processo é reforçado nas crises de superprodução, nas quais os menores capitalistas sofrem mais do que os grandes, tendo em vista o maior poder de mercado bem como a maior capacidade de mobilização de recursos destes últimos. Nesse ponto, vale diferenciar os conceitos marxistas de concentração e centralização do capital. Segundo Marx, centralização “pressupõe uma mudança na distribuição de capital já à mão, e funcionando”, e, portanto, esse processo “não depende de forma alguma de um crescimento positivo na magnitude do capital social. E esta é uma diferença específica entre centralização e concentração, o último sendo somente um outro nome para reprodução numa escala ampliada”.²⁶

Além desses aspectos assinalados por Marx, deveríamos argumentar acerca do crescente papel do progresso técnico na dinâmica do capitalismo contemporâneo. Na medida em que o progresso tecnológico contribui, de forma significativa, para o crescimento econômico, e as exigências em termos de tempo e recursos nesta atividade tornam-se cada vez maiores, poderíamos esperar que as grandes empresas desempenhariam um papel cada vez mais importante no processo de inovação tecnológica devido à maior disponibilidade de recursos nessas empresas. Assim, as grandes empresas com as maiores taxas de crescimento, devido aos esforços de diversificação e produção de inovações, teriam uma crescente importância no desenvolvimento do capitalismo contemporâneo. Na realidade, esta é a segunda parte de nossa análise, que desenvolvemos logo adiante.

É importante mencionar que não precisamos trabalhar com uma forte hipótese subconsumista a fim de entender o investimento externo. Na realidade, a ET pode existir sem transferência de capital na forma de capital-dinheiro. O controle de uma empresa no exterior pode ocorrer por meio da cessão de ativos na forma de tecnologia e capacidades gerencial, mercadológica e organizacional. O problema da “realização do capital” é básico em qualquer estrutura capitalista, mas não deve ser superenfatizado na explicação de certos aspectos da dinâmica do capitalismo, incluindo o entendimento do IED. Deve-se notar, no entanto, que há evidência empírica recente favorável ao argumento, que mostra o efeito da insuficiência da demanda agregada sobre os fluxos de saída de IED do Reino Unido.²⁷

Retornando à questão do processo de concentração e centralização do capital assinalado por Marx, verifica-se que este tem várias dimensões, algumas foram previstas por ele e outras não. Para ilustrar, a expansão global do capitalismo estava presente nos textos clássicos de Marx, enquanto o desenvolvimento desigual e heterogêneo do capitalismo global estava ausente.²⁸ Essas dimensões, bem como a evidência empírica não fazem parte do escopo deste trabalho. A evidência empírica aponta para um processo de concentração e centralização, assinalado por Marx, que se constitui na primeira parte da nossa perspectiva geral da dinâmica capitalista.²⁹

17.4.2 DESTRUIÇÃO CRIADORA

Na análise realizada por Schumpeter da instabilidade econômica inerente ao desenvolvimento capitalista, que ele sempre enfatizou como um processo histórico e orgânico, podemos encontrar elementos estruturais adicionais da dinâmica capitalista.

Segundo Schumpeter, “o impulso fundamental que coloca e mantém a máquina capitalista em movimento vem dos novos bens de consumo, novos métodos de produção ou transporte, novos mercados, e novas formas de organização industrial que a empresa capitalista cria”.³⁰ Na realidade, a realização de “novas combinações” de recursos e forças que aparecem de forma descontínua é a base do conceito schumpeteriano de desenvolvimento.

Assim a introdução de novos produtos ou produtos aperfeiçoados, novos métodos de produção, novas formas de organização, bem como o uso de novas fontes de matérias-primas e componentes e a penetração em novos mercados domésticos e externos, levam à expansão industrial e à mudança econômica. Inovações tecnológicas, abertura de novos mercados e mudanças na organização industrial criam instabilidade no sistema capitalista, na medida em que surgem de forma descontínua. Esse processo de mudanças descontínuas que aparecem substituindo antigas estruturas por novas é chamado por Schumpeter de processo de destruição criadora. Na fase competitiva do capitalismo, até o final do século XIX, este processo derivou-se de inovações realizadas por novas empresas, que substituíram antigas empresas, sob a liderança do empresário, ao passo que no capitalismo monopolista contemporâneo as inovações são principalmente realizadas dentro das grandes empresas.

No capitalismo contemporâneo, que Schumpeter chamou de capitalismo trustificado, a competição de mercado ocorre por meio, não somente da competição de preços, mas, principalmente, de inovações tecnológicas. Assim, empresas que procuram manter ou expandir suas parcelas de mercado necessitam inovar, engajar-se num processo de inovação tecnológica porque, de outra forma, por meio da introdução de novos produtos e novos métodos de produção outras empresas afetarão a estrutura de mercado. Nesse sentido, o desenvolvimento capitalista decorre de um processo de destruição criadora, pelo qual novas combinações substituem antigas combinações. E as posições de mercado monopólicas ou oligopólicas alcançadas por qualquer empresa não têm um caráter permanente devido às incessantes atividades tecnológicas realizadas por outras empresas, isto é, existe uma rivalidade entre empresas dentro ou fora da indústria. Essas incessantes atividades de pesquisa aplicada exigem não somente volumes elevados de recursos financeiros e não financeiros para enfrentar a crescente especialização e complexidade da pesquisa básica existente, como também o compromisso de tempo nas atividades de pesquisa e as incertezas envolvidas nessas atividades.

Esses aspectos fazem com que, no sistema schumpeteriano, as grandes empresas, geralmente com posições de mercado monopólicas ou oligopólicas, sejam os principais agentes do processo de inovação. Suas posições de mercado monopólicas ou oligopólicas capacitam-nas, por meio de um certo controle sobre o mercado, a se apropriar dos ganhos relativos às inovações e, em consequência, os riscos relacionados a esta atividade encontram-se minimizados.

Assim, no sistema schumpeteriano, as grandes empresas, que dominam os mercados no capitalismo contemporâneo e, em geral, têm poderes monopólicos ou oligopólicos nestes mercados, são as principais responsáveis pela realização de inovações e mudanças estruturais. Por conseguinte, as grandes empresas têm o papel mais importante na expansão industrial e no processo geral de crescimento econômico por meio do progresso técnico.³¹

17.5 Macrodinâmica Capitalista

A ET deve ser entendida a partir de uma perspectiva histórica, conforme ressaltado na seção anterior. Isto significa que a ET é parte de um processo histórico caracterizado por uma concentração e centralização do capital, que foi identificado por Marx nos anos 1960 do século XIX. A ET tem suas raízes na primeira fase da internacionalização do capital nos anos 1970 do século XIX – a chamada expansão imperialista –, na qual o capital financeiro desempenhou um papel importante, bem entendido por autores marxistas e não marxistas nos seus aspectos econômicos e não econômicos.³² Na realidade, a ET surge no início da fase de predominância dos interesses do grande capital na última década do século XIX.

A ET também resulta da aceleração do processo identificado por Schumpeter como destruição criadora, principalmente depois da Segunda Grande Guerra, que tem tido como seu principal agente as maiores empresas nos polos do sistema capitalista. Este processo caracteriza-se por incessantes e descontínuos movimentos no progresso tecnológico. Este progresso tecnológico incessante, juntamente com a necessidade de abertura e ampliação de novos mercados domésticos e externos, caracterizam o recente processo de concentração e centralização do capital, e o processo de internacionalização da produção; isto é, o processo de concentração e centralização do capital à escala mundial.

Assim, a ET surge como o resultado de dois movimentos gerais que ocorrem no desenvolvimento do capitalismo. A ideia central é que o capitalismo, ao exigir a abertura de novos mercados domésticos e externos, vai levar à criação e expansão da ET. A natureza da ET é determinada pela convergência do processo de concentração e centralização do capital e do processo de destruição criadora. Estes dois processos, o primeiro mostrado por Marx e o segundo por Schumpeter, convergem no sentido de transformar a ET na quintessência desse movimento, fazendo com que ela esteja entre as maiores empresas do mundo capitalista e, adicionalmente, torne-se “agressor”, o patrocinador do processo de desenvolvimento tecnológico, e o principal responsável pela dinâmica capitalista.

A evidência empírica disponível parece dar suporte, por um lado, à nossa perspectiva marxista-schumpeteriana da macrodinâmica capitalista. Observa-se um movimento de centralização e concentração do capital, com as grandes empresas capitalistas tendo uma crescente participação na produção, por meio seja da maior taxa de acumulação, seja do processo de fusões e aquisições. Adicionalmente, a grande empresa tem tido um papel significativo no processo de inovação tecnológica.

17.6 Conclusão

A nossa visão abrangente da internacionalização do capital consiste em entender o fenômeno da ET como um epifenômeno da própria dinâmica de desenvolvimento capitalista. A ET surge, então, como a resultante da convergência dos processos de concentração e centralização do capital e de destruição criadora. A ET é o principal objeto da macrodinâmica capitalista, mas também é o seu principal sujeito. A ET é a grande empresa capitalista e o principal agente de realização dos processos de internacionalização da produção, concentração e centralização do capital, e de destruição criadora.

A ET tem como característica básica o fato de ser uma grande empresa e possuir vantagens específicas. No entanto, a posse de vantagens específicas à propriedade surge como condição necessária, porém não suficiente, para explicar a internacionalização da produção em termos do seu movimento inicial, das formas assumidas, e da distribuição espacial (escolha de territórios). A “condição de suficiência” seria determinada pela interação de um conjunto de fatores locacionais específicos. Interagindo com os elementos sistêmicos e específicos às empresas (microeconômicos), os fatores locacionais aparecem como variáveis determinantes para a escolha da forma do movimento de internacionalização da produção.

O fato é que os próprios fatores locacionais específicos (operando no território, no espaço-nacional) dependem de fatores sistêmicos, pois eles surgem em função das necessidades de expansão do sistema capitalista em escala mundial. No caso de fatores sistêmicos que afetam os fatores locacionais específicos, o exemplo óbvio é o processo de desenvolvimento dependente dos países do Sul. Nos países do Norte, podemos destacar os gastos militares dos governos nacionais que estimulam o progresso tecnológico.

A visão abrangente da natureza da ET procura ter um caráter histórico na medida em que aceita o fato de que a cada momento a interação de fatores econômicos, políticos, sociais e culturais define um novo conjunto de pesos específicos para as variáveis determinantes. Em consequência, a resultante das variáveis muda de uma forma incessante, o que nos remete a uma análise dinâmica do processo de internacionalização da produção por meio da expansão das empresas transnacionais. Em síntese, a ET é sujeito e objeto dos processos de internacionalização da produção, centralização e concentração do capital e destruição criadora.

Notas

1. Como exemplo recente de omissão, ver Baumol, W. J. (2000) “What Marshall didn’t know: On the twentieth century’s contributions to Economics”, *The Quarterly Journal of Economics*, v. 115, n. 1, fevereiro.
2. Como exemplo de minimização, ver o livro-texto de Krugman, P. R.; Obstfeld, M. (1994) *International Economics. Theory and Policy*, Harper Collins, 3^a edição, 1994, p. 161.
3. Gonçalves, R. (1999) *Globalização e desnacionalização*, São Paulo, Ed. Paz e Terra.
4. Gonçalves, R. (1991) “Grupos econômicos: Uma análise conceitual e teórica”. *Revista Brasileira de Economia*, v. 45, n. 4. p. 491-519.
5. Santos, M. (2000) *Por uma outra globalização*. Rio de Janeiro, Ed. Record, p. 96-97.
6. Chesnais , F. (1994) *La Mondialisation du Capital*. Paris: Syros.
7. Pitelis, C. N.; Sugden, R. (orgs.), (2000) *The Nature of the Transnational Firm*, Londres, Routledge, 2a edição. Jacquemot, P. (1990), *La Firme Multinationale: Une Introduction Économique*, Paris, Economica; e Hood, N.; Young, S. (1979), *The Economics of Multinational Enterprise*, Londres, Longman.
8. Penrose, E. T. (1956) “Foreign investment and the growth of the firm”, *Economic Journal*, v. 66, p. 220-235; reproduzido em Dunning, J. H. (org.) (1972), *International Investment*, Reino Unido, Penguin Books Ltd., p. 243-264.
9. Stephen Hymer (1976) The International Operations of National Firms: A Study of Direct Foreign Investment. (*Tese de Doutorado, MIT, 1960*). Cambridge: The MIT Press, p. 41-42.
10. Dunning, J. H. (1977) “Trade, location of economic activity and the MNE: A search for an eclectic approach”, em Bertil Ohlin et al., *The International Allocation of Economic Activity*, Londres, The Macmillan Press.
11. Michalet, C. A. (1988) “Les accords inter-firmes internationaux: un cadre pour l’analyse”, in Arena, R. (1988) et al., *Traité d’Economie Industrielle*, Paris, Economica, p. 278-290.
12. Dunning, J. H. (1993) *The Globalization of Business*. Londres: Routledge, p. 134-165.
13. Knickerbocker, F. T. (1973) *Oligopolistic Reaction and Multinational Enterprise*. Boston: Harvard University Press.
14. Vernon R. (1966) “The international investment and international trade in the product cycle”. *Quarterly Journal of Economics*, 80, v. p. 190-207.
15. Peter, J., Buckley e Mark Casson (1976) *The Future of the Multinational Enterprise*. Londres: The Macmillan Press.
16. Casson, M. (1982) The theory of foreign direct investment. In Black, J.; Dunning, J. H. (Eds.), *International Capital Movements*. Londres: The Macmillan Press.
17. Richard, E.; Caves (1996) *Multinational Enterprise and Economic Analysis* (2^a edição). Cambridge: Cambridge University Press, cap. 1.
18. Kojima, K. (junho, 1982) Macroeconomic versus international business approach to direct foreign investment. *Hitotsubashi Journal of Economics*, v. 23, n. 1, p. 1-19.

19. McClain, D. (1983) "Foreign direct investment in the United States: Old currents, 'new waves', and the theory of direct investment". In Kindleberger, C. P.; Audretsch, D. B. (Eds.), *The Multinational Corporation in the 1980s* (p. 295). Cambridge: The MIT Press.
20. Buckley, P. J. (1983) "New theories of international business. Some unresolved issues". In Casson, M. (Ed.), *The Growth of International Business* (p. 34). Londres: George Allen and Unwin Ltd.
21. Gonçalves, R. (1997) "A teoria do comércio internacional: Uma resenha", em *Economia. Ensaios*, v. 12, n. 1, dezembro, p. 3-20.
22. Marx, K. (1970) *O Capital*, 3 volumes, Londres, Lawrence & Wishart.
23. Ibid. v. III, p. 246.
24. Ibid. v. I, p. 585.
25. Ibid. v. III, p. 219.
26. Ibid. v. I, p. 586-587.
27. Pitelis, C. N. "The TNC: an all-weather company", em Pitelis e Sugden, *op. cit.*, p. 193-209.
28. Fernandes, L. (1997) "O Manifesto Comunista e a dialética da globalização", in *O Manifesto Comunista 150 Anos Depois*, Rio de Janeiro, Editora Contraponto, p. 109-119.
29. Bottomore, T. (1997) *Dicionário do Pensamento Marxista*, Rio de Janeiro, Ed. Zahar (edição inglesa de 1983), p. 55-56.
30. Schumpeter, J. A. *Capitalism, Socialism and Democracy*, Londres, George Allen & Unwin Ltd., (5a edição, 1976), p. 83.
31. Uma síntese da evidência empírica é apresentada em Gonçalves, R. (1992), *Empresas transnacionais e internacionalização da produção*, Rio de Janeiro, Ed. Vozes, p. 28-31.
32. Cohen, B. J. (1976) *A questão do imperialismo*, Rio de Janeiro, Ed. Zahar. (edição inglesa de 1973).

Bibliografia

- Agarwal, J. P. (1980) "Determinants of foreign direct investment: a survey". *Weltwirtschaftliches Archiv*, v. 116, p. 739-773.
- Aliber, R. Z. (1983) "Money, multinationals and sovereigns". In: Kindleberger, C. P.; Audretsch, D. (orgs.). *The Multinational Corporation in the 1980s*. Cambridge: The MIT Press, p. 245-259.
- Bain, J. S. (1968) *International Differences in Industrial Structure*. New Haven: Yale University Press.
- Balassa, B. (1966) "American direct investment in the Common Market". *Banca Nazionale del Lavoro Quarterly Review*, n. 76, June, p. 134-135.
- Baranson, J. (1970) "Technology transfer through the international firm". *The American Economic Review*, v. LX, n. 2, May.
- Baumol, W. J. (2000) "What Marshall didn't know: On the twentieth century's contributions to Economics". *The Quarterly Journal of Economics*, v. 115, n. 1, February.
- Brash, D. (1966) *American Investment in Australian Industry*. Cambridge, Harvard University Press.
- Brown, M. B. (1972) *Essays on Imperialism*. Londres, Spokerman Books.
- Buckley, P. J. (1983) "New theories of international business. Some unresolved issues". In: Casson, M. (org.), *The Growth of International Business*. Londres, George Allen and Unwin (Publishers) Ltd, p. 34-50.
- Buckley, P. J. [1981] "A Critical review of theories of the multinational enterprise". *Aussenwirtschaft*, v. 36, p. 70-87.
- Buckley, P. J.; Casson, M. (1976) *The Future of the Multinational Enterprise*. Londres, The Macmillan Press Ltd.
- Bukharin, N. (1915) *Imperialism and World Economy*. Londres, The Merlin Press.
- Calvet, A. (1981) "A synthesis of foreign direct investment theories and theories of the multinational firm". *Journal of International Business Studies*, v. 12, p. 43-60.
- Cantwell, J. (1988) "Theories of International Production". *Discussion Papers in International Investment and Business Studies*, n. 122, University of Reading, England.
- Casson, M. C. (1979) *Alternatives to the Multinational Enterprises*. Londres, The MacMillan Press Ltd.
- Casson, M. C. (org.) (1983) *The Growth of International Business*. Londres, George Allen and Unwin (Publishers), Ltd.
- Casson, M. C. (1982) "The theory of foreign direct investment". In: Black, J.; Dunning, J. H. orgs.) (1982) *International Capital Movements*. Londres, The MacMillan Press Ltd.
- Caves, R. E. (1971) "Industrial Economics of foreign investment. The case of the international corporation". *Journal of World Trade Law*, v. 5, n. 3, May-June.
- Caves, R. E. (1971) "International corporations: the industrial economics of foreign investment". In: Dunning, J. H. (org.), *International Investment*. Harmondsworth: Penguin Books Ltd., p. 265-301.
- Caves, R. E. (1974) "Multinational firms, competition and productivity in host-country markets". *Economics*, v. 41, n. 162, May.
- Caves, R. E. (1974) "Industrial Organization". In: Dunning, J. H. (org.) (1974) *Economic Analysis and the Multinational Enterprise*. Londres: George Allen & Unwin Ltd, p. 115-146.
- Caves, R. E. (1982) *Multinational Enterprise and Economic Analysis*. Cambridge University Press.
- Cohen, B. J. (1976) *A Questão do Imperialismo*. Rio Janeiro: Zahar Editores.
- Dunning, J. H. (org.) (1972) *International Investment*. Harmondsworth: Penguin Books Ltd.
- Dunning, J. H. (1973) "The determinants of international production". *Oxford Economic Papers*, v. 25, n. 3, p. 289-336, November.
- Dunning, J. H. (org.) (1974) *Economic Analysis and the Multinational Enterprise*. Londres: George Allen & Unwin Ltd.
- Dunning, J. H. (1977) "Trade, location of economic activity and the MNE: A search for an eclectic approach". In: Ohlin, B. et al. *The International Allocation of Economic Activity*, Londres: MacMillan Press Ltd.
- Dunning, J. H. (1979) "Explaining changing patterns of international production: In defence of the eclectic theory". *Oxford Bulletin of Economics and Statistics*, v. 41, n. 4, p. 269-295.

- Dunning, J. H. (1981). *International Production and the Multinational Enterprise*. Londres: George Allen and Unwin.
- Evely, R. Little, I. M. D. (1960) *Concentration in British Industry*. Cambridge University Press.
- Fajnzylber, F.; (1983) *La Industrialización Trunca de America Latina*. México: ET/Editorial Nueva Imagen.
- Furtado, C. (1967) *Development and Underdevelopment*. University of California Press.
- Galbraith, J. K. (1967) *The New Industrial State*. Boston: Houghton Mifflin.
- Gordon, L.; Grommers, E. (1962) United States Manufacturing Investment in Brazil. The Impact of Brazilian Government Policies 1946-1960. Boston: Harvard University.
- Hilferding, R. (1910) *El Capital Financiero*. Madrid: Editorial Tecnos.
- Hirsch, S. (1976) "An international trade and investment theory of the firm". *Oxford Economic Papers*, v. 28, p. 258-270.
- Hobson, J. A. (1968) *Imperialism. A Study*. Londres: George Allen & Unwin Ltd., (1902).
- Hood, N.; Young, S. (1979) *The Economics of Multinational Enterprise*. Londres: Longman.
- Horst, T. (1972) "Firm and industry determinants of the decision to invest abroad: an empirical study". *The Review of Economics and Statistics*, n. 54, p. 258-266.
- Hufbauer, G. C. (1975) "The multinational corporation and direct investment". In: Kenen, P. B. (org.) *International Trade and Finance. Frontiers for Research*. Cambridge University Press.
- Hymer, S. (1960) The international operations of national firms: a study of direct foreign investment. Ph.D. Thesis, MIT.
- Hymer, S. (1972) "The internationalization of capital". *Journal of Economic Issues*, v. 6, n. 1, March, p. 91-111.
- Hymer, S. (1975) "The multinational corporation and the law of uneven development". In: H. Radice (org.) *International Firms and Modern Imperialism*. Harmondsworth: Penguin Books Ltd., p. 37-62.
- Kindleberger, C. P. (1967) "The international firm and the international capital market". *The Southern Economic Journal*, v. 34, n. 2, outubro.
- Kindleberger, C. P. (1969) *American Business Abroad. Six Lectures on Direct Investment*. Yale University Press.
- Knickerbocker, F. T. (1973) *Oligopolistic Reaction and Multinational Enterprise*. Boston: Harvard University Press.
- Kojima, K. (1982) "Macroeconomic versus international business approach to direct foreign investment". *Hitotsubashi Journal of Economics*, v. 23, n. 1, junho p. 1-19.
- Lall, S.; Streeten, P. *Foreign Investment and Developing Countries*. Londres: The Macmillan Press Ltd.
- Lenin, V. I. (1975) *Imperialism, the Highest Stage of Capitalism*. Pequim: Foreign Languages Press, 1917).
- R. (1971) *The Accumulation of Capital*. Londres: Routledge and Kegan Paul Ltd., (1913).
- Magdoff, H. (1972) "Imperialism without colonies". In: Owen, R.; Sulcliffe, B. (orgs.) *Studies in the Theory of Imperialism*. Londres, Longman Group Ltd.
- Magee, S. P. (1977) "Information and the multinational corporation: an appropriability theory of direct foreign investment". In: Bhagwati (org.), *The New International Economic Order*, Cambridge, Massachusetts: MIT Press.
- Marx, K. (1970) *Capital*, 3 volumes. Londres: Lawrence & Wishart.
- McClain, D. (1983) "Foreign direct investment in the United States: old currents, 'new waves', and the theory of direct investment". In: Kindleberger, C. P.; Audretsch, D. B. (orgs.) *The Multinational Corporation in the 1980s*. Cambridge, Massachusetts: The MIT Press, p. 278-333.
- Orr, D. (1973) "Foreign control and foreign penetration in the Canadian manufacturing industries". University of British Columbia, *mimeo*.
- Ozawa, T. (1975) "Peculiarities of Japan's Multinationalism: facts and theories". *Banca Nazionale del Lavoro Quarterly Review*, n. 115 December, p. 404-426.
- Parker, J. E. (1974) *The Economics of Innovation: The National and International Enterprise in Technological Change*. Londres: Longman.
- Penrose, E. T. *Foreign investment and the growth of the firm*. In: Dunning, J. H. (1972) (org.) p. 243-264.
- Phelps, D. M. (1939) *Migration of Industry to South America*. Greenwood Press Publishers.
- Rugman, A. M. (1980) "Internationalization as a general theory of foreign direct investment: a re-appraisal of the literature". *Weltwirtschaftliches Archiv*, v. 114, n. 2, p. 365-379.
- Schneider, F.; Frey, B. (1985) "Economic and political determinants of foreign direct investment". *World Development*, v. 13, n. 2, p. 161-175.
- Schumpeter, J. A. (1984) *Impérialisme et Classes Sociales*. Paris: Flammarion, (1919).
- Schumpeter, J. A. (1928) "The instability of Capitalism". *Economic Journal*, September, p. 361-386.
- Schumpeter, J. A. (1976) *Capitalism, Socialism and Democracy*. Londres: George Allen & Unwin Ltd. 5^a edição, (1943).
- Shepherd, W. G. (1964) "Trends of concentration in American manufacturing Industry", 1947-1948. *Review of Economics and Statistics*, May.
- Stevens, G. V. C. (1974) "The determinants of investment". In: Dunning , J. H. (org.) *Economic Analysis and the Multinational Enterprise*. Londres: George Allen & Unwin Ltd.
- Teece, D. J. (1983) "Technological and organizational factors in the theory of the multinational enterprise". In: Casson, M. org.) (1983) *The Growth of International Business*, Londres: George Allen and Unwin (Publishers), Ltd., p. 31-62.
- UNCTAD. *World Investment Report*. Genebra: United Nations Conference on Trade and Development. Publicação anual. Disponível: www.unctad.org.
- U.S. Department of Commerce (1975) *Historical Statistics of the United States. Colonial Times to 1970*. Washington D.C.: USGPO.
- Vaupel, J. W. (1971) *Characteristics and Motivations of US Corporations with Manufacture Abroad*. Paris: Atlantic Institute.
- Vernon, R. (1979) "The product cycle hypothesis in a new international environment". *Oxford Bulletin of Economics and Statistics*, v. 41, November, p. 255-267.
- Vernon, R. (1979) "The international investment and international trade in the product cycle". In: Savasini J. A. et al. (orgs.) *Economia Internacional*. São Paulo: Saraiva S.A., p. 89-107.
- Weiss, L. W. (1971) "Quantitative studies of industrial organization". In: Intriligator, M. (org.) *Frontiers of Quantitative Economics*. North-Holland Publishers.
- Wolf, B. M. (1977) "Industrial diversification and internationalisation: some empirical evidence". *Journal of Industrial Economics*, v. 26, n. 2, p. 177-191.



PARTE V

ESTRATÉGIAS EMPRESARIAIS

Concorrência Schumpeteriana

Mario L. Possas

18.1 Introdução

O objeto deste capítulo é a teoria da concorrência esboçada em suas linhas básicas por J. Schumpeter na primeira metade do século XX e aperfeiçoada nas últimas duas décadas por autores da corrente neo-schumpeteriana, em sua maioria também identificados pela abordagem evolucionista ou evolucionária, que vem se desenvolvendo sob inspiração daquele autor e por analogia aos avanços da biologia evolucionista moderna. A característica mais destacada deste enfoque é que, em contraste com o enfoque estático tradicional, a concorrência na economia capitalista passa a ser vista como um processo evolutivo, e portanto dinâmico, gerado por fatores endógenos ao sistema econômico, notadamente as *inovações* que emergem incessantemente da busca de novas oportunidades lucrativas por parte das empresas em sua interação competitiva.

A próxima seção contém uma síntese das três grandes visões de concorrência anteriores à schumpeteriana: a clássica, a marxista e a neoclássica. Por contraste a essas é possível destacar, na Seção 18.3, os principais traços da teoria schumpeteriana da concorrência, incluindo elementos da contribuição neo-schumpeteriana mais recente. A Seção 18.4 procura identificar algumas implicações normativas e de política econômica – especialmente para a política de concorrência – dessas proposições teóricas. Segue-se uma breve conclusão do capítulo.

18.2 Diferentes Visões da Concorrência na Teoria Econômica

Não há propriamente, na tradição da Economia como ciência, nenhuma “teoria da concorrência” anterior ao advento da obra do economista austríaco J. Schumpeter no século XX.¹ Mas há diversas “noções” de concorrência que, mesmo sem constituir alguma teoria sistemática, tiveram um papel importante em relegar a concorrência a uma posição meramente acessória na teoria econômica.

Passemos brevemente em revista as mais importantes dentre essas noções: a *clássica*; a de *Marx*; e a *neoclássica*, ainda hoje dominante – para melhor situar a revolução teórica representada pela abordagem *schumpeteriana*.

18.2.1 A NOÇÃO “CLÁSSICA” DE CONCORRÊNCIA

Como apresentado nos Capítulos 1 e 7, a noção “*clássica*” de concorrência – adotada por Smith, Ricardo e seus contemporâneos – está associada à livre *mobilidade do capital* entre diferentes indústrias, implicando a *livre entrada* (livre iniciativa) ou ausência de “barreiras à entrada”. No início do capitalismo, essas barreiras estavam relacionadas com privilégios monopolistas,

ou seja, restrições institucionais ou legais à livre concorrência e à livre iniciativa. A concorrência era vista como um *processo* que se desenrola ao longo do tempo, pelo qual os investimentos são atraídos pelas indústrias que proporcionam maior taxa de lucro, afastando-se das de menor rentabilidade. Seria esse contínuo fluxo intersetorial de capitais, possibilitado justamente pela *concorrência entre capitais* – ou seja, por sua *mobilidade* entre indústrias –, o responsável pela suposta *tendência à igualação* das taxas de lucro entre distintas atividades nas economias capitalistas.

Entretanto, é característico do enfoque teórico da Economia Clássica, especialmente em sua vertente ricardiana (e, hoje, “neoricardiana”), sua preocupação maior com o *resultado* desse processo – ou seja, com a formação de uma *taxa de lucro uniforme* entre indústrias, e os respectivos preços “naturais” ou de equilíbrio intersetorial – do que com o *processo da concorrência em si*, já prenunciando o viés *estático* da Ciência Econômica tal como viria a se consolidar com a teoria do equilíbrio geral, em fins do século XIX. Nesse quadro, a concorrência não é objeto de análise em si, mas só interessa pelos seus *efeitos tendenciais* ou de *longo prazo*, associados à teoria da *determinação dos preços* e da *taxa de lucro de equilíbrio*.

18.2.2 A CONCORRÊNCIA EM MARX

Para Marx, a concorrência também era considerada mais como um processo auxiliar – embora importante – para atingir determinados fins previstos pela teoria, do que como um objeto em si mesmo digno de ser teorizado. Na concepção de Marx a concorrência não tem o *status* de gerar *por si mesma* efeitos relevantes na economia capitalista; ela cumpre apenas um papel intermediário de “executar” as “leis de movimento” (para usar sua expressão) dessa economia, determinadas em nível mais “fundamental” (das relações de produção e das leis do capital) que o nível “superficial” da concorrência, incapaz de criar ou afetar essas determinações. Isso não exclui o reconhecimento da importância da concorrência, em Marx, como um *pressuposto* para a teoria do valor e do capital; bem como sua aceitação, no essencial, da visão clássica da concorrência como “mobilidade dos capitais”, atuando como o mecanismo básico para a tendência à formação de uma taxa de lucro uniforme.

Por outro lado, é importante notar que Marx *também* tinha uma percepção aguda da concorrência como um mecanismo permanente de introdução de *progresso técnico*, capaz de tornar *endógena* à economia capitalista a capacidade de mudança estrutural via inovações – na sua linguagem, de tornar o “desenvolvimento das forças produtivas” uma “lei de movimento” básica da economia capitalista. Este é um elemento crucial para uma *teoria dinâmica da concorrência*, que será retomado por Schumpeter muito mais tarde.

18.2.3 A NOÇÃO NEOCLÁSSICA DE CONCORRÊNCIA

A concepção clássica foi também adotada – e estendida – por Marshall, um dos pioneiros da tradição neoclássica e fundador da Microeconomia. Sua principal contribuição nesse terreno foi dar contornos mais precisos à noção *neoclássica* de concorrência, predominante ainda hoje. Na sua versão atual, trata-se da noção de *concorrência perfeita*, associada ao atomismo de mercado (tanto na oferta como na demanda), em que as empresas individuais são tomadoras de preço (*price takers*), ou seja, incapazes de afetar o preço de mercado, determinado pelo equilíbrio entre oferta e demanda, com *preço de mercado* igual a seu *custo marginal*. As implicações normativas dessas hipóteses são fundamentais para o enfoque estático de *eficiência alocativa*, predominante em praticamente todas as aplicações de política econômica. Implicações normativas idênticas, mas que dispensam o atomismo de mercado, são extraídas, ainda no campo neoclássico, pela recente teoria de “*mercados contestáveis*” (1982), apresentada no Capítulo 7, cujos resultados podem ter algum interesse teórico, mas cujo interesse empírico, em função de pressupostos pouco realistas, dependem de uma avaliação de suas aplicações.

18.3 A Teoria Schumpeteriana da Concorrência e o Papel das Inovações

18.3.1 A VISÃO DE SCHUMPETER

Em que consiste a noção – mais que isso, a “teoria” – de concorrência proposta por Schumpeter? Trata-se de uma noção não ortodoxa, mas potencialmente a mais interessante de todas, hoje conhecida como *concorrência schumpeteriana*. Sua principal característica é que ela se insere numa visão *dinâmica e evolucionária* do funcionamento da economia capitalista. Por ela, a *evolução* desta economia é vista *ao longo do tempo* (e por isso é *dinâmica e evolucionária*) como baseada num *processo* ininterrupto de introdução e difusão de *inovações* em sentido amplo, isto é, de quaisquer mudanças no “espaço econômico” no qual operam as empresas, sejam elas mudanças nos produtos, nos processos produtivos, nas fontes de matérias-primas, nas formas de organização produtiva, ou nos próprios mercados, inclusive em termos geográficos.

Por sua vez, qualquer *inovação*, nesse sentido amplo, é entendida como resultado da busca constante de *lucros extraordinários*, mediante a obtenção de *vantagens competitivas* entre os agentes (empresas), que procuram se *diferenciar* uns dos outros nas mais variadas dimensões do processo competitivo, tanto os tecnológicos quanto os de mercado (processos produtivos, produtos, insumos, organização; mercados, clientela, serviços pós-venda). No Capítulo 8 tivemos a oportunidade de discutir amplamente esse tema.

18.3.2 CONCORRÊNCIA SCHUMPETERIANA

Os principais traços da teoria schumpeteriana da concorrência, em síntese, são discutidos a seguir.

A concorrência schumpeteriana caracteriza-se pela busca permanente de *diferenciação* por parte dos agentes, por meio de *estratégias* deliberadas, tendo em vista a obtenção de vantagens competitivas que proporcionem *lucros de monopólio*, ainda que temporários.

Por isso mesmo, *concorrência não é o contrário de monopólio*. Se bem-sucedida, a busca de novas oportunidades, ou *inovações* em sentido amplo, deve gerar monopólios, em maior ou menor grau e duração. Se eles serão ou não eliminados eventualmente, por meio de novos concorrentes e/ou imitadores, é algo que não pode ser preestabelecido. Na ocorrência de retornos crescentes à escala, como é frequente, vantagens monopolísticas tendem a se *consolidar*, em vez de desaparecer.

A concorrência é um processo (*ativo*) de *criação* de espaços e oportunidades econômicas, e não apenas, ou principalmente, um processo (*passivo*) de *ajustamento* em direção a um suposto equilíbrio, nem supõe qualquer estado tendencial “normal” ou de *equilíbrio*, como nos enfoques clássico e neoclássico. O desfecho do processo de concorrência não é predeterminado, mas depende de uma interação complexa de forças que se modificam ao longo do mesmo processo – mecanismos dependentes da trajetória (*path dependence*), como são chamados na literatura –, tornando muitas vezes impossível prever a própria existência, que dirá as características de um estado terminal. Esse, por sinal, é um traço típico de processos evolutivos.

Nessa concepção, concorrência implica o surgimento *permanente* e *endógeno* de *diversidade* no sistema econômico capitalista, também como convém a um processo evolutivo. Importa mais a *criação* de diferenças, por meio das *inovações* em sentido amplo, do que sua *eliminação*, mesmo que tendencial, como nos enfoques clássico e neoclássico.

Há muitas *formas* ou *dimensões* da concorrência, sendo a concorrência em preços apenas a mais tradicional e mais simples, mas não a mais importante ou mais frequente. A concorrência se dá também por *diferenciação do produto* (inclusive qualidade) e, especialmente, por *inovações*, que no sentido schumpeteriano – muito amplo, como se viu – envolve toda e qualquer criação de novos espaços econômicos (novos produtos e processos, novas formas de organização da produção e dos mercados, novas fontes de matérias-primas, novos mercados).

Essa ênfase na *diferenciação* dos agentes e na *multiplicidade* dos instrumentos de concorrência e dos ambientes concorrenciais implica destacar a importância da *diversidade* dos fatores *microeconômicos* na caracterização dos esforços e resultados competitivos; em particular, a *diversidade estratégica* e a *variedade tecnológica* como elementos centrais na análise da concorrência.

A *empresa* é a unidade de análise da concorrência schumpeteriana, por ser a unidade de decisão e de apropriação dos ganhos. O *mercado* é o seu *locus*, definido como o espaço de interação competitiva principal entre as empresas (pode haver outros) em sua rivalidade e orientação estratégica; há, portanto, um componente subjetivo – de *avaliação estratégica* – nesta definição de “mercado”. É claro que fatores objetivos relacionados à demanda e à oferta dos produtos e serviços – seu grau de substituibilidade, sua afinidade tecnológica etc. – são critérios que balizam essa definição.

Embora a unidade de análise seja a empresa, as condições *ambientais* são decisivas – seja no nível *de mercado*, onde se dá efetivamente o processo de concorrência, seja no nível mais geral, *sistêmico*, onde se definem as *externalidades* e as *políticas* que afetam a concorrência.

A interação, ao longo do tempo, entre as *estratégias das empresas* – não apenas de inovação, mas também de investimento, de preços etc. – ou seja, *estratégias competitivas*, de um modo geral – e as *estruturas de mercado* preexistentes gera uma *dinâmica industrial*, pela qual a configuração de uma indústria, em termos de produtos e processos (tecnologias) utilizados, de participações no mercado das empresas, de rentabilidade, de crescimento etc., vai se transformando ao longo do tempo.

Assim, as *estruturas de mercado* são relevantes, mas não algo único nem imutável. Tanto podem condicionar, com maior ou menor intensidade, as condutas competitivas e as estratégias empresariais, como podem ser por estas modificadas, de forma deliberada e às vezes até profunda (no caso de inovações chamadas “radicais”, que afetam fortemente

o funcionamento de vários mercados). Tais mudanças devem ser consideradas como inteiramente normais, e não excepcionais, podendo apresentar características evolutivas mais ou menos regulares, como nas situações tratadas pelas noções de ciclo industrial e ciclo de produto. Em outras palavras, essas estruturas são em grande medida *endógenas* ao processo competitivo, e sua evolução deve ser vista no contexto da *interação* dinâmica entre *estratégia* empresarial e *estrutura* de mercado.

Nesse enfoque, concorrência não é um “dado” ou um conjunto de “precondições” – atomismo de mercado, racionalidade otimizadora dos agentes, informação completa – necessários para o equilíbrio competitivo, como na moderna ortodoxia axiomática da teoria econômica neoclássica. Tampouco é um processo de ajustamento a posições de equilíbrio, com *eliminação de lucros anormais* e de desvios considerados fortuitos, como no enfoque clássico e no neoclássico mais tradicional. É, na verdade, um processo de interação entre empresas voltadas à *apropriação de lucros* – ou, em outros termos (não usados por Schumpeter), à valorização dos ativos de capital.

Essa apropriação de lucros *não pressupõe* nem conduz a algum *equilíbrio* – como, por exemplo, a igualação entre taxas de retorno do capital, presente tanto na teoria clássica como na neoclássica. Ao contrário, está relacionada a *desequilíbrios* oriundos do esforço de diferenciação e criação de vantagens competitivas pelas empresas, que se esforçam por retê-las na forma de *ganhos monopolistas*, ainda que temporários e restritos a segmentos específicos de mercado.

Assim, os *lucros não são “normais”* em qualquer acepção – eles são conceitualmente mais próximos de *quase-rendas* (como na acepção de Marshall – rendimentos derivados da *escassez* temporária de determinados ativos) do que de rendimentos de um fator *em equilíbrio*, como na tradição neoclássica. Mais ainda, a própria noção de uma *taxa de lucro “ex post”* – razão entre o lucro obtido e o capital aplicado – *perde o sentido*, uma vez que o lucro resultante de um esforço competitivo bem-sucedido não tem qualquer relação causal com a *magnitude* do capital aplicado, e tampouco há qualquer tendência à homogeneização dessa relação entre os agentes econômicos do mesmo mercado ou de diferentes mercados, *ainda que* haja livre mobilidade do capital entre indústrias.

Da mesma forma, as situações monopolísticas criadas a partir de inovações bem-sucedidas *não* devem ser vistas como intrinsecamente anticompetitivas, pois constituem o objetivo mesmo, e o resultado esperado, do processo competitivo, ainda que de forma temporária e restrita. É por isso que, como já visto, *monopólio* (uma configuração de mercado, às vezes temporária) e *concorrência* (um processo) *não* são incompatíveis entre si – muito *ao contrário!*

18.3.3 A ABORDAGEM NEO-SCHUMPETERIANA

O quadro de referência teórico descrito na seção anterior, tão diferente do ortodoxo, vem sendo retomado e desenvolvido nas últimas quase três décadas por economistas da corrente neo-schumpeteriana, que em sua vertente mais microeconômica também adotam (e são conhecidos por) uma perspectiva “*evolucionária*” (e portanto dinâmica) da teoria e da análise microeconômica. Entre seus principais expoentes destacam-se R. Nelson e S. Winter, que fincaram o marco inicial dessa perspectiva com seu livro clássico *Uma teoria evolucionária da mudança econômica*.²

Esse autores – acompanhados por muitos outros participantes dessa corrente –, além de seguir as principais ideias de Schumpeter a respeito da dinâmica da concorrência e da inovação e sua importância na economia capitalista, propõem romper com os pressupostos metodológicos tradicionais (neoclássicos) – particularmente o de *equilíbrio*, substituído pela noção mais geral de *trajetória*; e o de *racionalidade maximizadora ou substantiva*, substituído pelo de *racionalidade limitada (“bounded”)* ou *processual (“procedural”*, na terminologia adotada por Herbert Simon).

Na analogia *evolucionária* proposta por Nelson e Winter são introduzidas as noções básicas de *busca (“search”*) de *inovações*, procedidas pelas *empresas* a partir de estratégias; e de *seleção (“selection”*) dos resultados econômicos dessas mesmas inovações, realizada pelo *mercado* – o *ambiente de seleção* por excelência – e, secundariamente, por outras instituições (centros de pesquisa, universidades etc.), conforme desenvolvido no Capítulo 8.

Este tipo de análise microdinâmica e os modelos desenvolvidos por esses e outros autores da corrente evolucionária neo-schumpeteriana baseiam-se na interação temporal entre as *estratégias* empresariais, que envolvem o referido processo de *busca* de inovações – mas abrangendo ainda outras estratégias competitivas e decisões (produção, investimento, preços) – e o processo de *seleção* pelo mercado dessas mesmas inovações.

A *trajetória* resultante – a evolução temporal da indústria, em que se vai modificando *endogenamente*, por meio das *inovações* e de sua seleção pelo mercado, a configuração ou a estrutura da indústria em termos de produtos, tecnologias, participações e concentração de mercado etc. – é o principal objeto de análise. Geralmente são utilizadas técnicas de modelagem por *simulação*, em vez de tentar obter soluções analíticas unívocas, como nos modelos neoclássicos tradicionais, que só examinam soluções de equilíbrio e sua estabilidade.

18.3.4 SÍNTESE

Em suma, a concorrência schumpeteriana é uma teoria em que a criação de novas oportunidades lucrativas – a dimensão *ativa* da concorrência, capaz de promover incessantemente *diferenciação* entre os agentes e *transformações* na esfera econômica – é *tão ou mais importante* que a tendência à eliminação de vantagens ou de diferenças entre os agentes – a dimensão *passiva* da concorrência, na verdade *a única* que havia sido implicitamente focalizada em todas as análises anteriores a Schumpeter (com exceção parcial de Marx, como vimos).

O destaque dado no enfoque schumpeteriano ao conceito de *inovações em sentido amplo* reflete essa ideia crucial: não se trata apenas de enfatizar a mudança *tecnológica* – como às vezes se supõe ao interpretar erroneamente, de forma reducionista, a *concorrência schumpeteriana* –, mas *toda e qualquer mudança no espaço econômico*, promovida pelas empresas em busca de vantagens e consequentes ganhos competitivos. É *esta* – a dimensão “ativa” da concorrência, criadora de todo tipo de *variedade* dentro do sistema econômico capitalista, e não eventuais “ajustamentos” (se houver algum) a uma nova posição de equilíbrio, como nas tradições anteriores –, que importa para fundamentar uma *teoria dinâmica* da concorrência capitalista. Isto porque é ela que permite explicar a notável capacidade que a economia capitalista apresenta – *por si mesma* e não por “choques exógenos”, como preferem pensar os economistas ortodoxos – de gerar *mudança qualitativa*, isto é, transformações em todo o espectro de atividades capazes de produzir lucros; o que só é compreensível quando se analisa a concorrência e seus efeitos *ao longo do tempo* – em suma, o *processo (dinâmico) de concorrência* –, em vez de contentar-se com supostos “*estados de equilíbrio*” (análise estática) que, também supostamente, representariam de forma adequada e suficiente o funcionamento dessa economia.

Apesar dos ganhos em realismo teórico do enfoque schumpeteriano, seu conteúdo mais complexo e seu instrumental de análise mais incipiente o tornam um referencial, embora promissor, ainda longe de ampla aceitação no campo econômico da corrente principal (*mainstream*) e, por extensão, nas aplicações normativas. Sejam essas aplicações voltadas para a defesa da concorrência ou para a política industrial, em geral os economistas supõem que dependam de referenciais de análise mais unívocos e mais “precisos”, ainda que essa precisão seja não raro ilusória.

18.4 Implicações Normativas e de Política Econômica

A análise normativa empregada universalmente em teoria econômica é, ainda hoje, a baseada nos conceitos de *bem-estar social* e correspondente *eficiência social ou alocativa* formulados por Pareto no início do século XX. Como se sabe, a noção de eficiência alocativa ou de Pareto propõe que uma alocação social – por exemplo, uma certa distribuição de bens e serviços, promovida pelo sistema de preços (ou algum outro sistema alocativo) – é “superior de Pareto” a outra – e portanto socialmente mais eficiente – se for “fracamente” preferível (sua utilidade não for inferior) a esta outra para todo indivíduo – consumidor, no caso –, e estritamente preferível (utilidade maior) para pelo menos um. O “ótimo de Pareto”, ou uma alocação “eficiente de Pareto”, é tal que não há outra alocação que lhe seja superior de Pareto. Ou, em termos mais simples, sempre que só for possível melhorar a posição de alguém piorando a posição de outrem.

Os chamados “teoremas fundamentais do bem-estar” demonstram que, sob certas condições restritivas (que asseguram a existência do equilíbrio geral competitivo – preferências convexas e insaciáveis, concorrência perfeita, ausência de retornos crescentes na produção etc.), uma alocação de equilíbrio geral competitivo é eficiente de Pareto, e vice-versa; o que estabelece uma relação biunívoca entre *concorrência perfeita* (entre outras restrições) e *eficiência alocativa* ou *ótimo(s) de Pareto*.

Essa vinculação entre concorrência perfeita e eficiência alocativa, que só pode ser estabelecida no nível da *economia como um todo* – por meio do modelo de equilíbrio geral competitivo e dos teoremas fundamentais do bem-estar a ele associados – foi rapidamente absorvida pela Microeconomia. Essa migração pouco crítica – uma vez que há várias restrições na passagem do conceito de eficiência alocativa desde o nível de equilíbrio econômico *geral*, em que foi formulado, para o de equilíbrio *parcial* em mercados individuais – acabou por induzir a análise normativa microeconômica a tratar como *anomalias* ou distorções, sob o título de “*falhas de mercado*”, os desvios dos mercados e/ou indústrias do mundo real em relação tanto à norma competitiva “ideal” (concorrência perfeita *versus* “imperfeições” e poder de mercado), quanto às demais premissas necessárias ao equilíbrio geral – entre outras, a presença de externalidades, economias de escala e de escopo, problemas de coordenação e custos de transação.

Como as formas de mercado concentradas são tratadas pelo enfoque normativo de eficiência alocativa aplicado a mercados específicos? Façamos uma brevíssima recapitulação, seguida de uma crítica desde a perspectiva schumpeteriana.

Posições *monopolísticas* são tratadas, na microeconomia tradicional, como associadas a uma restrição da oferta e a preços acima do nível competitivo, supondo que a empresa esteja maximizando lucros a curto prazo. A noção de poder de mercado é diretamente definida por essa capacidade de fixar preços acima dos custos marginais e unitários, obtendo lucros acima do “normal” (identificado aqui com o nível de preços competitivo, ou seja, que já embute um custo de oportunidade para o empresário). Nesse enfoque estático, o poder de mercado permite ao monopolista apropriar-se de parte do excedente do consumidor (efeito *distributivo*) e acarreta redução de eficiência alocativa para o conjunto da sociedade (efeito *alocativo* conhecido como perda de bem-estar social ou ônus do monopólio, que apresentamos no Capítulo 1).

Nesse quadro, supõe-se que a preservação de posições monopolistas tem por efeito, assim como por desígnio, exercer um tal poder discricionário sobre preços de forma mais ou menos automática.

Os casos de *oligopólio* recebem um tratamento essencialmente semelhante, embora menos unívoco. Uma importante tradição de análise econômica de oligopólio, iniciada por Chamberlin, sustentava que em princípio agentes oligopolistas racionais deveriam operar como um monopólio com maximização conjunta de lucros, que por sua vez seriam repartidos por quotas. A incorporação progressiva e recente da teoria dos jogos na análise da interação estratégica oligopolista permitiu reformular a questão de modo mais rigoroso, buscando explicar a rigidez de preços em níveis acima do competitivo por procedimentos de colusão tácita de preços, mediante formação de “preços focais” (preços convencionais) ou liderança de preços. Tais condutas concertadas, explícitas ou tácitas, implicam preços e lucros supracompetitivos, embora não necessariamente no nível de maximização conjunta, e em geral em nível indeterminado. Seja como for, o resultado é tratado analogamente ao monopólio, isto é, como manifestação e exercício de poder de mercado, com prejuízo líquido alocativo para a sociedade.

As exceções amplamente aceitas são apenas os chamados monopólios naturais – e, por extensão, “oligopólios naturais” –, decorrentes de escalas mínimas eficientes das empresas que, devido a uma presença importante de economias de escala e/ou de escopo, são significativos em comparação com o mercado. Tais situações tendem a ser aceitas como um custo social em perda de bem-estar a ser concedido em troca do benefício social de maior eficiência estática (custos unitários mais baixos), desde que o monopólio seja submetido à regulação pública para assegurar que os preços de fato socializem tal benefício. No Capítulo 23 essas estruturas são analisadas em detalhe.

Em todo esse tratamento analítico convencional perpassa a ênfase no *preço* e o viés *estático*, inclusive para expressar poder de mercado; ambos objeto de crítica original por parte de Schumpeter (1942) e dos autores neo-schumpeterianos, que vêm buscando recentemente tratar das implicações normativas da teoria schumpeteriana da concorrência.

Quanto ao primeiro aspecto (*preço*), vale ressaltar que no enfoque schumpeteriano a concorrência não se reduz à concorrência em preços, assim como o poder de mercado não se expressa apenas neles. A concorrência, especialmente no âmbito das grandes empresas oligopolistas, se dá com instrumentos muito mais poderosos e eficazes, capazes de criar todo tipo de diferenciação (tecnológica, produtiva, comercial, organizacional, de mercado, estratégica) entre os concorrentes na busca pela apropriação de ganhos diferenciais ou quase-rendas monopolistas, e que Schumpeter, como vimos, caracterizou como *inovações lato sensu*. Nesse contexto, o poder de mercado é bem mais diversificado nas causas e formas de manifestação, assim como nas possíveis formas de controle, do que o simples mecanismo de preços.

Quanto ao viés *estático*, Schumpeter argumentou que os comportamentos restritivos associados convencionalmente às grandes empresas monopolistas e oligopolistas são apenas um momento, não raro temporário, do processo concorrencial, por ele visto dinamicamente como de “*destruição criativa*” das estruturas econômicas preexistentes. Nessa perspectiva, eles representam

(...) *incidentes, muitas vezes inevitáveis, de um processo de expansão de longo prazo que eles mais protegem do que impedem. Não há mais paradoxo nisto do que em dizer que os automóveis podem correr mais porque são dotados de freios.*³

Patentes e outros mecanismos legais de proteção temporária e condicional a monopólios, associados à propriedade intelectual, não diferem em substância econômica de procedimentos defensivos destinados à proteção de vantagens competitivas conquistadas mediante esforços inovativos que, muito frequentemente (ainda que nem sempre), acarretam aumento de bem-estar social em perspectiva dinâmica.

Evitar que esses ganhos monopolísticos sejam rapidamente exauridos por imitação fácil e difusão precoce é condição indispensável para assegurar retorno econômico aos investimentos (de P&D, principalmente) voltados às inovações bem-sucedidas, viabilizando um fluxo razoável destas e os efeitos dinâmicos de bem-estar decorrentes. Distinguir entre esses casos e os de mero abuso de posição dominante no mercado é, em princípio, difícil, provavelmente não comportando regras simples e gerais. No entanto, é essencial que uma *política de concorrência*, e por consequência a *análise econômica antitruste*, reconheça a importância do problema, para melhor enfrentá-lo, como veremos no Capítulo 22.

Uma política de concorrência – às vezes chamada de política de “defesa da concorrência” – tem por finalidade precípua tanto proteger como estimular a concorrência nos mercados onde ela esteja ameaçada, seja por intermédio da própria legislação antitruste, seja pela ação dos órgãos por ela incumbidos desse tipo de intervenção nos mercados. Portanto, nessa perspectiva teórica schumpeteriana, ela não pode ser vista como *intrinsecamente* antagônica à existência de posições monopolistas ou oligopolistas, ou seja, de poder de mercado diferenciado.

A criação e ampliação de poder de mercado, em primeiro lugar, tanto ou mais que a atenuação de assimetrias, em segundo lugar, constituem aspectos inseparáveis do processo de concorrência. No primeiro caso, por assim dizer, sua dimensão que chamamos de “ativa”, transformadora das estruturas econômicas mediante inovações e reproduutora das desigualdades de poder econômico entre os agentes; e no segundo – este, o único tratado pelas teorias tradicionais – sua dimensão que denominamos “passiva”, de ajustamento a eliminação de diferenças, por imitação, inovações secundárias e entrada nos mercados. Em qualquer caso, nem o equilíbrio, nem a equalização das assimetrias, nem a supressão do poder econômico são características do processo concorrencial, que dirá suas propriedades principais, como suposto pela teoria tradicional.

18.5 Conclusão

Não poderíamos encerrar este capítulo sem uma necessária referência ao âmbito da concorrência e de um de seus atributos principais, a *competitividade*. Apesar de ser usada quase sempre com viés empírico, esta última noção pode ganhar um contorno mais rigoroso a partir da teoria schumpeteriana da concorrência.

Embora tanto concorrência como competitividade tenham como unidade a *empresa* no nível da ação estratégica, o *mercado* é de fato, como reconhecido na tradição antitruste, um espaço privilegiado tanto na teoria como na intervenção normativa e reguladora. Do ponto de vista teórico, isto se deve não só ao fato de o mercado ser o *locus* da concorrência, mas especialmente porque os instrumentos da disputa competitiva são definidos por características técnico-produtivas específicas de cada indústria e por características do produto associadas à demanda; isto é, no âmbito do mercado; podendo mesmo configurar determinados padrões de concorrência – quando apresentem alguma regularidade.

Além disso, aspectos regulatórios, infraestruturais, sociais e mesmo macroeconômicos – *sistêmicos*, em suma – agem de forma decisiva para calibrar a intensidade do processo competitivo e eventualmente reforçar a *competitividade* das empresas ali atuantes e, por extensão, da indústria correspondente. Preservar e fortalecer a concorrência, nesse quadro, implica a criação/reprodução de um *ambiente competitivo*. Este compreende: (1) estratégias empresariais *inovativas* e a adoção de critérios de eficiência produtiva, no plano das *empresas*; e, (2) no plano do *mercado*, a presença sistemática de *pressões competitivas* internas e potenciais (ameaça de entrada) e de *fatores sistêmicos* favoráveis à concorrência e à competitividade, seja oferecendo externalidades positivas (infraestrutura adequada, mão de obra qualificada etc.), seja assegurando condições macroeconômicas favoráveis ao crescimento e ao financiamento, seja mesmo por meio de legislação adequada e outros instrumentos de defesa da concorrência e da política industrial (como, por exemplo, os instrumentos cambiais e de grau de proteção tarifária e não tarifária, de política comercial).

A implicação mais clara dessas considerações é que a *concorrência* e a *competitividade* não surgem de forma *espontânea* – como supõem frequentemente os economistas com viés mais liberal –, mas dependem de modo crucial da adequação das condições ambientais e, por extensão, de medidas de política econômica. Em outras palavras, *concorrência* e *competitividade* devem ser *construídas*, tanto por iniciativa da *política econômica* – em particular, as políticas industrial e de concorrência (desde que tratadas de forma não antagônica) – com apoio da própria legislação antitruste e regulatória, quanto pelas *estratégias* das próprias empresas, pressionadas pela concorrência local ou (principalmente) pela mundial, num contexto globalizado como o atual.

Nesse sentido, fortalecer a concorrência não implica obrigatoriamente “enfraquecer” (reduzir seu tamanho e/ou sua lucratividade) as empresas, como por vezes se depreende do antigo e muito citado slogan segundo o qual “as leis antitruste foram criadas para proteger a concorrência, e não os concorrentes”. No enfoque schumpeteriano, concorrência fortalecida requer um ambiente intensamente competitivo, o qual, por sua vez, supõe empresas competitadoras fortes, isto é, *empresas competitivas*, por sua capacitação e por sua eficiência técnica, produtiva e organizacional.

Na visão schumpeteriana da concorrência, um mercado atomístico, composto de empresas economicamente insignificantes e desprovidas de qualquer poder de mercado, como paradigma competitivo, é uma lamentável ficção da ortodoxia econômica que, se verdadeira, debilitaria o ambiente competitivo e o processo de concorrência ao ponto de tornar este último *inoperante*, com consequentes prejuízos ao consumidor e ao bem-estar social, quando visto em perspectiva dinâmica.

Notas

1. Sua primeira obra importante, Teoria do Desenvolvimento Econômico, foi publicada em alemão em 1911 e traduzida para o inglês em 1934 (existe tradução em português: S. Paulo, Ed. Abril Cultural, 1982, col. “Os Economistas”). Embora já contivesse todos os elementos básicos para uma teoria da concorrência, esta ainda estava ofuscada por dois outros temas centrais para o autor – a inovação e o desenvolvimento econômico. Que ambos representam os principais efeitos da concorrência na economia capitalista só se torna claro num de seus últimos e mais importantes livros, Capitalismo, Socialismo e Democracia, 1942, caps. 7 e 8 (trad. port. Rio de Janeiro, Zahar Ed., 1984).
2. Nelson, R.; Winter, S. (1982) *An Evolutionary Theory of Economic Change*. Cambridge, Mass: Harvard University Press.
3. Schumpeter (1942), p. 88.

Bibliografia

- Nelson, R.; Winter, S. (1982) *An Evolutionary Theory of Economic Change*. Cambridge, Mass: Harvard University Press.
- Schumpeter, J. (1911) Teoria do Desenvolvimento Econômico. Trad. port. S. Paulo: Ed. Abril Cultural, col. “Os Economistas”, 1982.
- Schumpeter, J. (1942) Capitalismo, Socialismo e Democracia. Trad. port. Rio de Janeiro: Zahar Ed., 1984.

Estratégias de Inovação

Lia Hasencllever e Paulo Tigre

19.1 Introdução

O estudo das estratégias de inovação é um tema relativamente recente na literatura sobre Economia Industrial. A principal corrente teórica que aborda o assunto é a institucionalista-schumpeteriana, que focaliza sua análise na empresa, nos setores industriais e nas relações em rede com outros agentes econômicos. A empresa é concebida como um organismo vivo em permanente mutação que recebe influências de seu ambiente (mercado), mas ao mesmo tempo é capaz de transformá-lo ou criar novos mercados ou indústrias a partir da introdução de inovações tecnológicas. Esta visão contrasta com a visão exposta no Capítulo 8, onde se discutiu a questão da intensidade da inovação em função do padrão de concorrência existente no mercado em que a empresa estava situada. Em outras palavras, a empresa era tomada como um agente passivo, que apenas reagia aos estímulos do mercado, podendo ou não ser incitada por este.

Na concepção da escola institucionalista-schumpeteriana algumas inovações radicais que são acompanhadas de inovações organizacionais e institucionais são caracterizadas como “sistêmicas”. Por exemplo, o processo de difusão da energia elétrica no final do século XIX exigiu a montagem de redes de distribuição, novas formas de compra e venda e uma institucionalidade (legislação, agentes, sistema de financiamento) distinta da adotada quando as principais fontes energéticas eram o carvão e o querosene. As empresas, que na visão desta escola são consideradas organismos vivos, utilizam as inovações para introduzirem variedades na estrutura industrial existente e criarem novas estruturas. Além disso, a concepção e a implementação da estratégia de inovação de uma empresa levam em conta a sua organização interna e as suas relações externas com o sistema de inovação mais amplo em que está inserida.

As estratégias de inovação neste capítulo serão abordadas pelo estudo de três temas principais. Na primeira seção, serão focadas as estratégias de entrada e permanência das empresas em uma nova indústria por meio da introdução de inovações e sua relação com o ambiente tecnológico. Na segunda seção, as formas de organização da inovação nas empresas, em particular a coordenação interna exercida para implementar as estratégias de inovação. Na terceira seção abordaremos as formas de organização externa adotadas pelas empresas para acessar e difundir as inovações surgidas no ambiente tecnológico. Finalmente, à guisa de conclusão, serão abordadas as implicações dinâmicas que a visão estratégica da inovação pode representar para a transformação das estruturas industriais, avançando na interação dinâmica subjacente aos conceitos de estratégia e estrutura.

19.2 Estratégias de Entrada, Inovação e Ambiente Tecnológico

O tema estratégias de entrada das empresas no mercado não despertou atenção da abordagem adotada pela corrente principal da teoria econômica, fortemente influenciada pela visão neoclássica, uma vez que as empresas, segundo esta tradição, não agem na criação de mercados e interagem muito pouco entre si, limitando-se a reagir às mudanças ambientais. Isto significa supor que o comportamento da empresa é resultante de fatores externos a ela – grau de concentração do mercado, barreiras estruturais à entrada, entre outros. As empresas, nesta tradição, são desprovidas de poder discricionário, de capacidade de influenciar a concorrência em seu favor, de rivalizar diretamente com as demais empresas estabelecidas e potenciais entrantes. Em consequência, a eficiência e o desempenho de uma indústria estão principalmente referidos ao padrão de concorrência existente em um mercado.

No início da década de 1980, entretanto, a literatura de Economia Industrial evoluiu no sentido de enfatizar o fato das oportunidades e custos de entrada e saída influenciarem a eficiência e o desempenho de uma indústria. Os autores W. J. Baumol, J.C. Panzar e R.D. Willig mostram que são os custos irrecuperáveis (*sunk costs*), estudados nos Capítulos 3 e 12, e não as economias de escala ou barreiras estáticas, estudadas no Capítulo 7, que constituem a principal barreira à entrada de novas empresas em uma indústria.

Uma questão relacionada com este aspecto e não estudada pelos autores anteriormente mencionados, mas que nos interessa analisar, é em que medida a decisão de entrada de uma empresa em uma indústria pode estar associada com a inovação tecnológica, como o supôs J. Schumpeter. Ou, se mesmo que a decisão de entrada não esteja associada à inovação tecnológica, poder-se-ia relacionar a entrada com as forças de renovação tecnológica dos setores? Ou, ainda, se a introdução de inovações tecnológicas traz irreversibilidades e introduz custos irrecuperáveis que explicam as barreiras à entrada?

A entrada de empresas novas em uma indústria pode ser analisada de forma diferente se usamos a abordagem da concorrência como um processo de seleção (ver no Capítulo 8 o modelo de seleção e no Capítulo 18 a noção de concorrência schumpeteriana). Como visto, nesta perspectiva as empresas entrantes que estiverem incorporando inovações poderão, ou não, obter sucesso dependendo de como as empresas estabelecidas respondem à entrada.

A seguir serão analisadas algumas evidências empíricas sobre o processo de entrada em uma indústria associado à introdução de inovações usando uma abordagem econometrônica e uma abordagem evolucionista.

19.2.1 ESTUDO DAS EVIDÊNCIAS EMPÍRICAS: ABORDAGEM ECONOMÉTRICA

É importante observarmos que o estudo da relação entre a entrada de empresas (geralmente de pequeno porte) em uma indústria e a propensão a inovar é muito difícil de ser inferida por meio de testes econometríticos. O acúmulo de evidências sobre o papel das pequenas empresas na inovação tecnológica é ainda muito pequeno. Além disso, a causalidade inversa da mesma relação – a existência de oportunidades tecnológicas em uma indústria pode ser mais bem aproveitadas pelas pequenas empresas que são flexíveis – também é uma hipótese totalmente plausível.

Subjacente a esta questão, podem ser percebidos vestígios do debate acerca da capacidade da estrutura de mercado determinar a evolução da tecnologia ou de estabelecer as chamadas trajetórias tecnológicas. Entretanto, de maneira geral, é muito difícil construir um modelo operacional porque as condições de entrada na indústria e as oportunidades tecnológicas são dificilmente observáveis ou, mais precisamente, não são perceptíveis por todas as empresas ao mesmo tempo devido às diferentes distribuições do conhecimento entre elas.

Apesar destas dificuldades todas, Paul Geroski, em 1991, fez uma tentativa de mensurar a relação entre as maiores e comercialmente bem-sucedidas inovações e a taxa de entrada de novas empresas em cada indústria, utilizando o banco de dados do *Science Policy Research Unit* da Universidade de Sussex, e entrevistas com 79 indústrias inglesas. Os principais resultados observados foram:

1. Os impactos que as empresas entrantes causaram sobre as condições médias das indústrias (medidos em termos de aumento de faturamento provocado pela entrada de novas empresas) são modestos.
2. O fluxo de entrada de empresas no mercado é bem mais fraco que o fluxo de saída.

O autor observou ainda que, se há grandes diferenças interindustriais, é importante ressaltar que estas diferenças nas atividades de inovação são mais marcadas e estáveis ao longo do tempo, ao passo que as diferenças nas taxas de entrada não o são. Assim podemos resumir os seguintes fatos estilizados quanto à caracterização das relações entre inovação e entrada:

1. A taxa de entrada de novas empresas em uma determinada indústria já existente é dificilmente previsível. Isto sugere que as condições de entrada são provavelmente muito sensíveis a um vasto espectro de fatores ligados às condições particulares de cada mercado.
2. As barreiras à entrada são menos eficientes como uma arma de dissuasão à entrada do que como mecanismos de regulação; elas poderiam, entretanto, explicar as taxas de saídas de empresas relativamente altas.
3. A correlação existente entre entradas de novas empresas e inovações é positiva, mas modesta, e não permite identificar o sentido da causalidade. A coincidência observada entre taxa de entrada elevada e forte propensão a inovar decorre de indústrias para as quais as barreiras à entrada são fracas e as oportunidades tecnológicas particularmente ricas.

Como visto, a abordagem econométrica enfatiza as estruturas de mercado e as características tecnoeconômicas da indústria para explicar a relação da taxa de entrada de novas empresas e a propensão a inovar. Entretanto, os resultados econometríticos são incapazes de encontrar uma correlação significante entre estas duas variáveis e também de precisar o sentido da causalidade para explicar se a entrada de novas empresas em uma indústria é consequência ou causa da introdução de inovações.

19.2.2 ESTUDO DAS EVIDÊNCIAS EMPÍRICAS: ABORDAGEM EVOLUCIONISTA

Quando adotamos a abordagem institucionalista-schumpeteriana, os aspectos pertinentes não são mais as estruturas de mercado ou as características tecnoeconômicas da indústria, mas sim aquelas relacionadas ao ambiente econômico ou regime das empresas, conforme propôs S. Winter (1986). Esta concepção tem origem nos dois modelos de Schumpeter sobre a empresa inovadora: o modelo da pequena empresa inovadora fundada pelo “gênio criador” do empreendedor e o modelo de gestão da inovação organizado pela grande empresa que procura rotinizar o processo de criação tecnológica. Estes dois modelos nos remetem às caracterizações da empresa capitalista identificadas historicamente por J. Schumpeter: a pequena empresa emergente, criada pelo empreendedor, e a grande empresa estabelecida, que introduz inovações rotineiramente a partir de suas atividades de P&D.

Ainda que estes dois modelos possam se apresentar como marcando duas épocas históricas sucessivas – o surgimento das indústrias e as suas maturidades, representando a criação de novas pequenas empresas e seu crescimento e consolidação –, a abordagem evolucionista os considera como dois tipos ideais, duas instituições, dois setores de criação tecnológica que podem coexistir.¹

Um exemplo da coexistência destes dois modelos em uma indústria já madura é a criação da empresa Microsoft, que introduz uma inovação não rotinizada, isto é, contra o padrão ou trajetória tecnológica vigente na indústria de computadores, e revoluciona os padrões desta indústria, sem criar uma nova indústria, mas introduzindo significativas barreiras à entrada.

Assim, em vez de raciocinar a partir do conceito clássico de indústria, visto no Capítulo 2, os trabalhos de S. G. Winter enfatizam as características do ambiente econômico das empresas. Dois regimes econômicos podem ser identificados. O *regime empreendedor* é aquele que se apresenta favorável às empresas que entram (ou que nascem) na indústria e é pouco receptivo às atividades de inovação das empresas existentes na indústria. Ele se opõe ao *regime tradicional* ou *rotineiro*, onde as grandes empresas estabelecidas são as principais responsáveis pela introdução de inovações. Dessa forma, pode-se concluir que o primeiro modelo enfatiza a importância do papel da pequena e média empresa na introdução de inovações e o segundo principalmente o papel das grandes empresas.

Alguns autores, como G. Dosi em seu trabalho de 1988, sugerem que os dois regimes tecnológicos representam, em linhas gerais, a evolução de uma indústria, em particular a sua emergência e a sua fase de maturidade. Segundo o autor, na fase de *emergência da indústria* há uma fase de tentativa e erro que pode afetar os aspectos técnicos, econômicos e comerciais da inovação. Nesta fase os empreendedores trabalham com uma taxa de risco elevada, conforme havia previsto Schumpeter. Muitas vezes, neste período de emergência de novas tecnologias, constata-se o aparecimento de novas empresas. Neste período também é frequente que o progresso técnico seja produzido de forma associada entre empresas e instituições técnicas.

Na fase de *maturidade*, quando as empresas já se consolidaram e cresceram constituindo um oligopólio, a mudança técnica é a principal arma da concorrência. Elas são capazes de mudar as regras do jogo concorrencial, derrubar certas barreiras à entrada e criar outras, provocar o desaparecimento imediato de certos concorrentes e demandar o aparecimento de outros produtores complementares, tais como fornecedores e clientes. A inovação, a difusão e a exploração comercial não são separadas. Quanto mais a estrutura tecnológica é coerente, mais o mecanismo de criação da inovação e das novas técnicas torna-se endógeno aos mecanismos econômicos, sendo internalizados pelas empresas.

Uma indústria nascente está sempre alicerçada sobre um modelo empreendedor, já que nenhuma empresa tem uma posição sólida no mercado. Já em uma indústria madura, podem ocorrer as duas situações polares de acordo com a importância das

QUADRO 19.1 REGIMES TECNOLÓGICOS E TRAJETÓRIAS DE CRESCIMENTO

Regime	Crescimento	Crescimento da Produtividade
Rotineiro	Mais forte	Mais forte
Empreendedor	Mais fraco	Mais fraco

Fonte: *Le Bas, 1995, p. 79.*

QUADRO 19.2 DINÂMICA DA INOVAÇÃO E REGIMES TECNOLÓGICOS

Regime	Volume da inovação	Impacto sistêmico da inovação	Velocidade da difusão	Crescimento da produtividade
Rotineiro	Grande	Fraco	Forte	Forte
Empreendedor	Fraco	Grande	Fraco	Fraco

Fonte: *Le Bas, 1995, p. 79.*

barreiras à entrada. Se as barreiras são altas, um pequeno número de grandes empresas estabelecidas gera o essencial do progresso técnico, como no modelo de P. Sylos Labini (1984). No caso das barreiras à entrada serem baixas, a necessidade de inovação é maior, pois as empresas estabelecidas são sucessivamente ameaçadas por novas gerações de empresas concorrentes.

Assim, é necessário fazer distinção entre os dois regimes (empreendedor e rotineiro) em relação à origem das inovações: são os novos entrantes que inovam no regime empreendedor e são as empresas estabelecidas que o fazem no regime rotineiro. No regime rotineiro existem duas variantes. Na primeira, as empresas estabelecidas inovam protegidas por elevadas barreiras à entrada. Na segunda variante, são criadas novas empresas inovadoras porque as barreiras à entrada estão ausentes.

A conclusão essencial será que o regime empreendedor gera menos concentração – a mesma conclusão apontada por Schumpeter e já vista no Capítulo 8 – mas, da mesma forma, menos crescimento em termos de produto. Se acrescentarmos o fato de que neste regime a difusão leva tempo, podem-se associar a cada um destes regimes tecnológicos, trajetórias de crescimento muito específicas, conforme Quadro 19.1.

Os dois regimes têm as mesmas condições em termos de crescimento potencial dos conhecimentos por ser o crescimento da produtividade potencial nos dois regimes idêntico. Da mesma forma, os dois regimes se igualam no que diz respeito às condições de imitação (por hipótese, não existe nem proteção nem patentes).

No que diz respeito à dinâmica de inovação, há algumas distinções entre os dois regimes, como se pode observar a partir do Quadro 19.2.

O regime rotineiro, apesar de produzir um maior número de inovações, produz principalmente inovações incrementais, ou seja, com impacto sistêmico fraco. O elevado grau de concentração do regime rotineiro garante a velocidade da difusão das inovações entre as grandes empresas estabelecidas e um aumento de produtividade forte.

19.3 Inovações e Estruturas de Organização das Empresas

As empresas, concebidas como organizações, ao escolherem as suas estratégias tecnológicas terão que introduzir estruturas organizacionais compatíveis para suportá-las. Iremos abordar duas análises bastante conhecidas entre os autores schumpeterianos-institucionalistas que se opõem ao tradicional modelo linear e sequencial da inovação, subjacentes à maioria das análises neoclássicas sobre a organização das atividades de P&D no interior das empresas.

O primeiro é o modelo concebido por S. J. Kline e por N. Rosenberg, em 1986, que está mais referido às interações entre as atividades de P&D e às demais funções da empresa, tais como produção e marketing. O modelo está voltado para a geração de inovações, mas também pode ser utilizado para captar as interações que têm lugar no interior de uma cadeia de valor. Isso inclui um conjunto de atividades envolvendo redes de empresas com distintas especializações, universidades e centros de pesquisa.

O segundo é o modelo concebido por M. Aoki, em seus trabalhos de 1986 e 1988, que está direcionado para compreender as estruturas de organização das empresas como um todo e as principais diferenças entre as estruturas em contextos socioculturais

diferentes. Aoki recorre à descrição de dois tipos idealizados de empresas – a empresa americana e a empresa japonesa – concentrando sua análise no fluxo de informações e nos processos de aprendizagem internos e externos.

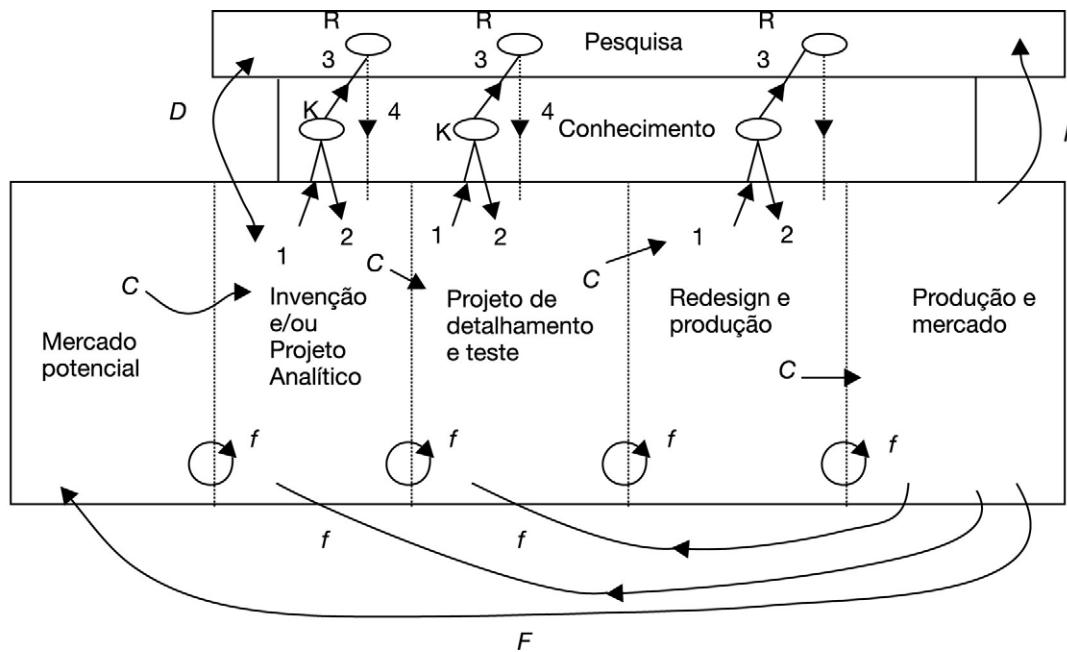
19.3.1 O MODELO DE KLINE-ROSENBERG

Em oposição ao modelo linear de inovação, que supõe que a empresa aplica o conhecimento científico para descobrir novos processos e produtos, a partir de uma visão unidirecional, S. J. Kline e N. Rosenberg, em 1986, criaram um modelo que enfatiza as ligações existentes entre as diferentes atividades de pesquisa e as atividades industriais e comerciais. Como visto no Capítulo 8, as diferentes atividades de pesquisa são divididas em básica, aplicada e experimental e o processo de inovação nasce destas atividades e de suas interações com as demais funções da empresa, tais como as atividades de produção, marketing e vendas.

Este modelo, que pode ser visualizado na Figura 19.1, possui uma cadeia central de interações e quatro outros pontos de interatividade intensa. Na cadeia central, a direção parte de uma invenção ou concepção analítica para os estados de concepção detalhada (protótipo de produto e processo), de produção e de comercialização. Este seria o caminho para geração de conhecimento que já era admitido pelo modelo linear de inovação.

Entretanto, o modelo de Kline-Rosenberg especifica outras relações que estão além desse caminho central e enfatizam a natureza sistêmica do processo de inovação:

1. Fluxos de informação curtos e longos entre as várias fases, permitindo a troca de informações, por exemplo, entre os usuários (departamento de produção) e os inventores (departamento de P&D) para a identificação de novas oportunidades.
2. Uma interatividade entre ciência e inovação que não se restringe ao departamento de P&D, mas se estende a toda cadeia central de inovação.
3. A existência de conhecimento acumulado (competências) no interior das empresas que lhes permite “ler” melhor as oportunidades tecnológicas que se apresentam.



C: cadeia central de inovação

f: realimentação de fluxos de informação curtos

F: realimentação de fluxos de informação longos

K-R: ligações do conhecimento à pesquisa e retornos. Se o problema é resolvido no nó K, a ligação 3 até R não é ativada. O retorno vindo da Pesquisa (ligação 4) é problemático e, portanto, representado por linha tracejada.

D: ligação direta entre a pesquisa e problemas na invenção e projeto.

I: apoio à pesquisa científica por instrumentos, máquinas, ferramentas e procedimentos da tecnologia.

Figura 19.1 Fluxo de Informação e Cooperação no Modelo Kline-Rosenberg. (Tradução a partir de Le Bas, 1995, p. 81.)

É importante ressaltar ainda que o modelo mostrado tanto poderia estar representando uma única empresa, que desenvolve desde a pesquisa básica até a comercialização dos produtos e seus vários departamentos – pesquisa corporativa onde é desenvolvida a pesquisa básica, pesquisa aplicada e desenvolvimento experimental, onde são detalhados os protótipos de cada área de negócios; produção e comercialização –, ou um conjunto de empresas especializadas que se inter-relacionam como clientes e fornecedoras, ou, ainda, um conjunto de empresas especializadas, institutos de pesquisa e universidades que fazem parcerias com o único propósito de produzir inovações.² Essas redes de cooperação foram estudadas no Capítulo 16.

19.3.2 MODELO DE AOKI (EMPRESA A E EMPRESA J)

M. Aoki procurou determinar a importância de alguns determinantes de eficácia das organizações industriais. Ele destacou que a capacidade de aprendizagem, as formas de coordenação internas às organizações e as estruturas informacionais contribuem fortemente para explicar a eficiência das empresas. Para ilustrar suas observações criou dois tipos ideais e polares de empresa: a empresa A, que representa genericamente as empresas americanas, e a empresa J, que representa o modelo adotado pelas empresas japonesas.³

A empresa A possui uma estrutura principalmente vertical de circulação de informações e exprime um desequilíbrio entre a distribuição do aprendizado bastante mais concentrado nos altos escalões ou hierarquias. Esta centralização da informação e esta hierarquia de controle induzem a importantes ineficiências, tais como a baixa capacidade dos gerentes em captar a informação e tratá-la adequadamente; simetricamente nenhum mecanismo incita os assalariados do chão de fábrica a partilhar informações pertinentes. Dessa forma, a empresa tende a apresentar dificuldades de adaptação quando o ambiente está evoluindo rapidamente devido à sua inflexibilidade e dificuldade em interpretar as necessidades de mudança. O grau de integração é muito mais elevado na empresa A, ao passo que a empresa J apela para a coordenação contratual de curto e de longo prazos.

Na empresa J, a organização do trabalho e a coordenação entre os diversos departamentos ou células de produção são estruturadas em torno de um tipo particular de sistema de informações, que é distinto dos modelos clássicos de hierarquia funcional. A empresa J baseia-se em uma comunicação horizontal entre as suas várias unidades funcionais ou departamentos, bem como valoriza a solução autônoma de problemas em cada uma das unidades funcionais individuais, permitindo o espírito de equipe e cooperação entre os seus vários funcionários (operários e gerentes), em contraste com as especializações excessivas e fragmentadas estabelecidas na empresa A. Esta estrutura descentralizada favorece adaptações repentinhas e rápidas das atividades da empresa a um mercado e a um ambiente tecnológico em evolução constante.

Em outras palavras, este tipo de estrutura descentralizado permite a rapidez das comunicações internas necessárias a uma adaptação organizada contra os choques globais que assolam o mercado e uma ação descentralizada contra os choques locais. Ainda que este sistema de coordenação dependa em grande parte das competências (ou das capacidades intelectuais do tratamento das informações) do conjunto dos funcionários, incluindo operários e assalariados. M. Aoki, em seu trabalho de 1988, enfatiza como ponto central de sua análise o sistema de incitação que permite manter as competências no interior da empresa ou desenvolvê-la por meio do mecanismo horizontal de transmissão da informação. Tratando-se de estímulos individuais, o autor afirma que o operário mais produtivo em uma empresa do tipo J é aquele cujas competências são generalistas, possui uma função menos definida.

Outras qualidades deste operário, que lhe permitiriam ampliar a produtividade do trabalho, seriam um espírito cooperativo, grande facilidade de adaptação às novas tarefas, uma capacidade de comunicação com os demais colegas de equipe e de tomar iniciativa para resolver os problemas. O autor chama atenção ainda para o fato de que o rodízio de tarefas entre os vários operários pode ser um estímulo muito interessante para ampliar o espírito de cooperação entre os operários e suas competências.

A empresa J criou um sistema que permite avaliar a aquisição de conhecimentos, as motivações e a eficácia dos assalariados: a hierarquia por escalonamento, por meio da qual os assalariados da empresa concorrem. Destacaremos aqui a título de exemplo a estrutura das atividades de P&D nas empresas da indústria japonesa e, em particular, nas empresas da indústria elétrica e eletrônica que são intensivas nestas atividades. As empresas dispõem de um laboratório central de pesquisa e em cada divisão uma célula de P&D ou departamento de engenharia. Foi este último que jogou um grande papel no processo de criação técnica das grandes empresas japonesas. As características do sistema são: ligação estreita entre a concepção do produto e do processo (engenharia) e a fabricação (produção), transferência de informações garantida pela forte interatividade entre estes departamentos e mobilidade dos trabalhadores.

O laboratório central de pesquisa, todavia, viu seu papel aumentar durante a segunda metade da década de 1970. Sua missão principal, em conjunto com as células de P&D das divisões ou os departamentos de engenharia, era ampliar o potencial de conhecimentos técnicos de base das empresas e aportar novos conhecimentos. A absorção de conhecimentos tecnológicos vindos do exterior deve ser combinada com o conhecimento tecnológico acumulado no interior das empresas.

Existe, assim, uma espécie de isomorfismo (correspondência biunívoca entre elementos de dois grupos que preserva as operações de ambos) entre a forma que as atividades de P&D são organizadas e dirigidas e a forma que as tarefas operacionais são executadas e coordenadas. Isomorfismo entre as comunicações horizontais não hierárquicas na coordenação das operações e a parceria de conhecimentos entre o laboratório central de P&D e o departamento de P&D da divisão.

A hierarquia por escalas, que é um modo de gestão das competências coerente com a estrutura de informação horizontal e descentralizada, irá igualmente reger os recursos humanos investidos em P&D. A saída de um pesquisador ou de um engenheiro é permanentemente desincentivada. A consequência é a fraca mobilidade dos pesquisadores entre as empresas e a difusão mais lenta ou mais dificultada de conhecimentos entre elas. Esta circulação de conhecimentos mais fraca entre as empresas favorece, todavia, uma forte circulação na empresa que pode ser associada a uma forte comunicação que se estabelece entre pesquisadores, entre pesquisadores e engenheiros, entre departamento técnico e de produção, entre outros. Esta excessiva especialização da empresa em uma competência específica poderá dificultar a absorção de conhecimentos de outras disciplinas não relacionadas com as competências de cada empresa, se isto for necessário.

Podemos deduzir que a empresa J não está preparada para gerar projetos pluridisciplinares que ultrapassem as competências estabelecidas na própria empresa. Para tanto ela busca se fundir com outras empresas para provocar um choque de visões apesar desta estratégia trazer dificuldades de incitação de dois escalonamentos de hierarquia distintos e independentes.⁴

19.4 Inovação e Formas de Coordenação Externas

A cooperação entre as empresas, além de favorecer a eficácia produtiva, permite avançar mais rapidamente na criação tecnológica, agregando competências não disponíveis em uma empresa. Este tema alimenta hoje vários eixos de pesquisa em Economia Industrial e pode ser abordado segundo dois aspectos particulares que interessam às estratégias de inovação: a coordenação vertical das atividades e a padronização ou normatização das atividades industriais.

19.4.1 INOVAÇÃO E COORDENAÇÃO VERTICAL DAS ATIVIDADES

Realizar uma inovação tecnológica significa também poder se apropriar dos benefícios econômicos do progresso técnico, tanto nos processos quanto nos produtos, sob a forma de “quase-rendas” ou “sobrelucros”. Portanto, assim que estudamos as diferentes restrições que pesam sobre a empresa inovadora, destacamos aquelas vindas das empresas concorrentes, que podem ter estratégias de imitação, e das empresas potenciais entrantes, que podem dispor de produtos ou de tecnologias que desloquem as anteriores, por meio das estratégias de preços dos fornecedores e dos clientes.

De fato, as quase-rendas da inovação podem ser absorvidas pelas empresas fornecedoras ou por aquelas que comercializam o produto. O saber tecnológico novo baseado em informação é um bem público: toda a venda de um novo produto revela uma informação a um agente econômico que a utiliza para fazer eventualmente concorrência ao inovador. As formas de implementar esta estratégia são várias: licenciamento, cópias, imitação, entre outras.

O principal problema que se apresenta para o inovador é que a apropiabilidade deste conhecimento, ou seja, a capacidade do inovador reter em seu benefício as quase-rendas ou os sobrelucros gerados pela inovação não está totalmente assegurada pelo sistema de propriedade intelectual. De fato, ainda que o inovador esteja protegido, o sistema de propriedade intelectual apresenta limites. Estes limites estão relacionados, entre outros fatores, com o grau de conhecimento tácito embutido em cada tecnologia e com a velocidade de difusão do conhecimento. Pode-se dizer que, quanto maior o conteúdo de conhecimento tácito de uma tecnologia, maior é a proteção que o sistema de propriedade intelectual pode oferecer e quanto maior é a velocidade de difusão do conhecimento em uma área, menos importante é a proteção do sistema de propriedade intelectual.

Outro problema que se coloca é estimar o valor da inovação para que as bases dos contratos de licenciamento e parcerias possam ser estabelecidas. Em geral, a estimativa do valor da inovação foge a algumas regras econômicas estabelecidas para, às vezes, revelar as relações de forças econômicas nas negociações de preços. As empresas com maior poder de mercado obterão melhores preços. É neste momento que se dá a repartição da quase-renda de inovação entre o produtor e os outros agentes econômicos (empresas fornecedoras e compradoras).

D. J. Teece, em 1986, mostrou em um artigo que as empresas que tinham ganhado a competição tecnológica, fracassaram, todavia, em se apropriar dos benefícios econômicos de seus avanços tecnológicos. Ele recomendava que as empresas aumentassem suas atenções para os ativos complementares, ou seja, conhecimentos e competências necessárias à fabricação e à comercialização de bens rentáveis para as empresas inovadoras, que na maioria das vezes são de propriedade das empresas estabelecidas fornecedoras e clientes.

Dois aspectos precisam ser destacados para melhor entendimento da relação entre apropriação das quase-rendas de inovação e a coordenação vertical. O primeiro aspecto está relacionado ao fato de que a empresa inovadora não poderá se beneficiar de forma completa de seus investimentos para gerar a inovação se sua gestão sobre os ativos complementares não for apropriada.

O segundo aspecto diz respeito à forma de propriedade que a empresa inovadora tem que ter em relação aos ativos complementares. Gerenciar ativos complementares não significa obrigatoriamente ter a propriedade destes ativos, como era comum se imaginar no passado onde o grau de integração vertical das empresas era bem mais intenso. Autores como Erick von Hippel, em seu trabalho de 1988, ou como Michael Porter, em seu livro de 1989, trouxeram outra perspectiva, analisando que a chave do entendimento do sucesso da inovação estava nas relações estreitas estabelecidas entre a empresa inovadora, as empresas usuárias da inovação e as empresas que comercializam a inovação. Dessa forma, uma boa adequação entre as características técnicas do produto (e da inovação) e as necessidades dos usuários tem frequentemente mais importância do que as características do processo de produção da inovação em si, como o tamanho dos investimentos e a estrutura do departamento de P&D. Em outras palavras, os autores estão querendo enfatizar que a fonte das inovações pode não estar localizada no interior das empresas, mas no seu exterior. E. von Hippel, neste sentido, mostrou, por meio de vários estudos de caso, a importância dos usuários como fonte de inovação.

Todos esses fatores mostram a importância da estratégia vertical das empresas que buscam combater duas imperfeições básicas de mercado – incapacidade de se apropriar da inovação e comportamentos oportunistas dos fornecedores e dos clientes. Neste sentido, as estratégias verticais são consideradas muito importantes no estímulo à inovação e, em geral, causam poucos efeitos indesejáveis de monopólios espúrios. Por estratégia vertical entendemos a escolha de certa coordenação entre diferentes empresas da cadeia produtiva em oposição ao sistema de trocas no mercado onde a incerteza é mantida.

Uma vez que a opção da coordenação vertical esteja escolhida, gerando troca de informações e conhecimento tácito entre diferentes agentes, ainda é preciso fazer uma segunda escolha entre integração vertical pura e as várias formas de cooperação: quase-integração, parceria, ligação em rede, franquia, *joint-venture*, entre outras. Estas formas de cooperação foram estudadas em detalhe no Capítulo 16.

Uma integração vertical pura tende a unificar as diferentes empresas fornecedoras e clientes em uma única empresa, reduzindo a diversidade da troca de informações entre fornecedores e clientes e a possibilidade de estabelecer uma saudável concorrência entre fornecedores ou clientes. Estas desvantagens têm levado as empresas a cada vez mais utilizar outras formas de cooperação como alternativa à integração vertical pura.

19.4.2 INOVAÇÃO E PADRONIZAÇÃO

Uma tecnologia, para ser amplamente difundida, precisa adotar padrões de produção e de uso. Em outras palavras, precisa dispor de técnicas de produção e de técnicas de uso, de um sistema técnico já constituído que vai permitir sua utilização, mas que ao mesmo tempo vai introduzir certas restrições ao seu uso. Um exemplo é o sistema operacional e os componentes utilizados pelos computadores. Neste caso, a tecnologia necessita de um padrão técnico que vai permitir a utilização de aplicativos, criando, ao mesmo tempo, determinadas restrições ou aprisionamentos. Estas restrições são variáveis conforme as tecnologias e seus ciclos. Dessa forma, no caso das tecnologias de rede, que não podem funcionar sem uma forte interatividade com outras tecnologias, a restrição de compatibilidade é mais forte. Os exemplos mais marcantes são as tecnologias de informação e telecomunicações, mas outras tecnologias mais antigas como a ferrovia e a energia elétrica também precisaram estabelecer padrões para poderem se conectar.⁵

Nos períodos de nascimento das tecnologias há lugar para uma competição entre padrões concorrentes que tentam se impor sobre os demais no sentido de estabelecer um “padrão de fato” para o mercado. Nem sempre apenas um padrão se impõe, mas não há espaço para muitos. Por outro lado, o padrão dominante nem sempre é tecnicamente melhor. Nos dois casos identificam-se ineficiências econômicas.

A literatura econômica sobre normas e padrões é bastante extensa e aborda normas técnicas em oposição a normas de comportamento humano. Reconhece também que certas normas técnicas são impostas pela regulamentação, indicando que nem sempre o mercado é o agente mais eficiente para permitir uma maior harmonia entre interesses dos produtores e consumidores (ver Parte VI).

Para resumir a controvérsia que o tema da padronização suscita, podem-se enfatizar duas grandes abordagens. A primeira supõe que a compatibilidade é uma variável estratégica da empresa; ela é vista como uma iniciativa no sentido de maximizar uma função de utilidade. Esta é a visão adotada, por exemplo, no livro de C. Shapiro e H. Varian, intitulado, *A economia da informação*, onde os autores mostram a ideia, também consagrada por outros autores, de que a padronização é, em geral, preferível para duas empresas em concorrência, a menos que uma das duas empresas domine a outra. Mostram também que as motivações

privadas e sociais para realizar a padronização ou adotar uma norma podem divergir. Isto se daria quando as empresas adotassem a compatibilidade para reduzir a concorrência entre elas e prejudicassem os consumidores (ver Capítulo 23 a respeito).

A segunda abordagem concebe a padronização como um processo mais complexo, no qual os usuários e os avanços tecnocientíficos intervêm por meio de um processo mútuo de retroalimentação. Nesta linha de reflexão estão autores mais heterodoxos, como, por exemplo, os institucionalistas-neo-schumpeterianos. Para estes autores a racionalidade do sistema é mais importante que a racionalidade individual, mostrando como uma tecnologia muitas vezes de menor qualidade e desempenho econômico pode se impor como a norma e tornar-se, ao longo do processo, cada vez mais eficiente devido aos efeitos de aprendizagem.⁶ Outros autores salientam que o fato da nova tecnologia considerar o capital instalado e as externalidades de rede geradas pelas tecnologias atuais atrasa a adoção de novas tecnologias.

19.5 Conclusão

Quais as implicações dinâmicas que a visão estratégica da inovação pode representar para a transformação das estruturas industriais? Na visão schumpeteriana-institucionalista, a empresa e o seu ambiente ou estrutura industrial estão em permanente troca. As estratégias de inovação da empresa implicam mudanças nas estruturas tecnoeconômicas existentes e nas estruturas organizacionais e institucionais, criando e destruindo mercados.

Notas

1. Ver Langlois, R. (1998) a este respeito.
2. Para uma taxonomia dos fluxos de conhecimento entre as empresas e as universidades consultar o artigo de Pavitt (1984).
3. O modelo de Aoki estava refletindo a passagem do fordismo para o pós-fordismo, por volta do final da década de 1980, liderada pelas empresas japonesas. Hoje a maioria das empresas dinâmicas incorpora as mudanças organizacionais que facilitam a coordenação de informações possibilitadas pelo uso das novas tecnologias.
4. A esse respeito, ver Nonaka e Takeuchi (1997).
5. Consultar para um exemplo brasileiro da importância da padronização para a difusão de tecnologia a tese de Souza, 2008.
6. O artigo de Paul David de 1985 sobre o modelo do teclado de uma máquina de escrever, conhecido como modelo QWERTY, é um excelente exemplo.

Bibliografia

- Aoki, M. (1986) "Horizontal versus Vertical Information Structure of the Firm". *American Economic Review*, v. 76.
- Aoki, M. (1988) *Information, Incentives and Bargaining in the Japanese Economy*. Cambridge: Cambridge University Press.
- Baumol, W. J.; Panzar, J. C.; Willig, R. D. (1982) *Contestable Market and the Theory of Industrial Structure*. San Diego: Harcourt Brace Jovanovitch.
- David, P. (1985) Clio and the Economics of QWERTY. *The American Economic Review*, v. 75, n. 2, p. 332-337.
- Dosi, G. (1988) "Sources, Procedures and Economic Effects of Innovation", *Journal of Economic Literature*, v. 26, p. 1120-1171, set. 1988.
- Kline, S. J.; Rosenberg, N. (1986) "An Overview of Innovation". In: Landau, R.; Rosenberg, N., (orgs.), (1986) *The Positive Sum Strategy*. Washington: National Academy Press.
- Langlois, R. N. (1998) Schumpeter and Personal Capitalism. In: Eliasson, G.; Green, C., (orgs.). *Microfoundations of Economic Growth. A Schumpeterian Perspective*. United States of America: The University of Michigan Press.
- Le Bas, C. (1995) *Economie de l'Innovation*. Paris: Ed. Economica.
- Nonaka, I.; Takeuchi, H. (1997) *Criação de conhecimento na empresa*. Rio de Janeiro: Editora Campus.
- Pavitt, K. (1984) "Sectorial Patterns of Technical Change: Towards a Taxonomy and a Theory". *Research Policy*, v. 13, p. 343-373.
- Porter, M. (1989) *Vantagem competitiva. Criando e sustentando um desempenho superior*. Rio de Janeiro: Editora Campus.
- Shapiro, C.; Varian, R. H. (2000) *Economia da informação*. Rio de Janeiro: Editora Campus.
- Souza, T. L. (2008) Metrologia e padronização técnica como ferramentas para a competitividade e inovatividade industrial: uma análise a partir da indústria brasileira de etanol combustível. Tese (Doutorado em Economia). Rio de Janeiro: IE/UFRJ.
- Teece, D. J. (1986) "Profiting from Technological Innovation: implications for integration, collaboration, licensing and public policy". *Research Policy*, n. 5.
- Von Hippel, E. (1988) *The Sources of Innovation*. Nova York: Cambridge University Press.

Estratégias de Propaganda e Marketing

Edmar L.F. de Almeida e Luciano Losekann

20.1 Introdução

Um elemento fundamental na busca de rentabilidade pelas empresas consiste no processo de diferenciação em relação ao comportamento médio da indústria. A propaganda é um instrumento muito utilizado para a diferenciação de produtos. O objetivo básico da empresa ao realizar propaganda é influenciar na formação das preferências dos consumidores, aumentando a demanda de seu produto. Portanto, a propaganda é um elemento relevante da estratégia competitiva das empresas.

A propaganda é uma atividade econômica muito relevante. O montante gasto em propaganda nos Estados Unidos é superior ao produto interno bruto (PIB) de muitos países (Quadro 20.1). Ao gerar receita para os meios de comunicação, a propaganda possibilita a penetração desses por toda a sociedade contribuindo para o bem-estar coletivo. Só assim, as emissoras de televisão podem deixar de cobrar de seus consumidores (TV aberta), os jornais podem ter preços de banca inferiores ao seu custo e os sites da Internet podem ser gratuitos.

Apesar da relevância da propaganda, esta é tradicionalmente negligenciada na análise econômica. Este capítulo tem como objetivo analisar a influência da propaganda sobre o processo de concorrência, avaliando seus efeitos no desempenho das empresas e, consequentemente, na estrutura da indústria. Neste capítulo, serão apresentados os modelos econômicos que analisam os determinantes e efeitos dos gastos em propaganda.

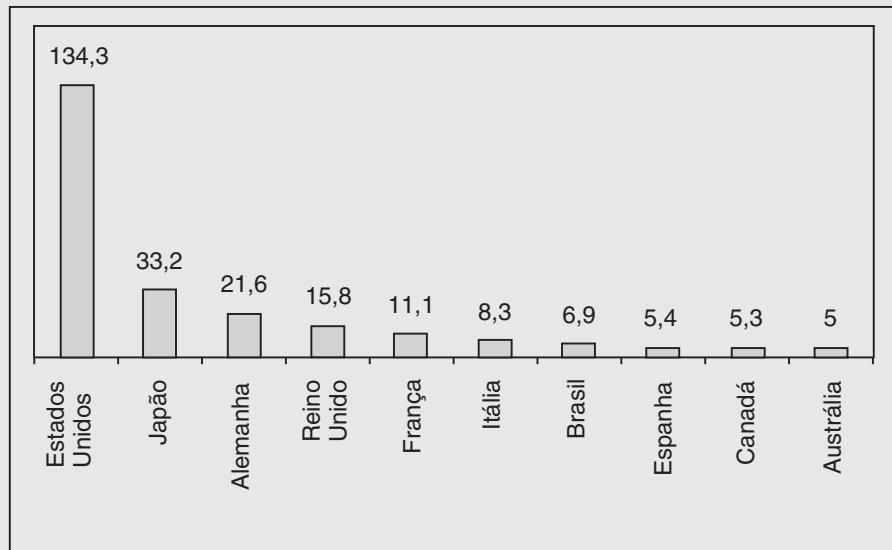
20.2 A Propaganda como Instrumento de Diferenciação

O principal objetivo da propaganda é proporcionar um aumento das vendas por meio da modificação das preferências dos consumidores ou aumento do nível de informação destes últimos sobre o produto comercializado. A propaganda pode ser vista como um instrumento para aumentar a transparência do mercado por meio da divulgação de informações sobre fabricantes, produtos e preços, constituindo um meio para a implementação de estratégias de comercialização diferenciadas e mais bem adaptadas ao ambiente de seleção da indústria.

QUADRO 20.1 O MERCADO GLOBAL DE PROPAGANDA

Os Estados Unidos são o país onde os gastos de propaganda são maiores, totalizando US\$ 134 bilhões em 2000. O mercado brasileiro é considerável, ocupando a sexta posição no ranking mundial de gastos em propaganda. O Brasil é o maior mercado da América Latina e o único país do grupo dos países em desenvolvimento a se posicionar entre os dez primeiros. Apesar desta posição, o gasto brasileiro é vinte vezes inferior ao americano.

Gastos em Propaganda em 2000*
(Top 10 em US\$ bilhões)



*Projeções

Fonte: Meio e Mensagem

Cada empresa busca estabelecer uma estratégia de comercialização mais adaptada ao mercado. Isto se materializa pela escolha de uma forma de propaganda adequada ao tipo de produto, às características do mercado e ao posicionamento estratégico do conjunto de empresas concorrentes. As empresas podem lançar mão de diversas formas de propaganda: (1) anúncios nos meios de comunicação de massa; (2) promoções por meio da concessão de incentivos pecuniários à experimentação de um novo produto; (3) criação de marcas e logotipos para produtos e serviços de fácil memorização por parte dos consumidores; (4) ou mesmo por meio de um esforço de vendas especial com vendedores de porta em porta. O tipo de propaganda adotada dependerá basicamente do tipo do produto ou serviço e do público-alvo.

20.3 Anúncios nos Meios de Comunicação de Massa

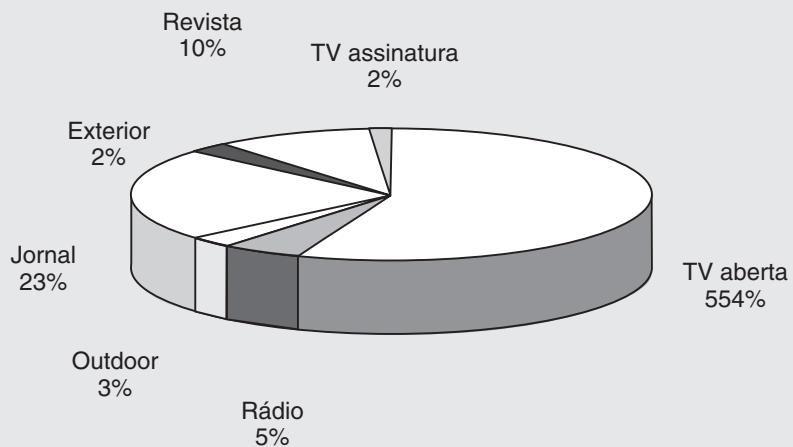
A comunicação de massa realizada por meio do rádio, da televisão e da Internet, entre outros, representa um eficiente instrumento de difusão de informação entre os consumidores. Uma propaganda na televisão é capaz de atingir milhões de consumidores, informando-os e convencendo-os sobre a qualidade e vantagens de produtos de determinadas marcas. O poder de influência dos meios de comunicação de massa nas preferências dos consumidores leva as empresas a alocar um grande volume de recursos com gastos com propaganda. O desenvolvimento da televisão a partir dos anos 1950 foi acompanhado pelo crescimento da propaganda como uma atividade econômica importante das economias industriais. Ver o Quadro 20.2 sobre os tipos de mídia e gastos em propaganda no Brasil.

QUADRO 20.2 OS TIPOS DE MÍDIA

A propaganda é efetuada por meio de diversos tipos de meios de comunicação. Os meios de comunicação que formam a mídia brasileira são divididos em sete grupos: televisão aberta, TV por assinatura, rádio, outdoor, jornal, mídia exterior e revista. Mais recentemente, a Internet vem se consolidando como um novo meio para a realização de propagandas. É consensual que a Internet deve ocupar um papel crescente na estrutura dos gastos em propaganda, mas o volume de gastos nesta mídia no Brasil é ainda baixo, cerca de 0,5% do total.

O total investido em propaganda em todo tipo de mídia no Brasil alcança cerca de R\$ 10 bilhões. A maior parte deste montante é direcionada para a televisão aberta, que representa 56% do total. A TV por assinatura, apesar de receber apenas 2% do total, é a mídia que apresenta a maior taxa de crescimento, 160% no último ano.

Participação dos meios de comunicação nos gastos em propaganda no Brasil (2000)



Fonte: Meio e Mensagem

Várias empresas (agências) atuam no mercado de propaganda oferecendo diversos tipos de serviços especializados. As agências de propaganda são especializadas em tipo de mídia ou/e por tipo de produto anunciado – por exemplo, agências especializadas em anúncios de outdoor ou agências especializadas em marketing político. Assim, apesar do mercado publicitário ser relativamente “atomizado”, algumas empresas detêm elevado poder de mercado em nichos específicos de mercado. Ver no Quadro 20.3 as características do mercado publicitário brasileiro e o posicionamento das agências por participação nas vendas.

A propaganda em meios de comunicação de massa visa construir uma boa impressão do produto anunciado. As campanhas publicitárias podem prover informações sobre a qualidade e preço do produto aos consumidores, ou associando o consumo daquele produto com determinados resultados ou efeitos positivos. Frequentemente estamos sendo alvo de campanhas publicitárias que tentam associar, por exemplo, o consumo de cigarros com atividades que dão a sensação de liberdade e sucesso; o consumo de refrigerante ou cerveja com a sensação de alívio do calor; a compra de automóveis com a imagem de sucesso profissional.

As campanhas publicitárias, quando bem-sucedidas, são capazes de criar uma imagem positiva do consumidor em relação à marca do produto. Isto é conhecido como imagem de marca de um produto ou de uma linha de produtos. Esta imagem de marca é um ativo da empresa proprietária e seu valor é tão maior quanto maior a sua relevância para a diferenciação positiva da empresa no processo de concorrência. Tendo em vista que a marca é um ativo para a empresa, existem leis que protegem os proprietários destas marcas da fraude e imitação. No Brasil, esta proteção é garantida pelo Instituto Nacional da Propriedade Intelectual (INPI). Ver o Quadro 20.4.

QUADRO 20.3 O MERCADO PUBLICITÁRIO BRASILEIRO

Na indústria de propaganda nacional atuam aproximadamente duzentas empresas. Os índices de concentração da indústria são baixos, o CR(4) é de apenas 20% e o índice Herfindahl-Hirshman é de 238.

Outra característica da indústria de propaganda é a forte presença de empresas multinacionais. Apenas duas das dez principais agências que atuam no mercado nacional são genuinamente brasileiras. A empresa líder do setor, McCann-Erikson, já ocupa esta posição a seis anos consecutivos e seu mercado é quase o dobro da segunda colocada, conforme a Tabela 20.1, a seguir.

TABELA 20.1 Ranking de Agências

Posição	Agência	Receita (R\$ milhões)
1º	McCann-Erikson	153
2º	DM9DDBB	78
3º	Grupo Total	62
4º	AlmapBBDO	61
5º	DPZ	60
6º	Young & Rubican	55
7º	Giovanni	55
8º	Propeg	46
9º	Carillo Pastore	43
10º	Talent	42

Obs.: Algumas das principais agências, como Salles D'arcy e W/Brasil, não participaram do estudo da Meio e Mensagem

Fonte: Meio e Mensagem

QUADRO 20.4 INSTITUTO NACIONAL DA PROPRIEDADE INTELECTUAL – INPI

O INPI é uma Autarquia Federal, criada em 1970, vinculada ao Ministério da Indústria, do Comércio e do Turismo. Tem por finalidade principal executar as normas que regulam a propriedade industrial. Suas principais funções são a concessão de marcas e patentes, a responsabilidade pela averbação dos contratos de transferência de tecnologia e o registro de programas de computador, contratos de franquia empresarial, registro de desenho industrial e de indicações geográficas.

20.4 A Propaganda de Informação e de Persuasão

Basicamente, as campanhas publicitárias podem se dividir em dois tipos: a propaganda de informação e a de persuasão. No primeiro caso, a campanha publicitária é centrada na apresentação das características dos produtos, buscando dar maior transparéncia ao mercado quanto às vantagens do produto. Já no segundo caso, a campanha busca transformar a preferência dos consumidores, associando o consumo do produto a determinados resultados ou sensações. O tipo de campanha publicitária escolhida dependerá do tipo de produto.

Os produtos que podem ter suas principais características apreendidas antes mesmo de seu consumo são conhecidos como bens de busca (*search goods*). Os principais atributos destes produtos podem ser facilmente revelados pela uma apresentação visual e oral. Alguns exemplos seriam roupas, materiais esportivos etc. Outros produtos só podem ter suas características inteiramente reveladas por meio do seu consumo ou uso. Este é caso de produtos como bebidas, cigarro, perfumes ou comida, que são conhecidos como bens de experiência (*experience goods*).

No caso dos produtos de busca, a propaganda via informação pode ter efeitos importantes sobre as escolhas dos consumidores. Neste caso, a propaganda oferece informações sobre os produtos por meio de fotos ou descrições. A propaganda pela informação teria pouco efeito em produtos que requerem a experiência do consumidor para ter seus atributos revelados. Dessa forma, no caso dos produtos de “experiência”, o tipo de campanha publicitária empregada é principalmente via persuasão. O objetivo é mudar as preferências dos consumidores, levando-os a experimentar o produto. Dada a dificuldade de informar sobre os produtos de tipo “experiência”, muitas vezes as campanhas publicitárias são centradas em marcas. O caso dos sanduíches McDonald's é um exemplo clássico em que o objetivo da campanha é levar o consumidor a provar os sanduíches da marca.

Não é por acaso que as franquias se difundiram concomitantemente ao desenvolvimento da propaganda em meios de comunicação de massa. As franquias representam um tipo de relação comercial entre empresas cada vez mais comuns nas economias industrializadas. Neste tipo de relação, uma empresa licencia para uma outra empresa uma tecnologia de processo e produto testada e aprovada e o direito de utilizar a mesma marca da empresa licenciadora. As franquias abrangem principalmente produtos do tipo “experiência”, uma vez que neste caso uma marca pode estar associada a uma qualidade de produto ou serviço já conhecida do consumidor. A criação de franquias representa uma forma de valorização da marca por meio do aproveitamento de economias de escala na propaganda. Isto é, na medida em que uma grande quantidade de empresas se estabelece com uma mesma marca por meio da franquia,¹ o volume de recursos disponíveis para investimento em propaganda torna-se expressivo, a um custo relativamente baixo para as empresas individualmente.

20.5 Relevância da Marca para o Processo de Concorrência

Os modelos relacionados com a teoria do consumidor e da empresa na microeconomia tradicional partem do pressuposto de que os agentes detêm informação plena e perfeita sobre as características dos produtos e tecnologias. Neste sentido, a propaganda não contribuiria para o aumento das vendas dos produtos na abordagem microeconômica tradicional, uma vez que os consumidores não se tornariam mais informados sobre determinados produtos e não teriam suas preferências modificadas por ela.

Modelos econômicos mais recentes começaram a analisar o papel da propaganda no mercado a partir do reconhecimento de que os consumidores possuem informações imperfeitas sobre os produtos disponíveis no mercado. Esta imperfeição da informação é vista como uma falha do mercado e está associada a diversos problemas: (1) fontes de informação não são plenamente isentas; (2) o elevado custo de obtenção da informação;² (3) a capacidade limitada dos consumidores de estocar informações disponíveis nos mercados. Modelos que reconhecem falhas do mercado no processo de informação aos consumidores passaram a analisar a propaganda como um instrumento disponível para melhorar o nível de informação dos consumidores e maximização dos lucros das empresas.

A utilização de marcos analíticos baseados na ideia de informação imperfeita abre espaço próprio para a incorporação da propaganda em modelos microeconômicos. A hipótese básica do conceito de informação imperfeita é que o processo de decisão dos agentes é limitado pela falta de informações sobre as variáveis do mercado. Além do problema de informação imperfeita, alguns modelos levam em consideração a racionalidade limitada dos agentes. A partir deste novo fundamento microeconômico, a informação passa a ser um componente importante da estratégia de comercialização das empresas, uma vez que, ao informar consumidores de racionalidade limitada, pode-se influir sobre suas preferências. A utilidade dos produtos passa a não depender somente de suas características intrínsecas, mas pode ser variável de acordo com o nível e tipo de informações que os consumidores recebem que podem estar associados a marcas. Neste sentido, a explicação do comportamento dos consumidores, em contexto de racionalidade limitada, leva em consideração outros sinais do mercado que não o preço.

No contexto de informação imperfeita, a marca passa a ter um papel importante no comportamento dos consumidores, uma vez que associa características de produtos, níveis de qualidade de fabricação e de prestação de serviços a um nome e logotipo. O estabelecimento de uma imagem de marca pode ser visto como um contrato de confiança entre fabricantes e clientes, construído ao longo do tempo, garantindo um nível de qualidade de práticas e produtos.

A marca pode se desenvolver a partir da propaganda, que tenta difundir mais rapidamente as informações que os consumidores só poderiam adquirir após muito tempo de experiência acumulada no consumo de produtos e serviços. Este processo de informação e convencimento resulta na agregação de valor ao produto. Não é por acaso que empresas tentam associar marcas a produtos de difícil diferenciação. Tomates, por exemplo, não podem ser diferenciados segundo o produtor, mas podem ser classificados de acordo com a qualidade. A propaganda pode associar uma marca a tomates de melhor qualidade. O consumidor que, por meio da propaganda obteve a informação que os tomates de determinada marca são de melhor qualidade, pode ser induzido a pagar mais pelo produto desta marca, para não perder seu tempo escolhendo seus tomates no supermercado.

20.6 Nível Ótimo de Propaganda

Vários modelos e estudos econôméticos buscam analisar a determinação do nível ótimo de propaganda e seus efeitos sobre o desempenho das empresas e o processo competitivo. Neste item serão apresentados os principais modelos em formulações simplificadas.³

O modelo de Dorfman e Steiner (1954) é o primeiro a estudar o nível ótimo de propaganda. O modelo considera custo unitário constante da propaganda T por mensagem. A quantidade de mensagens de propaganda, A , em conjunto com os preços, P , determina a quantidade demandada, Q .

$$Q = Q(A, P) \quad \frac{\partial Q}{\partial A} > 0 \quad \frac{\partial Q}{\partial P} < 0$$

Considerando $C(Q)$ os custos excluindo os gastos em propaganda, a função de lucros pode ser definida como:

$$\Pi = PQ(A, P) - C(Q(A, P)) - AT$$

O nível de propaganda que maximiza os lucros é determinado pela primeira derivada da função de lucros em relação à quantidade de propaganda igualada a zero:

$$\frac{\partial \Pi}{\partial A} = \left(P - \frac{\partial C}{\partial Q} \right) \frac{\partial Q}{\partial A} - T = 0$$

Dessa forma, a razão ótima de gastos em relação à receita da empresa será:

$$\frac{AT}{PQ} = \left(\frac{P - \frac{\partial C}{\partial Q}}{P} \right) \frac{A}{Q} \times \frac{\partial Q}{\partial A}$$

Interpretando o resultado, a razão ótima dos gastos em propaganda é igual à margem preço custo multiplicada pela elasticidade da resposta da demanda à realização de propaganda $\left(a = \frac{A}{Q} \times \frac{\partial Q}{\partial A} \right)$.

Considerando que a empresa fixa preços visando maximizar lucros de forma que a margem preço custo se iguala ao inverso da elasticidade-preço da demanda ($\frac{1}{\epsilon}$), pode-se reescrever a equação anterior:

$$\frac{AT}{PQ} = \frac{a}{\epsilon}$$

Este modelo tem dois defeitos principais. Os efeitos da propaganda são duradouros, não se esgotando em um único período. Portanto, a análise deve incorporar os efeitos dinâmicos. Além disso, o modelo não considera a resposta dos competidores.

O modelo de Arrow e Nerlove (1962) realiza uma análise dinâmica da propaganda.⁴ Para considerar o efeito da propaganda ao longo do tempo é necessário incorporar uma taxa de desconto (α). Por outro lado, a propaganda perde efeito ao longo do tempo pelo esquecimento dos consumidores à taxa δ . Assim, a elasticidade de longo prazo da demanda em relação à propaganda pode ser definida como:

$$a^* = \int_0^\infty ae^{-(\alpha+\delta)t} dt = \frac{a}{\alpha + \delta}$$

Substituindo esta equação no resultado de Dorfman e Steiner temos:

$$\frac{AT}{PQ} = \frac{a}{\epsilon(\alpha + \delta)}$$

Cable (1973) incorpora a resposta de outras empresas por meio da propaganda. Pode-se redefinir a função de lucros incorporando na demanda a quantidade de propaganda de outras empresas, \bar{A} .

$$\Pi = PQ(A, \bar{A}, P) - C(Q(A, \bar{A}, P)) - AT$$

$$\frac{\partial \Pi}{\partial A} = \left(P - \frac{\partial C}{\partial Q} \right) \left(\frac{\partial Q}{\partial A} + \frac{\partial Q}{\partial \bar{A}} \cdot \frac{\partial \bar{A}}{\partial A} \right) - T = 0$$

$$\frac{AT}{PQ} = \left(\frac{P - \frac{\partial C}{\partial Q}}{P} \right) \left(\frac{\partial Q}{\partial A} \cdot \frac{A}{Q} + \frac{\partial Q}{\partial \bar{A}} \cdot \frac{\bar{A}}{Q} \cdot \frac{\partial \bar{A}}{\partial A} \cdot \frac{A}{\bar{A}} \right) = \left(\frac{P - \frac{\partial C}{\partial Q}}{P} \right) (a + \bar{a}\eta)$$

onde $\bar{a} = \frac{\partial Q}{\partial \bar{A}} \cdot \frac{\bar{A}}{Q}$ é a elasticidade da demanda em relação à propaganda de competidores e $\eta = \frac{\partial \bar{A}}{\partial A} \cdot \frac{A}{\bar{A}}$ é a elasticidade da resposta de propaganda de competidores com relação à propaganda da empresa.

Este conjunto de modelos aponta que os gastos em propaganda são determinados pela margem de preço-custo, as elasticidades da demanda em relação à propaganda da empresa e a propaganda de competidores, a elasticidade da resposta da propaganda dos competidores em função de propaganda da empresa, a taxa de perda de efeito da propaganda ao longo do tempo e a taxa de desconto da empresa.

Como foi exposto anteriormente, existem dois tipos de propaganda, de informação ou persuasão. George Stigler, em 1961, analisou a propaganda do primeiro tipo. No seu modelo, os consumidores não conhecem o produto da empresa até receber informações. A propaganda é uma forma da empresa informar o consumidor sobre as características do produto. O modelo considera que a cada período novos consumidores passam a fazer parte do mercado, ao passo que outros esquecem as informações do produto ou saem do mercado, perdendo interesse sobre o produto. Assim, para um mercado de tamanho N , a proporção de consumidores informados (λ) é uma função do alcance da propaganda – isto é, a proporção de consumidores que a propaganda consegue informar no período (C), considerando um dado nível de propaganda (A) – e do número de novos consumidores no mercado (b). Este último conjunto é constituído por consumidores que efetivamente estão chegando no mercado e por consumidores que “esqueceram” a informação recebida em períodos anteriores.

No primeiro período existem CN consumidores informados. No segundo período, a quantidade de consumidores informados será formada por três conjuntos:

1. Consumidores do primeiro período que permanecem informados – $C(I - b)N$
2. Novos consumidores informados – CbN
3. Consumidores que já estavam no mercado no primeiro período, mas que não foram informados sobre o produto, sendo informados somente no segundo período – $CN(I - C)(I - b)$

O total de consumidores informados no segundo período é $CN[1 + (I - b)(I - C)]$.

No k -ésimo período, o total de consumidores informados será:

$$CN[1 + (1 - b)(1 - C) + \dots + (1 - b)^{k-1}(1 - C)^{k-1}]$$

Após um número elevado de períodos, esta série pode ser aproximada por:

$$\frac{C}{1 - (1 - C)(1 - b)} N = \lambda N$$

A elasticidade da demanda em relação à realização de propaganda pode ser definida como:

$$\alpha = \frac{A}{Q} \cdot \frac{\partial Q}{\partial \lambda} \cdot \frac{\partial \lambda}{\partial C} \cdot \frac{\partial C}{\partial A}$$

$\frac{\partial Q}{\partial \lambda}$ é uma medida da sensibilidade da demanda em relação à proporção de consumidores informados. $\frac{\partial \lambda}{\partial C}$ é a sensibilidade da proporção de consumidores informados em relação ao alcance da propaganda. $\frac{\partial C}{\partial A}$ representa a sensibilidade do alcance da propaganda em relação a uma unidade extra de propaganda.

O autor explora o termo $\frac{\partial \lambda}{\partial C}$ da equação da elasticidade, interpretado como um fator de atratividade da propaganda. Quanto maior seu valor, o gasto com propaganda tende a aumentar, já que seu retorno se eleva. Stigler aponta que a realização de propaganda apresenta retornos decrescentes em relação à quantidade de consumidores potenciais que recebe informações sobre o produto. Isto ocorre porque apesar dos termos λ e C terem relação positiva, esta também é decrescente ($\frac{\partial^2 \lambda}{\partial C^2} < 0$). Ou seja, quanto mais o mercado está informado, mais difícil é informar um consumidor adicional.

Outro ponto tratado no modelo é relacionamento entre a atratividade da propaganda $\left(\frac{\partial \lambda}{\partial C}\right)$ e a proporção de consumidores no mercado, ou a taxa de rotatividade do mercado, b . Para valores de b pequenos em comparação a C , o aumento da rotatividade estimula a realização de propaganda, porque a empresa deseja informar os novos consumidores sobre o produto. No entanto, quando a rotatividade é muito elevada em relação ao alcance da propaganda, o gasto de propaganda é desestimulado, já que o efeito duradouro da propaganda diminui.⁵

20.7 A Construção de Barreiras à Entrada por meio da Propaganda

A estratégia competitiva de uma empresa pode visar a construção de barreiras à entrada de novos concorrentes. Os gastos com propaganda podem contribuir para a construção destas barreiras, impondo um custo irrecuperável (*sunk cost*) para a atuação na indústria. Ao investir no reconhecimento de suas marcas por meio da propaganda, as empresas estabelecidas podem impor um “pedágio” para a entrada de novos concorrentes. Na medida em que as marcas das estabelecidas se consolidam, os novos concorrentes deverão arcar com gastos de propaganda muito superiores aos das empresas estabelecidas para conseguir penetrar neste mercado. Este investimento é uma condição necessária para contornar o viés dos consumidores em favor das empresas já estabelecidas. Muitas vezes, as empresas entrantes não detêm capacidade financeira para arcar com estes custos irrecuperáveis ou estes tornam a entrada não atrativa em termos econômicos.

Uma vez que uma boa imagem de marca exige, além de grandes investimentos em propaganda, uma experiência acumulada por parte dos consumidores, as marcas estabelecidas contam com uma vantagem comparativa importante diante dos novos entrantes. Esta vantagem comparativa pode ser vista como um capital acumulado na forma de uma imagem de marca em favor das empresas estabelecidas. Dessa forma, as empresas detentoras de uma imagem de marca favorável usufruem de um certo grau de monopólio, que frequentemente estão associadas a taxas de lucro supranormais.

Comanor e Wilson (1967) apontam que o custo de penetração em mercados onde os gastos com propaganda são elevados é o principal determinante de barreiras à entrada. Em função de efeitos cumulativos, as empresas estabelecidas gastam menos em propaganda para proteger sua marca do que uma empresa entrante que tem de construir a marca.

O Gráfico 20.1 descreve o efeito da propaganda na construção de barreiras à entrada. A empresa estabelecida arca com custos de C_a , que incluem o gasto com propaganda para a manutenção da marca. Os custos das empresas entrantes (C_e) são compostos por C_a somado ao custo de penetração no mercado, C_p . O preço que previne a entrada, neste caso, será superior a P_e (o quanto superior é determinado pela relação entre a escala mínima eficiente da entrante, EME, e o tamanho do mercado).

Esta estratégia de criação de barreiras à entrada por meio de gastos em propaganda é comum em mercados de produtos relativamente homogêneos voltados para o grande público. Um exemplo é o mercado de produtos de higiene e limpeza que, apesar de ser um produto de baixa complexidade tecnológica e pouca possibilidade de diferenciação, é dominado por um pequeno grupo de empresas multinacionais (principalmente Unilever e Procter & Gamble). Para limitar a entrada de novos concorrentes, estas empresas alocam uma parcela importante de seus recursos para campanhas publicitárias (Quadro 20.5). Estas empresas investem maciçamente em propaganda não apenas para barrar a entrada de novos concorrentes, mas utilizam esta estratégia competitiva para disputar cada ponto percentual do mercado. Um novo entrante neste mercado teria que investir em propaganda de forma mais agressiva ainda para estabelecer sua imagem diante de outras marcas, configurando uma barreira praticamente intransponível para novos entrantes.

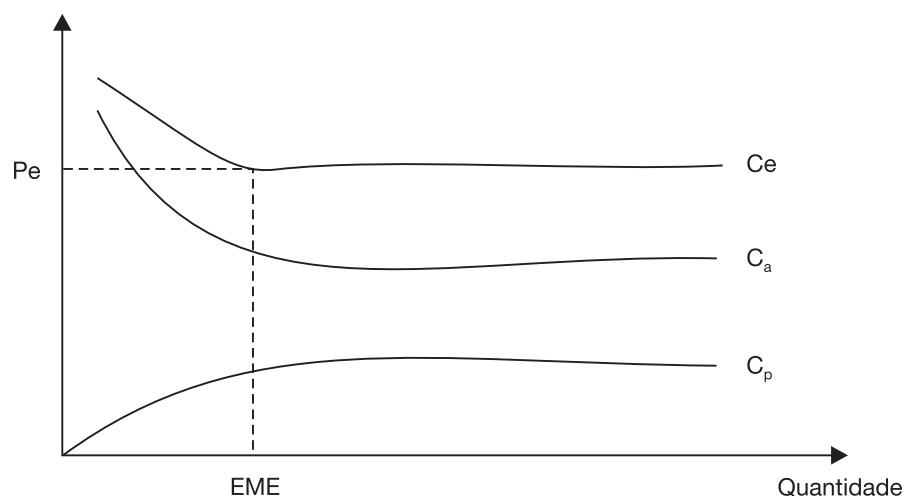


Gráfico 20.1 Barreiras à Entrada e Propaganda.

QUADRO 20.5 ANUNCIANTES

O setor comercial é o responsável pela maior parte dos gastos em propaganda no Brasil, cerca de 30% do total. Supermercados e cadeias de lojas ocupam grande espaço da mídia com seus anúncios. Empresas deste tipo normalmente realizam propaganda informativa, anunciando suas ofertas aos consumidores.

A empresa que mais gasta em propaganda no Brasil é a Gessy Lever. O setor de higiene e limpeza geralmente é um dos prepondérantes nos gastos em propaganda em todo o mundo. O grupo Unilever está entre os três principais anunciantes em cinco dos dez países com maiores gastos em propaganda. O interessante é que as características físicas destes produtos são relativamente homogêneas, e a propaganda se torna o principal meio de diferenciação.

Outro destaque no ranking brasileiro de maiores anunciantes são as montadoras de automóveis, que ocupam a segunda e terceira posições do ranking. O setor automotivo representa parte relevante dos gastos em propaganda no mundo inteiro. Nos Estados Unidos, é o setor que predomina em gastos com propaganda. Neste setor, a propaganda tem uma função diferente, já que os automóveis apresentam diferenças físicas substanciais, servindo tanto para informar o consumidor sobre novidades quanto para interferir na formação de suas preferências.

Ranking dos Principais Anunciantes no Brasil (2000)

Posição	Anunciante	Despesa (R\$ milhões)
1º	Gessy Lever	124
2º	Volkswagen	90
3º	General Motors	86
4º	Embratel	81
5º	Liderança Capitalização	77
6º	Procter & Gamble	76
7º	Itaú	67
8º	Ford	65
9º	Casas Bahia	63
10º	Nestlé	56

Fonte: Meio e Mensagem

20.8 A Propaganda Persuasiva e as Práticas Desleais de Concorrência

Muitos estudos apontam os efeitos negativos na alocação de recursos do mau uso da imagem de marca. Estes estudos argumentam que a propaganda pode direcionar as preferências dos consumidores para produtos inferiores, limitando a capacidade de empresas com produtos de melhor qualidade ou preços inferiores de se impor no mercado.

O caso da concorrência dos remédios genéricos com seus similares de marca ilustra bem as distorções possíveis na alocação de recursos em decorrência da propaganda. Neste caso, laboratórios fabricantes de medicamentos de marca investem em propaganda para passar a mensagem que seu produto é mais confiável que os concorrentes. Estas campanhas muitas vezes levam os consumidores a pagar muito mais caro pelo remédio de marca para não correr um suposto risco de comprar produtos genéricos, que na verdade são idênticos. Neste caso, a propaganda apenas será considerada propaganda persuasiva se afirmar que os produtos concorrentes são inferiores. Entretanto, existem muitas maneiras de sugerir esta mensagem sem colocá-la explicitamente. O consumidor acaba pagando mais caro para minimizar seus supostos riscos.

Estudos recentes tentam explicar as distorções econômicas associadas à propaganda por meio de uma análise dinâmica do processo de concorrência. Isto é, a difusão de produtos de pior qualidade se deve ao tempo necessário para que os consumidores adquiram e acumulem informações sobre as características e desempenhos dos produtos. No curto prazo, algumas marcas estabelecidas com maior volume de vendas podem estar associadas a produtos de qualidade inferior aos concorrentes. Entretanto, uma marca inferior não pode persuadir os consumidores de forma durável, pois estes, no médio e longo prazos, acabarão por “descobrir” as vantagens dos produtos concorrentes.

A propaganda pode ser um poderoso instrumento de práticas desleais de concorrência, uma vez que a propaganda de persuasão pode mudar as preferências dos consumidores para uma marca em detrimento de outra. As práticas desleais ocorrem quando uma campanha publicitária tenta não apenas melhorar a imagem de um produto, mas também denegrir a imagem dos concorrentes com informações negativas, muitas vezes não verídicas. Este tipo de prática normalmente é coibido pela legislação que regulamenta a propaganda. No Brasil a legislação impede inclusive a comparação entre produtos concorrentes, dando pouco espaço para este tipo de prática desleal. Ver o exemplo das restrições à propaganda de cigarros no Quadro 20.6.

QUADRO 20.6 RESTRIÇÕES À PROPAGANDA DE CIGARROS

Sabidamente a propaganda tem relevantes efeitos sobre a demanda de mercadorias. Mas como o governo deve agir em relação à propaganda de produtos cujo consumo é nocivo aos consumidores? Como já ocorre em outros países, o governo brasileiro restringiu a propaganda de cigarros, por meio da Lei Federal n. 10.167 de 2000. Esta lei proibiu os anúncios nos meios de comunicação (inclusive na Internet), anúncios em outdoors, placas e cartazes luminosos. Até então a indústria do tabaco tinha uma grande participação no mercado publicitário. No ano de 1999, a indústria gastou cerca de R\$ 60 milhões com propaganda em televisão, revistas e jornais.

Era particularmente importante a participação dos gastos em propaganda de empresas de cigarro para certos tipos de mídia. As marcas de cigarro realizavam pesados aportes em eventos culturais e esportivos.

Recentemente, foram lançadas peças publicitárias como parte da campanha antitabagista. O objetivo da campanha é revelar os artifícios das campanhas publicitárias de cigarros, que procuram relacionar o consumo de cigarros a atividades esportivas, pessoas bonitas e saudáveis e à sensação de liberdade. Em uma das peças, um “traficante” considera que uma propaganda de drogas pesadas deveria ter esta característica para que os consumidores não percebessem o vício e o malefício à saúde.

20.9 Conclusão

Neste capítulo foi ressaltado o papel da propaganda no processo de concorrência interempresas. Tentamos mostrar que a propaganda tem um papel central na estratégia de comercialização da empresa, que é um dos principais instrumentos de diferenciação no processo de concorrência. É por meio desta diferenciação que a empresa pode obter uma taxa de lucro acima da média da indústria, garantindo um aumento de sua participação no mercado.

A importância da propaganda no processo de concorrência das economias industriais resultou numa expressiva alocação de recursos para esta atividade. No Brasil, os gastos com propaganda se situam em cerca de 1% do PIB, atraindo um grande número de empresas para este mercado. Os serviços de propaganda variam segundo o tipo de mercado e de produto anunciado e o tipo de meio de comunicação utilizado para fazer o anúncio. Entretanto, o objetivo principal de toda campanha publicitária é o mesmo: atuar sobre as preferências dos consumidores em favor de um produto ou uma marca de produtos.

Basicamente, as campanhas publicitárias se dividem em dois tipos: as campanhas de informação que tentam mudar as preferências dos consumidores revelando informações sobre produtos que não são do conhecimento dos consumidores; e as campanhas de persuasão cujo objetivo é associar o produto a certos padrões de qualidade ou tipos de resultados que levem os consumidores a experimentá-lo. Neste capítulo mostramos que a marca é um dos principais instrumentos da propaganda. Uma das principais formas de atuação da propaganda é a promoção da fidelidade dos consumidores às marcas de produtos e serviços. Neste sentido, a imagem de marca é um ativo importante de uma empresa, uma vez que é por meio dela que a empresa implementará sua estratégia de diferenciação via comercialização, podendo inclusive impor barreiras à entrada de novos concorrentes quando atingir uma posição dominante nesta indústria.

Notas

1. Alguns exemplos de franquias importantes no mercado brasileiro são McDonald's, Bob's, Mister Pizza, Casa do Pão de Queijo, O Boticário, entre outros.
2. Por exemplo, o grande número de empresas oferecendo o produto inviabiliza a comparação dos preços e condições oferecidas por cada uma destas empresas em um determinado mercado.
3. As versões apresentadas seguem, em larga medida, a exposição de Hay e Morris (1979).
4. Os autores utilizam o conceito de “goodwill stock” para representar o efeito cumulativo da propaganda nas preferências dos consumidores.
5. Estes resultados podem ser verificados pela análise das primeiras e segundas derivadas de λ :

$$\frac{\partial \lambda}{\partial C} = \frac{b}{(C + b - Cb)^2} > 0$$

$$\frac{\partial^2 \lambda}{\partial C^2} = \frac{2b(1-b)}{(C + b - Cb)^3} < 0, \text{ pois } C > 0, b > 1$$

$$\frac{\partial^2 \lambda}{\partial C \partial b} = \frac{C - b + bC}{(C + b - Cb)^3} > 0 \text{ se } C - b + bC > 0, \text{ i.e. } \frac{1}{b} > \frac{1}{C} - 1.$$

Bibliografia

- Angelier, J. P. (1991) *Economie industrielle: éléments de méthode*. Presses Universitaires de Grenoble, 131 p.
- Cable, J. (1973) in Cowling, K. (1973) *Market Structure and Corporate Behaviour*. Londres, p. 105-124.
- Comanor, W.; Wilson, T. (1967) "Advertising, Market Structure and Performance". *Review of Economics and Statistics*, v. 49, novembro, p. 423-440.
- Dorfman, R.; Steiner, P. (1954) "Optimal Advertising and Optimal Quality". *A.E.R.*, v. 44, p. 826-836.
- Glais, M. (1992) *Économie Industrielle: Les Stratégies Concurrentielles des Firmes*. Éditions Litec, Paris.
- Hay, D.; Morris, D. (1979) *Industrial Economics*. Oxford: Oxford University Press.
- Meio e Mensagem*. Vários números.
- Nelson, R.; Winter, S. (1982) *An evolutionary theory of technical change*. Harvard University Press, Cambridge, Ma, p. 437.
- Nerlove, M.; Arrow, K. (1962) "Optimal Advertising Policy under Dynamic Conditions". *Economica*, N.S. 29, p. 129-142.
- Porter, M. (1996) *Estratégia Competitiva: Técnicas para Análise de Indústrias e Concorrência*. Rio de Janeiro: Editora Campus.
- Porter, M. (1999) *Vantagem Competitiva: Criando e Sustentando um Desempenho Superior*. Rio de Janeiro: Editora Campus.
- Shapiro, C. (1985) "Premiums for High Quality Products". *Quarterly Journal of Economics*, novembro.
- Stigler, G. (1961) "The Economics of Information". *J.P.E.* v. 69, p. 213-225.
- Williamson, O. (1961) "Selling Expenses as a Barrier to Entry". *Quarterly Journal of Economics*, v. 77, fevereiro, p. 112-128.



Estratégias de Financiamento

Helder Queiroz Pinto Jr.

21.1 Introdução

É muito frequente observarmos na literatura econômica uma separação formal estabelecida entre o conjunto de “fenômenos reais”, relativos à esfera produtiva, e os “fenômenos financeiros”. É bem verdade que alguns autores importantes, como Hilferding e Schumpeter, discutiram a importância do sistema financeiro, em particular a importância do crédito para o desenvolvimento industrial. Entretanto, a maior parte dos trabalhos de Economia Industrial não presta suficiente atenção à dimensão financeira.

Isto é explicado, em parte, pelo fato de o estudo dos problemas financeiros ter sido considerado, durante muitos anos, um campo de trabalho à parte, uma espécie de “primo pobre” da economia,¹ tratando basicamente dos aspectos técnicos e contábeis dos mecanismos de financiamento das empresas.

Entretanto, desde os anos 1950, observa-se um aumento considerável dos trabalhos teóricos versando sobre a interdependência entre o sistema financeiro e o sistema industrial. As preocupações com este tema se articulam em torno da seguinte questão principal: de que forma as condições de financiamento e a estrutura de financiamento das empresas influenciam as decisões de investimento das empresas?

Para aportar elementos de resposta, três abordagens privilegiam a análise das relações de causalidade entre sistema financeiro e sistema industrial. A primeira investiga o desenvolvimento do sistema financeiro como resultado das transformações do sistema industrial. Nesse caso, as decisões de investimento produtivo comandariam as mudanças do sistema financeiro, exigindo deste último a necessidade de adequação dos instrumentos de crédito à evolução do perfil do investimento produtivo. Cabe notar que, nesse tipo de abordagem, as relações de interdependência são frágeis e o papel do sistema financeiro é reativo: as mudanças constituem uma resposta à economia real.

A segunda vertente examina os problemas de direitos de propriedade dos grupos e conglomerados, ressaltando a importância do capital financeiro para o controle acionário de grandes corporações. Essa vertente remonta aos estudos de Berle e Means, publicados em 1932,² sobre a concentração industrial e financeira da economia norte-americana. Esse tipo de abordagem vem recentemente ganhando espaço nos estudos de teoria das organizações.

A terceira busca destacar as relações de interdependência entre os sistemas industrial e financeiro, caracterizados por um estado de interação permanente. Isto é, os limites ao desenvolvimento do sistema industrial não são independentes da estrutura do sistema financeiro. Sob esta ótica, as escolhas tecnológicas das empresas podem estar condicionadas por problemas relacionados com a oferta e os instrumentos de crédito.

Apesar dos diferentes focos de análise, todas essas abordagens destacam o problema de mobilização de recursos financeiros para o crescimento da empresa.

É importante notar que uma empresa sempre recorre a diversas fontes de financiamento para implementar seus programas de investimento. Elas podem ser classificadas da seguinte maneira:

1. Reinversão dos lucros retidos (autofinanciamento).
2. Emissão de ações (captação de recursos dos acionistas).
3. Emissão de títulos financeiros (mecanismos de financiamento direto).
4. Empréstimos bancários (mecanismos de financiamento indireto).

A estrutura de financiamento da empresa é definida como sendo a escolha de uma combinação das fontes de financiamento apresentadas anteriormente. A mobilização de recursos por meio de cada uma dessas fontes representa um custo distinto para a empresa. Essas fontes constituem o conjunto de instrumentos que são conhecidos na literatura de Economia Industrial como finanças corporativas (*corporate finance*).

Um dos problemas mais importantes de gestão econômico-financeira e da escolha da estrutura de financiamento diz respeito ao custo de cada uma dessas fontes de financiamento. Se a participação de cada uma delas nos programas anuais de investimento da empresa é estável, o custo médio do capital pode ser utilizado como referência para a determinação da taxa de desconto para a seleção de investimentos. Essa observação será analisada nas próximas seções, à luz das principais abordagens teóricas sobre as inter-relações entre decisões de investimento e decisões de financiamento das empresas.

Na Seção 21.2 será analisado o efeito de alavancagem e o teorema de Modigliani Miller e na Seção 21.3 a interdependência das decisões de financiamento e investimento. Finalmente na Seção 21.4 são apresentadas as principais conclusões.

21.2 O Efeito de Alavancagem e o Teorema de Modigliani e Miller

As decisões de financiamento e de investimento estão permanentemente ligadas às seguintes questões:

1. Que taxa de desconto utilizar?
2. Qual o custo de capital máximo que é possível suportar?
3. Como selecionar investimentos em situação de restrição de crédito?

Vamos ver de que forma a Economia Industrial tem tentado oferecer elementos de resposta a essas questões.

O estudo de Modigliani e Miller, de 1958, é considerado uma referência fundamental no debate teórico sobre as decisões de investimento e de financiamento das empresas. Na visão destes autores, o custo de capital das empresas independe das estruturas de financiamento das empresas, mensurada pelo indicador γ que representa a razão endividamento-capitais próprios (*debt-equities ratio*).

O teorema comprehende duas etapas de análise: a primeira pressupõe que as empresas desejam não apenas maximizar seus lucros, mas também maximizar seu valor de mercado. A segunda destaca que o valor de mercado (V) da empresa independe de γ .

O teorema de Modigliani e Miller (MM) contesta a visão tradicional sobre o efeito de alavancagem do endividamento sobre a rentabilidade dos capitais próprios. Vejamos, inicialmente, de que forma a visão tradicional aborda o problema.

21.2.1 VISÃO TRADICIONAL DE EFEITO SOBRE EFEITO DE ALAVANCAGEM

Para isso, é importante definir algumas variáveis fundamentais para a análise da estrutura de financiamento das empresas.

Vamos considerar, inicialmente, o valor de mercado V que é definido como sendo igual a:

$$V = S + D,$$

onde: S é o valor de mercado dos capitais próprios e D é o valor de mercado da dívida total da empresa. O indicador D/S reflete a estrutura de endividamento global da empresa.

A estrutura de financiamento pode ser definida como a participação relativa de cada modalidade de financiamento utilizada pela empresa. Considerando que a empresa se financia apenas com as modalidades S e D , o peso de cada modalidade na estrutura de financiamento é dado respectivamente por:

$$S/V \text{ e } D/V.$$

O passo seguinte é a definição do custo de capital das empresas. Esse é um tema controverso entre os estudiosos de teoria financeira. Se, por um lado, o custo dos capitais emprestados é normalmente determinado pela taxa de juros dos empréstimos contratados, a mensuração dos custos dos capitais próprios depende sempre de um conjunto de premissas e de critérios adotados para a sua definição, devendo refletir o custo de oportunidade da empresa. De uma maneira esquemática, podemos definir o custo médio ponderado de capital da seguinte maneira:

$$CMK = r(S/V) + i(D/V),$$

onde: r = custo dos capitais próprios e i = custo dos capitais de empréstimo.

Admitindo a ausência de incidência fiscal, uma formalização simples do efeito de alavancagem para um montante total de investimento a ser realizado é apresentada a seguir:

$$\alpha = RE / KI \quad (1)$$

$$\beta = (RE - CF) / KP \quad (2)$$

$$KI = KP + KE \quad (3)$$

$$CF = KE \times i \quad (4)$$

onde:

α = indicador de rentabilidade do capital investido

β = indicador de rentabilidade dos capitais próprios

RE = resultado operacional

KI = capital investido

KP = capital próprio

KE = capital de empréstimo

i = taxa de juros

D = dívida total

CF = custos financeiros

Consideremos duas situações (I e II) com relação ao modo de financiamento. A primeira situação é chamada de autofinanciamento total, isto é, a empresa financia seu programa de investimento apenas com capitais próprios. Na segunda, denominada efeito de alavancagem, a empresa recorre a capitais de empréstimos para financiar uma parte do seu programa de investimentos.

Assim, na situação I, α é exatamente igual a β , pois o montante dos capitais próprios corresponde ao total do capital investido. Em outros termos:

$$\alpha = \beta = RE / (KP + KE), \text{ como } KE = 0, \text{ logo } \alpha = \beta = RE / KI$$

Na situação II, é fácil encontrar, a partir de (1) e (2), que:

$$\beta = (\alpha KI - CF) / KP \quad (5)$$

Se substituirmos KI e CF em (5), observamos:

$$\beta = (\alpha (KP + KE) - KEi) / KP \quad (6)$$

Como o nosso objetivo é analisar as relações entre α e β , podemos, com a ajuda do indicador endividamento-capitais próprios γ ,³ encontrar:

$$\gamma = KE / KP \quad (7)$$

$$\beta = \alpha(1 + \gamma) - (\gamma i) \quad (8)$$

Duas hipóteses complementares são necessárias para demonstrar que a rentabilidade dos capitais próprios é mais elevada quando a empresa se endivida:

1. α é constante para as situações I e II;
2. $i < \alpha$.

A primeira hipótese sustenta que as características técnicas e econômicas do projeto não se alteram, ou seja, a rentabilidade econômica (não financeira!) esperada do projeto é a mesma, independente do seu modo de financiamento. A segunda indica que a taxa de juros referente ao empréstimo contratado é menor que a rentabilidade econômica do projeto: a empresa, por prudência, não contratará empréstimos com taxas de juros superiores à rentabilidade econômica esperada.

Dessa forma, o ponto onde $i = \alpha$ é chamado “ponto de apoio da alavancagem”. Assim, se $\gamma = 0$, a empresa não tem dívidas e reencontramos o resultado da situação I: $\beta = \alpha$. Porém se $\gamma > 0$, β assumirá valores maiores do que na situação I, admitindo

as duas hipóteses mencionadas anteriormente. Esse resultado indica que a rentabilidade dos capitais próprios é “alavancada” pelos recursos de empréstimos.

À luz desta abordagem, a gestão financeira da empresa deve buscar o ponto ótimo para γ , que coincide com o custo médio de capital mínimo.

21.2.2 TEOREMA DE MODIGLIANI E MILLER (MM)

As proposições fundamentais de MM contestam a visão anterior e estabelecem que o valor de mercado (V) e o custo de capital (CMK) são independentes da estrutura de financiamento da empresa. Isto significa que a visão de MM postula uma neutralidade entre decisões de investimento e de financiamento. Em outras palavras, o CMK independe da forma de financiamento escolhida.⁴ As principais hipóteses de MM são:

1. Existência de um mercado perfeito de capitais
2. Nível de risco conhecido
3. Ausência de restrição de solvência da empresa
4. Informação perfeita
5. Ausência de incidência fiscal

Para MM o valor de mercado da empresa depende apenas da capacidade de realização de lucros no longo prazo. Além disso, a taxa de desconto utilizada para a seleção de investimentos independe da modalidade de financiamento escolhida.

Consideremos a situação inicial, no instante $t=1$:

$$CMK_1 = r_1(S_1/V_1) + i_1(D_1/V_1).$$

Na visão de MM, se em $t=2$ a empresa decide investir contratando novos empréstimos, o aumento da dívida $D_2 > D_1$ não afeta CMK, pois a variação do nível de endividamento, mesmo alterando a razão D/V , compensada por um aumento de remuneração associada aos capitais próprios. Logo, no instante $t=2$, $r_2 > r_1$.

A Tabela 21.1 resume de forma simplificada, com um exemplo numérico, o principal resultado do teorema de MM. Tanto em $t=1$ como em $t=2$, CMK permanece igual a 0,08, apesar do aumento do nível de endividamento. Isso ocorre devido ao aumento da remuneração dos capitais próprios.

A Figura 21.1 ilustra as visões tradicional e a de MM sobre as relações entre CMK e γ .

TABELA 21.1 Custo Médio de Capital e a Abordagem de Modigliani-Miller

	$t=1$	$t=2$
V	100	120
S	60	60
D	40	60
r	0,10	0,11
i	0,05	0,05
$CMK = r(S/V) + i(D/V)$	0,08	0,08

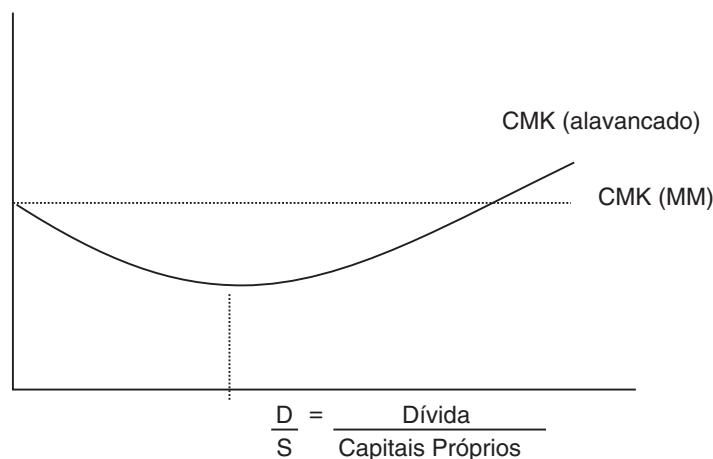


Figura 21.1 Custo Médio de Capital e o Indicador Dívida/Capitais Próprios.

21.3 A Interdependência das Decisões de Financiamento e de Investimento

A abordagem de MM provocou um intenso debate na literatura econômica sobre as relações entre decisões de financiamento e de investimento. Essas críticas estão centradas, em particular, no caráter pouco realista das hipóteses de funcionamento dos mercados de capitais.⁵ A lógica subjacente das proposições de MM está baseada na hipótese de que os meios de financiamento são perfeitamente substitutos entre si. Entretanto, a possibilidade da utilização de diferentes formas de financiamento está sujeita a aspectos referentes às condições de acesso ao crédito e às diversas técnicas de financiamento.

21.3.1 RACIONAMENTO DE CRÉDITO E DECISÕES DE INVESTIMENTO

Quanto às condições de acesso ao crédito, a abordagem de MM reserva um papel passivo ao sistema financeiro. Num mercado perfeito de capitais, a taxa de juros cumpre o papel de equilibrar oferta e demanda de crédito. Como observa Stiglitz (1969), não haveria lugar, nesse caso, para problemas de racionamento de crédito. Entretanto, esse é um problema concreto. As instituições financeiras levam em conta simultaneamente a rentabilidade esperada dos projetos a serem financiados e a capacidade da empresa de reembolsar as dívidas contratadas.

Desse ponto de vista, o risco financeiro de novos projetos exige a avaliação da capacidade de garantir o pagamento das dívidas.⁶ Esse tipo de informação afeta diretamente as condições de financiamento. Isto é explicado pelo fato de que os empréstimos considerados mais arriscados têm contratos de financiamento com condições mais restritivas: prazos mais curtos e taxas de juros mais elevadas. Este argumento leva à rejeição do pressuposto de neutralidade entre decisões de financiamento e de investimento. Quando a empresa está mais endividada, seus projetos se tornam mais arriscados e o financiamento mais caro. Ora, isso se reflete no custo médio de capital da empresa.

Vale lembrar que as diferenças de custo entre as diversas fontes de financiamento estão ligadas ao problema dos riscos financeiros assumidos pelos credores. De forma simplificada, podemos afirmar que o “departamento projetos” toma recursos emprestados do “departamento financeiro” a um custo médio de capital. Porém, em grandes empresas, o risco econômico e financeiro das múltiplas atividades pode ser distinto.

Dessa forma, importa avaliar a questão da escolha da taxa de desconto para o processo de seleção de investimentos. Como se sabe, uma taxa de desconto baixa implica a seleção de projetos mais capital intensivos. Em contrapartida, uma taxa de desconto alta conduz à seleção de projetos menos capital intensivos.

Em situações de racionamento de crédito, a taxa de desconto cumpre mais um papel na seleção de investimentos: a arbitragem de uma taxa de desconto irá então refletir escolhas estratégicas. Nestas condições, o “departamento projetos” não pode tomar emprestado ao “departamento financeiro” a uma taxa equivalente ao custo médio de capital, e os métodos de análise de risco devem ser mais sofisticados.

A técnica de avaliação de projetos, conhecida como método do valor presente, favorece a exploração de economias de escala e o desenvolvimento de grandes projetos num contexto onde não existam restrições de financiamento. Assim, dados dois projetos A e B, se o investimento inicial do projeto A (I_a) e do projeto B (I_b) são, respectivamente, iguais a 1.000 e 200, a seleção é feita, a partir do cálculo atualizado do fluxo de caixa (somatório das receitas menos despesas) descontado ao longo da vida útil do projeto, por meio da comparação do projeto de maior valor atual. Se o valor atual do projeto A (V_a) é maior do que o valor atual do projeto B (V_b), o projeto A é selecionado apesar do maior investimento inicial.

Mesmo sem entrar na análise detalhada da determinação da taxa de desconto a ser utilizada no cálculo econômico, pode-se constatar que o custo do investimento inicial e seu modo de financiamento não constituem uma restrição fundamental nesse tipo de método. Porém, num contexto de racionamento de crédito, esse tipo de método não é suficiente. Nesse caso, apesar de ser mais interessante, do ponto de vista estritamente econômico, o desenvolvimento do projeto A pode não se concretizar, pois o seu equacionamento financeiro será muito mais complexo, em função da diferença do custo do investimento inicial. Esse aspecto assume uma relevância ainda maior nas indústrias altamente intensivas em capital e com longo prazo de maturação dos investimentos. Se a empresa possui uma relação endividamento-capitais próprios elevada, a contratação de empréstimos adicionais pode comprometer a sua capacidade de solvência.

Isto posto, cabe notar que o problema de racionamento de crédito não está tão somente relacionado com a liquidez das instituições financeiras. O papel ativo do sistema financeiro confere um caráter seletivo às decisões de oferta ou de restrição de crédito. Nesse caso, o que importa são as condições de solvência do tomador de empréstimo que deve ser capaz de oferecer garantias de reembolso dos recursos emprestados.

21.3.2 COMPORTAMENTO DE AUTOFINANCIAMENTO, ESPECIFICIDADES SETORIAIS E O PROJECT FINANCE

Outra crítica comum ao teorema de MM se deve ao fato do não reconhecimento das especificidades dos programas de investimentos em diferentes setores da atividade econômica. Muitas vezes, o recurso ao autofinanciamento é absolutamente necessário. Tal

é o caso das indústrias extractivas minerais, como, por exemplo, a indústria de petróleo. A atividade de exploração e prospecção é usualmente financiada com capitais próprios devido aos riscos geológicos e ao caráter aleatório associados à atividade de descoberta de novas jazidas. Porém, uma vez comprovada a existência de reservas, a atividade de produção e as demais atividades da cadeia produtiva desta indústria podem combinar diferentes estratégias de financiamento.

Um *project finance* é uma técnica de financiamento baseada na atratividade de um projeto específico e não na análise do endividamento global da empresa, como os denominados *corporate finance*. O *project finance* é uma modalidade de financiamento garantido pelas receitas, ativos e colaterais (por exemplo, direitos de concessão) de um projeto específico.

Isso significa que quando os credores decidem financiar um *project finance*, eles se baseiam na capacidade do projeto gerar recursos que garantam o pagamento e remuneração de seu capital, independentemente dos outros fluxos que compõem o balanço das empresas empreendedoras do projeto. Para tanto, cria-se a figura de uma Sociedade de Propósito Específico (SPE),⁷ capaz de representar o projeto como entidade.

A constituição da SPE requer uma estrutura que delimita os compromissos dos agentes envolvidos. A SPE é um verdadeiro nó de contratos com: instituições financeiras, investidores institucionais, bancos, construtores, clientes, fornecedores, seguradoras, poder concedente, entre outros. Isto explica os elevados custos de transação para o desenvolvimento de um *project finance*. Esses custos refletem as despesas legais envolvidas na elaboração do projeto, pesquisa e gerenciamento de informações, aspectos fiscais, preparação de documentação. Quanto maior a incerteza quanto às variáveis que definem a rentabilidade do projeto, maiores serão o tempo de busca de instrumentos contratuais visando a assegurar a maior cobertura possível de contingências. Em muitos casos, a falta de históricos comparativos de projetos semelhantes e a impossibilidade de uso de determinadas ferramentas de garantia podem inviabilizar a opção por um *project finance*.

A estrutura contratual e financeira montada tem a finalidade principal de fundamentar os arranjos necessários para viabilizar a participação dos potenciais agentes econômicos e a perfeita identificação de riscos e respectiva alocação otimizada. Neste aspecto, a preocupação primordial é dar garantias e proteções contra os riscos a que estarão sujeitos os participantes.

Em relação aos riscos que podem influenciar o sucesso de um projeto, estes podem ser de inúmeras naturezas e de difícil mensuração. Dessa forma, quando se opta por um *project finance*, a gestão e alocação dos riscos entre os agentes envolvidos torna-se questão fundamental. Essa alocação não é simples e depende dos tipos de agentes que serão envolvidos.

Uma característica interessante de um *project finance* é a sua capacidade de alavancar capital. O empreendedor entra com uma parcela de capital próprio (em geral de 20% a 40%) e financia o restante com as garantias oferecidas pelo projeto. Isso possibilita ao empreendedor chegar a um arranjo de capital onde seu desembolso direto ou seu endividamento podem ser bastante reduzidos. A empresa aumenta, portanto, sua capacidade de investimento.

Como vimos anteriormente, quando uma empresa tem em sua estrutura de capital uma proporção de endividamento bastante elevada, o seu risco de crédito aumenta dificultando outros aportes de capital. Em tese, o *project finance* tem a vantagem de permitir alavancagem elevada de projetos permitindo endividamento das empresas empreendedoras na proporção de sua participação e não na proporção de investimento total efetuado no projeto. Isso ocorre porque um projeto com esse tipo de financiamento se apresenta, usualmente, como entidade independente com atividade contábil separada das demais da empresa. Esse arranjo é chamado de *off-balance-sheet* e, portanto, não consta na contabilidade da empresa como dívida. Essa vantagem tem grande efeito para as empresas empreendedoras que não têm seus riscos de crédito aumentados.⁸

Em paralelo, o custo do capital próprio é, em geral, mais elevado que o de recursos provenientes de terceiros. Por conseguinte, a necessidade de alavancar o empreendimento significa aumentar a remuneração do capital próprio (*equity*).

21.4 Conclusão

Apesar das diferentes abordagens teóricas, é possível afirmar que as inter-relações entre a esfera produtiva e a esfera financeira, interpretadas à luz dos problemas entre decisão de investimentos e de financiamento, revelam que as características técnicas e econômicas de um projeto de investimento podem ser influenciadas pela estrutura de endividamento da empresa e pelas condições de financiamento impostas pelo sistema financeiro. Essa constatação deixa claro os limites da abordagem de MM.

Em 1974 Stiglitz⁹ forneceu contribuições que permitem retomar o problema dessas inter-relações, classificando as decisões de expansão da empresa segundo:

1. As diferentes modalidades de financiamento
2. A forma de distribuição de suas receitas
3. Os montantes dos programas de investimento
4. Os projetos específicos a serem desenvolvidos e a seleção das tecnologias de produção

Segundo Stiglitz, (1) e (2) são decisões típicas de financiamento, ao passo que (3) e (4) se referem às decisões de investimento. Para este autor, é impossível tratar esses quatro grupos de decisão de forma separada, pois os sinais referentes, por exemplo, ao comportamento da política financeira da empresa constituem um sinal importante para a rentabilidade esperada dos investimentos. Dessa forma, a capacidade de mobilização de recursos depende da estrutura de financiamento, mas também do ambiente financeiro, isto é, da organização das instituições financeiras e bancárias. Na mesma direção de análise, o exame da interdependência entre sistema financeiro e a esfera produtiva deve buscar focalizar o problema da mitigação dos riscos industriais e financeiros.

Desse modo, não é possível ignorar que as instituições financeiras estão atentas às modalidades de financiamento das empresas e às escolhas do tipo de investimento a ser implementado.

A este respeito ver especialmente as críticas de Stiglitz, J. (1969), “A Re-examination of the Modigliani-Miller Theorem”, *American Economic Review*, v. 59, dezembro, p. 784-793, a respeito do debate provocado sobre as relações entre decisão de financiamento e investimento.

Notas

1. Expressão utilizada por Goldsmith, R.W., in “Prologomènes à l’analyse comparative des structures financières”, *Revue d’Economie Politique*, maio-junho, 1970, pp. 395-424.
2. Berle, A.; Means, G. (1932) *The modern corporation and private property*. Nova York: Mac Millan.
3. O indicador γ é a relação endividamento-capitais próprios para o programa de investimentos a ser realizado. Se a empresa possui apenas os ativos referentes a um único projeto, o valor de γ se confundirá com o valor de D / S.
4. Outra consequência importante: as decisões da empresa quanto à arbitragem entre a política de dividendos a serem distribuídos e o montante de recursos a serem reinvestidos (a parcela de autofinanciamento) são irrelevantes para o valor de mercado da empresa.
5. Ezra Solomon definiu, de maneira irônica, o mercado perfeito de capitais como aquele que satisfaz as condições estabelecidas por Modigliani-Miller. Ver Solomon, E. (1966) “What Should we Teach in a Course in Business Finance?”, *Journal of Finance*, v. XXI, maio.
6. O que explica, nos dias de hoje, a importância das chamadas agências de avaliação de risco, como por exemplo a Standard and Poor’s e a Moody’s. O papel dessas agências é precisamente tentar reduzir os problemas de informação imperfeita e assimétrica que são observados nos mercados de capitais.
7. Podendo também assumir outras denominações como: SPC (Special Purpose Company); CPE (Companhia de Propósito Específico) etc.
8. Vale observar que a *Securities and Exchange Commission* (SEC) americana tem sido mais rígida em relação às apresentações contábeis das companhias e, mesmo que os *project finance* não sejam incorporados aos balancetes, tem-se obrigado a apresentação desses empreendimentos em notas de rodapé. Com isso, essa vantagem que as empresas esperam obter em relação à sua proporção de endividamento pode ser ilusória e não transparente às empresas que divulgam ratings de riscos.
9. Stigliz, J. (1974) “On the Irrelevance of Corporate Financial Policy”, *American Economic Review*, v. 64, n. 6, p. 851-866.

Bibliografia

- Berle, A.; Means, G. (1932) *The modern corporation and private property*, New York, Mac Millan.
Goldsmith, R. W. (1970) “Prologomènes à l’analyse comparative des structures financières”, *Revue d’Economie Politique*, maio-junho, p. 395-424.
Solomon, E. (1966) “What Should we Teach in a Course in Business Finance?”, *Journal of Finance*, v. XXI, maio.
Stiglitz, J. (1969) “A Re-examination of the ‘Modigliani-Miller Theorem’”, *American Economic Review*, v. 59, dezembro, 1969, p. 784-793.
Stigliz, J. (1974) “On the Irrelevance of Corporate Financial Policy”, *American Economic Review*, v. 64, n. 6, p. 851-866.



PARTE VI

POLÍTICAS E REGULAÇÃO DOS MERCADOS

Defesa da Concorrência

Maria Tereza Leopardi Mello

22.1 Introdução

Uma política de defesa da concorrência tem por finalidade garantir a existência de condições de competição, preservando ou estimulando a formação de ambientes competitivos com vistas a induzir, se possível, maior eficiência econômica como resultado do funcionamento dos mercados. Em princípio, a concorrência pode (e deve) ser promovida e defendida em vários âmbitos de atuação do Estado, mas existem sistemas legais especificamente voltados para essa finalidade (no Brasil como em diversos países) – que consistem nas chamadas *leis de defesa da concorrência* (ou *leis antitruste*), conforme o Quadro 22.1.

Este capítulo fornece uma visão panorâmica dos objetivos, características e principais conceitos e parâmetros de análise envolvidos na aplicação da lei antitruste. O tema se situa numa intersecção do Direito e da Economia; sua análise constitui sempre

QUADRO 22.1 DEFESA DA CONCORRÊNCIA E LEI ANTITRUSTE NO BRASIL

Juridicamente, a lei antitruste no Brasil (Lei n. 12.529/2011) decorre de norma constitucional que determina a repressão ao “... *abuso de poder econômico que vise à dominação dos mercados, à eliminação da concorrência e ao aumento arbitrário de lucros.*” (Constituição, art. 173, § 4º). Mas “livre concorrência” também é princípio constitucional da Ordem Econômica (cf. art. 170, IV), o que significa uma diretriz geral que deve orientar todas as ações dos poderes públicos (tanto a produção legislativa quanto as ações de governo e as decisões judiciais). Portanto, defesa da concorrência não se resume apenas à lei antitruste e ao arcabouço institucional voltado para sua aplicação, mas também a todas as ações do Estado relacionadas a ela; todos os agentes do poder público que tomam decisões que possam afetar as condições de concorrência nos mercados devem levar em conta o princípio.

No âmbito do Executivo, o órgão atualmente encarregado da aplicação da lei antitruste é o Conselho Administrativo de Defesa Econômica (o CADE, uma autarquia federal). Também participa do sistema de defesa da concorrência a Secretaria de Acompanhamento Econômico do Ministério da Fazenda (SEAE), que é responsável por opinar sobre os aspectos concorrenciais de atos normativos (leis e regulamentos, inclusive aqueles emanados de agências reguladoras), exercendo um papel que pode ser caracterizado como de “advocacia da concorrência” junto à Administração Pública Federal.

O CADE atua na defesa da concorrência como *direito difuso*. Tem poder de determinar a cessação de uma prática anticompetitiva e aplicar multas, além de autorizar certos atos de concentração. Se, ademais, uma prática anticompetitiva acarretar também danos privados (a empresas concorrentes, por exemplo), uma ação de indenização pode ser proposta, mas no Judiciário, cf. art. 47 da Lei n. 12.529/2011. De qualquer modo, prejuízos privados só podem ser motivo de ação com base na legislação antitruste se a eles se associarem também prejuízos à concorrência (ver Seção 22.1 deste capítulo).

QUADRO 22.2 A LEGISLAÇÃO NORTE-AMERICANA DE DEFESA DA CONCORRÊNCIA

A primeira lei antitruste americana – conhecida como *Lei Sherman* – data de 1890. No essencial, essa lei buscou consolidar e sistematizar questões que já eram discutidas anteriormente pelo Judiciário sob o prisma privado (antes da lei, empresas prejudicadas por atos anticompetitivos de suas concorrentes já podiam contestar judicialmente tais atos alegando prejuízos individuais). A inovação mais importante da *Lei Sherman* foi permitir a contestação de contratos, acordos ou práticas comerciais por iniciativa do Estado ou de terceiros (*isto é*, permitiu tratar a concorrência como bem jurídico de interesse da sociedade em geral, e não apenas das empresas eventualmente prejudicadas por práticas anticompetitivas). Tal lei é composta de duas seções:

Seção 1: proíbe “contratos, combinações em forma de truste ou de outro tipo, ou conspirações para restringir o comércio ...” (essencialmente, proíbe cartéis explícitos);

Seção 2: proíbe tentativas de monopolizar mercados, por iniciativa individual ou combinada (o que não implica a proibição de monopólios em si, quando alcançados por meios competitivos normais).

A forma genérica da *Lei Sherman* levou os legisladores a detalhar melhor a repressão a certas condutas por uma nova lei – a *Lei Clayton*, de 1914 – que, no essencial, proíbe:

- na Seção 2, a discriminação de preços com efeitos anticompetitivos (esta seção foi posteriormente emendada pela *Lei Robinson-Patman*, de 1936);
- na Seção 3, as práticas de venda casada (*tie-ins*) e de acordos de exclusividade (*exclusive dealing*), quando gerem prejuízos à concorrência;
- na Seção 7, as fusões que possam prejudicar a concorrência (posteriormente emendada pela *Lei Celler-Kefauver*, em 1950);
- na Seção 8, o controle de empresas competidoras por meio de participações cruzadas nas respectivas direções executivas (*interlocking directorates*).

Também em 1914 foi criada por lei a *Federal Trade Commission (F.T.C.)*, responsável – ao lado do Departamento de Justiça (*Department of Justice – D.o.J.*) – pela aplicação das leis antitruste e pela promoção de ações de defesa da concorrência perante o Judiciário.

O caráter jurisprudencial do sistema jurídico nos EUA torna tão ou mais importante que a própria lei a jurisprudência dos tribunais criada pela sua aplicação ao longo do tempo.

um híbrido de aspectos jurídicos e econômicos;¹ por isso, referências à legislação são imprescindíveis. Algumas referências ao sistema antitruste americano também serão úteis, pois a lei desse país, embora não tenha sido a primeira, é, sem dúvida, uma das mais antigas no mundo e em torno da experiência de sua aplicação desenvolveu-se rica literatura econômica e jurídica que serve de referência para discutir certas questões antitruste, independentemente dos sistemas de cada país. O Quadro 22.2 contém uma breve história da legislação antitruste americana.

Este capítulo está organizado da seguinte forma, além desta introdução. Na seção seguinte são apresentados os principais objetivos e características da defesa da concorrência. Na Seção 22.3 os conceitos básicos da análise antitruste. Na Seção 22.4 são descritos os padrões da ação antitruste. E, finalmente, na Seção 22.5 são apresentados temas para discussão.

22.2 Defesa da Concorrência: Objetivos e Características

A política de defesa da concorrência busca limitar o exercício do poder de mercado pois, em princípio, empresas que detêm esse poder são capazes de prejudicar o processo competitivo, gerando ineficiências como resultado de seu exercício. Note-se, no entanto, que a lei antitruste não torna o poder de mercado – nem os monopólios – ilegais, mas apenas tenta controlar a *forma* pela qual esse poder é adquirido e mantido. A lei procura reprimir o exercício *abusivo* de poder de mercado, e não o poder em si.²

É implementada, normalmente, por meio de dois padrões básicos de ação: (a) aquele voltado para as condutas dos agentes no processo competitivo; (b) aquele relacionado aos parâmetros *estruturais* que condicionam tais condutas.

As regras relativas à *conduta* preveem punições às práticas anticompetitivas (*restritivas* da concorrência) derivadas do exercício abusivo de *poder de mercado*. Tais práticas abusivas podem ter natureza horizontal ou vertical, como será visto na Seção 22.3.1 e, independentemente de prejudicarem uma ou outra empresa concorrente, serão consideradas ilícitas se restringirem o *processo concorrencial*, prejudicando, em última análise, os consumidores. A essas regras atribui-se, em geral, o caráter *repressivo*.

As ações de caráter *estrutural* buscam evitar o surgimento de estruturas de mercado *mais* concentradas – que aumentem a probabilidade de exercício abusivo de poder de mercado – por meio do *controle preventivo* sobre os chamados *atos de concentração* (fusões, aquisições, *joint-ventures* etc.), que também podem apresentar natureza horizontal ou vertical.³

Tanto num caso como no outro, a lei de defesa da concorrência não impõe aos agentes econômicos a obrigação de efetivamente competir nem diz por qual forma os agentes devem fazê-lo. Apenas busca canalizar as forças de mercado e as estratégias das empresas na direção da competição e, com ela, da inovatividade e da eficiência, *evitando que o processo concorrencial*

seja restringido por agentes com poder suficiente para isso. Nesse sentido, não age diretamente sobre os resultados desse processo, mas sim nos meios que levam a esse resultado. Em outras palavras, a lei antitruste não impõe aos agentes obrigações que assegurem diretamente os resultados positivos associados idealmente à concorrência; trata-se, ao contrário, de um tipo de regulação reativa do Estado que impõe, ao agente, o dever de abster-se de praticar certos atos: “cumpre-se” a lei enquanto não se prejudica o processo concorrencial.

Daí duas características peculiares da política antitruste: (a) o objetivo de defender o processo de concorrência implica a necessidade de reprimir qualquer tipo de prática que provoque o efeito de restringir esse processo; será esse efeito que irá caracterizar um ato como proibido perante a lei antitruste;⁴ (b) as imposições decorrentes da lei são, substancialmente, abs-tencões (*não produzir efeitos anticompetitivos*).

Uma última observação é a de que o objetivo da lei antitruste consiste em defender o *processo concorrencial*, e não os concorrentes individualmente considerados e tampouco, de forma direta, os consumidores. A proteção do processo competitivo, cuja existência, supõe-se, gera resultados socialmente aproveitáveis (bem-estar), é de interesse de toda a sociedade – e não deste ou daquele concorrente em particular – e como tal é juridicamente protegido. Trata-se, assim, de um típico *interesse difuso*, de que é titular um grupo indeterminável de pessoas (Quadro 22.1).⁵

A observação é necessária para distinguir o âmbito de incidência da lei antitruste do de outros sistemas jurídicos que lhe são próximos – tais como o de repressão à *Concorrência Desleal* e o de *Defesa do Consumidor* – mas que protegem interesses de natureza diversa, e se diferenciam pelos objetivos visados, pelas técnicas de análise, pelos conceitos utilizados e pelos canais institucionais competentes para garantir os direitos protegidos. O sistema de repressão à Concorrência Desleal incide sobre as relações entre particulares (concorrentes); o interesse tutelado de forma imediata é o do concorrente que pode sofrer danos por atos *desleais* (como, por exemplo, divulgar informação falsa sobre produto de concorrente para desviar sua clientela). No caso da Defesa do Consumidor, os interesses tutelados pela lei são aqueles de consumidores *finals* – individual ou coletivamente. Embora interesses de consumidores possam ser afetados por práticas anticompetitivas ou pelo fato de a concorrência estar ausente num certo mercado, isso não é uma condição necessária para a aplicação da lei do consumidor ou para o reconhecimento de direitos (há direitos de consumidor independentemente de haver danos à concorrência).

22.3 Conceitos Básicos da Análise Antitruste

Toda análise antitruste gira em torno da noção de *poder de mercado*: danos ou restrições à concorrência só podem ser causados por empresas detentoras desse *poder*, que é, portanto, *condição necessária* para haver ilicitude do ponto de vista da lei. Mas não é suficiente, pois a ilicitude depende ainda da constatação de *efeitos anticompetitivos* (efetivos ou potenciais) que decorram de uma conduta ou de um ato de concentração.

Contudo, certas condutas restritivas ou certos atos de concentração, ainda que provoquem efeitos negativos sobre a concorrência, podem também gerar ganhos de *eficiência* que os compensem. Dentre os ganhos típicos de eficiência estão, por exemplo, reduções de custo associadas a economias de escala e escopo, aumentos de produtividade ou qualidade, aperfeiçoamentos tecnológicos, economias de custos de transação etc. Nesses casos, há consenso no sentido de que tais condutas ou atos de concentração não devem ser proibidos quando seus eventuais efeitos restritivos forem devidamente compensados pelas *eficiências* por eles geradas; caso contrário a aplicação da lei provocaria ineficiências nos mercados e teria um resultado contrário ao interesse social.

A partir das afirmações feitas nos dois parágrafos anteriores podemos identificar os conceitos usados em toda análise antitruste: em primeiro lugar, é necessário identificar a existência de *poder de mercado*, para o que se requer, como passo logicamente prévio, a delimitação do mercado em que tal poder é exercido (conceito de *mercado relevante*) e a análise das *condições de mercado* que tornam provável (ou não) o exercício desse poder de mercado. Tendo em vista que uma mesma conduta/ato pode apresentar efeitos ambíguos sobre a concorrência – restritivos, mas geradores de *eficiências* –, é necessário identificar esses dois tipos de efeitos e ponderá-los a fim de verificar qual deles prevalece, para só proibir condutas/atos que apresentarem *efeitos anticompetitivos líquidos*. À análise desses conceitos serão dedicadas as subseções seguintes.

22.3.1 MERCADO RELEVANTE

Delimitar o *mercado relevante* é um passo prévio essencial da análise antitruste, pois é em relação a ele que se calculam os indicadores de concentração e se analisam todas as demais condições necessárias à caracterização do poder de mercado e ao seu exercício e os danos à concorrência. Esse *mercado* deve ser definido caso a caso, e o adjetivo que o acompanha (*relevante*) se

QUADRO 22.3 MERCADO RELEVANTE – DEFINIÇÕES NOS EUA E NO BRASIL

Nos EUA, conforme o documento *Diretrizes para Análise de Fusões Horizontais (Horizontal Merger Guidelines)* do FTC/DoJ, de 1992, um mercado relevante para a análise antitruste é definido como

... um produto ou grupo de produtos e uma área geográfica na qual ele é produzido ou vendido, tal que uma hipotética empresa maximizadora de lucros, não sujeita a regulação de preços, que seja o único produtor ou vendedor, presente ou futuro, daqueles produtos naquela área, poderia provavelmente impor pelo menos um pequeno, mas significativo e não transitório aumento no preço, supondo que as condições de venda de todos os outros produtos se mantêm constantes. Um mercado relevante é um grupo de produtos e uma área geográfica que não excedam o necessário para satisfazer tal teste.

Na lei brasileira, todas as referências à posição dominante (poder de mercado) são feitas por relação a um *mercado relevante* que, entretanto, não é definido no texto legal. Usaram-se, durante muito tempo, as definições da literatura e a do sistema americano citadas acima. O conceito foi, finalmente, *institucionalizado* na Resolução 15/98 do CADE, nos seguintes termos: o mercado relevante do produto compreende “*todos os produtos/serviços considerados substituíveis entre si pelo consumidor, devido às suas características, preço e utilização*”. Em sua dimensão geográfica, o mercado relevante “*compreende a área em que as empresas ofertam e procuram produtos/serviços em condições de concorrência suficientemente homogêneas em termos de preços, preferências dos consumidores, características dos produtos/serviços*”. Considera-se que integram o mercado as *empresas* que possam iniciar a oferta se houver um pequeno, mas substancial aumento de preços; devendo-se identificar os “*obstáculos à entrada de produtos ofertados por empresas situadas fora dessa área*”.

refere à sua pertinência para o caso sob julgamento: trata-se de identificar o(s) mercado(s) em que atuam os agentes envolvidos e no(s) qual(is) ocorrem os supostos efeitos restritivos de uma conduta ou ato de concentração.

O mercado relevante é definido como um *locus* (produto/região) em que o poder de mercado possa (hipoteticamente) ser exercido. Os principais conceitos utilizados para isso são as *elasticidades-preço* da *demand*a e da *oferta*. Existe uma clássica definição estabelecida pelos órgãos antitruste dos EUA, que é universalmente adotada, inclusive no Brasil (embora com uma formulação diferente, conforme pode ser visto no Quadro 22.3).

A tarefa de delimitar um mercado relevante consiste num exercício hipotético que busca, por aproximações sucessivas, estabelecer um grupo de *produtos* e uma área geográfica na qual, em relação a tais produtos, um hipotético monopolista possa elevar preços e auferir *maiores* lucros com isso (ou, dito de outra forma, no qual o exercício de poder de mercado, que se busca prevenir ou reprimir, seja *logicamente* possível). Avalia-se a reação da *demand*a ao hipotético aumento de preços, pois este pode não ser lucrativo se os consumidores puderem buscar outros produtos ou produtores de outros locais.

Na dimensão *produto*, identificam-se os produtos que concorrem entre si, incluindo não apenas os produtos idênticos (substitutos perfeitos), mas também os *substitutos próximos*.⁶ Deve haver *alta substitutibilidade* dos produtos considerados *dentro* do mercado; ao mesmo tempo, deve ser *baixa* a substitutibilidade destes em relação aos produtos considerados *fora* do mercado. Em outras palavras, tanto a demanda pelo(s) produto(s) do mercado relevante quanto sua oferta devem ter *elasticidades-preço* suficientemente baixas para que um eventual aumento de preço resulte em *maiores* lucros – e não menores – para a empresa que hipoteticamente exerce poder de mercado.

Essa identificação é feita, em princípio, pelo lado da *demand*a,⁷ observando-se que não basta a mera possibilidade técnica de substituição para determinar a inclusão de um produto no mesmo mercado que outro, mas sim que os demandantes o façam habitualmente.

Em sua dimensão *geográfica*, o mercado relevante é definido como uma área na qual os produtos (e seus substitutos) são produzidos ou vendidos, o que varia conforme o tipo de produto e tecnologia, custos de transporte, sistema de distribuição, barreiras tarifárias e não tarifárias às importações etc. Procura-se delimitar a área sujeita à atuação de uma empresa hipoteticamente monopolista para detectar, na hipótese de aumentos de preços: (a) se os consumidores podem comprar o produto em outras localidades a custos acessíveis; ou (b) se concorrentes de outras localidades podem direcionar suas vendas para essa região a custos acessíveis.

Uma vez identificado um mercado [produto/área] possível, o passo seguinte é proceder ao *teste do monopolista hipotético*: verifica-se se uma empresa hipotética, maximizadora de lucros e detentora de um também *hipotético monopólio* da oferta – a hipótese mais pessimista para o bem-estar – no mercado considerado (produto/área), é capaz de impor um aumento de preço *significativo e persistente* para caracterizar um *exercício de poder de mercado*. O *mercado relevante* é definido, então, como o *menor* mercado possível (o menor agregado de produtos, combinado com a menor área) que satisfaça o critério anterior.

A delimitação do mercado relevante depende de quanto se supõe que deva aumentar o preço para configurar o suposto exercício abusivo de poder de mercado. No Brasil costuma-se usar o parâmetro de 10%.⁸ A definição de qual deva ser esse

percentual de aumento de preço é, do ponto de vista econômico, arbitrário, embora seja absolutamente imprescindível do ponto de vista jurídico para possibilitar a aplicação da lei. Esse ponto merece ser mais bem explicado.

Para se definir um mercado relevante, identificam-se, inicialmente, um certo grupo de produtos e uma área possíveis de serem considerados *um mercado*. Procede-se, então, ao teste do monopolista hipotético, tomando por parâmetro um certo percentual de aumento de preço (suponhamos 10%): verifica-se se um monopolista seria capaz de impor um aumento de 10% no preço [daqueles produtos naquela área]. O mercado [produto/área] passará no teste se o preço de monopólio (que maximiza os lucros do monopolista hipotético) corresponder tão exatamente quanto possível a esses 10% de aumento – nem mais nem menos. Dentre outras, duas hipóteses em que o mercado [produto/área] não atende às condições do teste podem ser pensadas:

1. Se a elasticidade-preço da demanda for muito alta, o suposto monopolista hipotético não conseguiria elevar o preço suficientemente (a 10%), o que denota que o mercado [produto/área] foi definido de forma muito restrita. Será necessário redefini-lo – pela inclusão de mais produtos possíveis substitutos e/ou pela ampliação da área geográfica considerada – até que se encontre uma combinação [produto/área] em relação à qual a elasticidade da demanda seja menor, tornando possível o tal aumento de preços em 10%.
2. Se a elasticidade-preço da demanda for muito baixa, o suposto monopolista hipotético seria capaz de elevar seu preço num percentual superior aos 10% tomados como parâmetro (o preço de monopólio que maximiza lucros estaria num nível superior ao preço competitivo aumentado em 10%). Isso denota que o mercado [produto/área] foi definido de forma demasiadamente ampla, pois existe uma outra combinação possível [de produto/área], mais restrita, na qual o aumento de 10% poderia ser imposto (o nível de preço que maximiza os lucros do monopolista corresponderia ao preço aumentado em 10%). Portanto, também nesse caso o mercado deve ser redefinido pois, de acordo com o conceito, busca-se o *menor* mercado possível para satisfazer o critério (“...que não excedam o necessário para satisfazer tal teste”). Deve-se, então, procurar uma combinação [produto/área] mais restrita, para a qual a elasticidade-preço da demanda seja maior.

Percebe-se, assim, que se alterarmos o percentual de aumento de preço considerado, os resultados poderiam ser diferentes. Se tomarmos como parâmetro 5%, por exemplo, a hipótese (A) poderia ter outra solução: o monopolista hipotético – que não seria capaz de impor um aumento de 10% – poderia ser capaz de aumentar seu preço em 5%, e assim o mercado [produto/área] estaria adequadamente definido (seria um *mercado relevante antitruste*).

As hipóteses anteriores apontam para um resultado geral importante: para uma dada função de demanda, quanto *mais alto* o limiar de aumento de preço tomado como referência, *menor* será a elasticidade-preço da demanda necessária para que o mercado [produto/área] passe no teste.⁹

A Figura 22.1 ilustra esse resultado, representando, de forma geral, o exercício que define o mercado relevante. A demanda (D) é suposta linear para simplificar.

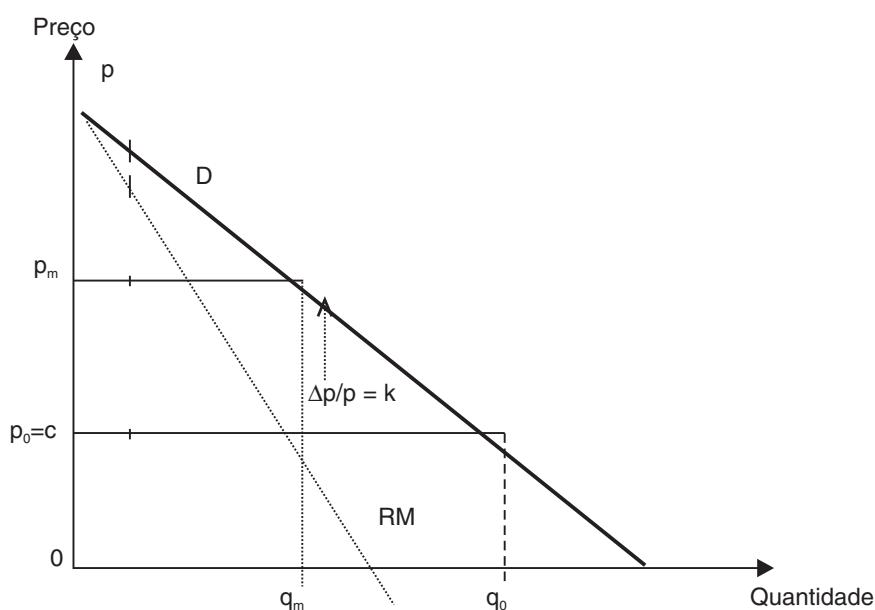


Figura 22.1 Mercado relevante antitruste.

Suponhamos uma posição inicial competitiva, em que o preço (p_0) se situa no nível competitivo (igual ao custo unitário c , no caso suposto constante e, portanto, igual ao marginal). Se monopolizado o mercado (ou cartelizado, com maximização conjunta de lucros), o preço será elevado ao nível que maximiza os lucros, p_m , ao qual a receita marginal iguala os custos marginais ($= c$). Se (e só se) tal elevação de preços se der à taxa considerada abusiva ($\Delta p/p = k$) – o que depende da elasticidade-preço da demanda –, então o mercado assim delimitado (nas dimensões produto e geográfica) será um *mercado relevante antitruste*.

Finalmente, como já mencionado, a elasticidade da oferta também deve ser considerada na delimitação do mercado relevante, embora não esteja incorporada no exercício anterior. Isso é feito num passo seguinte, depois de o mercado relevante ter sido definido (com base na demanda), incluindo-se a oferta *potencial* do(s) produto(s) considerado(s) – *isto é*, considerando o potencial de aumento da oferta pela reorientação de capacidade já existente para a produção dos produtos em questão – por empresas que, por algum motivo, não estejam produzindo, mas possam fazê-lo sem grandes investimentos e a curto prazo. Quanto menores estes (prazos e custos), maiores as possibilidades de que um aumento de oferta por acréscimo de ofertantes venha a contestar o poder de mercado das empresas estabelecidas.

A delimitação do mercado relevante afeta diretamente os resultados de um julgamento: para um dado volume de vendas da(s) empresa(s) envolvida(s), quanto menor um mercado, maiores as possibilidades de existir poder de mercado e, portanto, potencial de danos à concorrência; ao contrário, quanto maior, mais diluída será a participação da empresa investigada e menor a probabilidade de haver efeitos anticoncorrenciais.

Em princípio, o conceito é suficientemente adaptável a qualquer situação e a qualquer tipo de mercado; e não há limite de tamanho (nem para mais nem para menos) do mercado: tanto podemos ter um mercado *mundial* quanto um restrito a um município, tanto um mercado de um só produto quanto de vários etc.

Uma vez identificado o *mercado relevante* pelo exercício hipotético acima descrito, passa-se à análise de suas condições concretas – o grau de concentração, a parcela de mercado da *empresa* envolvida no caso etc. A utilização de medidas de concentração de mercado e a avaliação do nível das barreiras à entrada são os principais instrumentos para análise do *poder de mercado*.

22.3.2 PODER DE MERCADO

Poder de mercado está associado à capacidade de restringir a produção e aumentar preços para, não atrair novos competidores, obter lucros *acima do normal*; é definido como poder de fixar preços significativa e persistentemente acima do nível competitivo, *isto é*, dos custos médios. Ressalte-se que o poder de mercado não se expressa apenas em preços – grande parte das condutas consideradas anticompetitivas não ocorre via preços; mas a definição é utilizada por ser simples e de fácil aplicação e implica a suposição de que a empresa capaz de elevar preços significativa e persistentemente acima dos custos possui poder de mercado e pode exercê-lo por qualquer outro meio disponível.¹⁰

O critério de participação da *empresa* no mercado (market-share) é uma primeira aproximação para avaliá-lo. Além disso, recorre-se também aos indicadores de concentração de mercado, como o CR(4) e o HH vistos no Capítulo 5, para verificar a *possibilidade estrutural* de existência de poder de mercado.¹¹

A análise antitruste, coerente com o modelo *Estrutura-Conducta-Desempenho*, parte do suposto de que maior concentração da oferta (estrutura) implica maior probabilidade de colusão (conduta) e, por consequência, de preços e lucros mais elevados. Presupõe-se que o poder de mercado é função crescente da concentração, pois a existência de *pequeno número* de concorrentes e/ou sua *desigualdade* favorecem conluios tácitos ou explícitos, pelo menos entre as empresas líderes. Essa, entretanto, é uma aproximação precária, pois a concentração é apenas condição necessária, mas não suficiente, para o surgimento de poder de mercado; não há correlação perfeita entre ambas as variáveis.¹² Veja o Quadro 22.4 para uma definição de poder de mercado na lei brasileira.

QUADRO 22.4 PODER DE MERCADO NA LEI BRASILEIRA

A lei brasileira utiliza o termo *posição dominante* para designar, praticamente, o mesmo conceito que *poder de mercado*. Posição dominante é definida como *capacidade de alterar unilateral ou coordenadamente as condições de mercado* ou pela detenção de uma parcela de mercado de 20% ou mais de um mercado relevante.

A posição dominante também pode ser exercida coletivamente, caso em que a parcela de mercado a ser considerada deve ser calculada em relação ao grupo de empresas envolvidas num determinado caso.

Esse parâmetro, contudo, gera apenas uma *presunção* – não absoluta; nada impede que se discuta e comprove que, apesar de deter uma parcela de mercado superior a 20%, outras variáveis fazem com que a empresa não possa exercer poder de mercado (e, portanto, não detenha posição dominante).

Medidas de participação no mercado e indicadores de concentração isoladamente podem não significar muito, se não forem analisados em conjunto com outros fatores, entre os quais se destacam as condições de entrada na indústria, a existência de competidores potenciais e a dinâmica da concorrência.

*Barreiras à entrada*¹³ são uma condição decisiva para a avaliação de poder de mercado. Sabe-se que na ausência de barreiras à entrada não é possível fixar preços acima dos custos de forma persistente e significativa (portanto, não é possível exercer poder de mercado). Por isso, as *barreiras à entrada* são um elemento fundamental de análise antitruste, não só para *atos de concentração*, como também (quase sempre) para *condutas*.

Apesar de não permitir mensuração totalmente objetiva, o *nível* das barreiras à entrada num mercado concentrado (oligopólio ou monopólio) é o principal instrumento de avaliação do poder de mercado das empresas que nele atuam. Entretanto, numa visão dinâmica, ele deve incluir a *capacidade inovativa* da indústria e das empresas concorrentes potenciais, pois a *inovação* é o principal antídoto tanto às barreiras à entrada quanto à possibilidade de abuso de poder de mercado pelas empresas dominantes.

A existência de poder de mercado (e a possibilidade de seu exercício) é condição necessária para a aplicação da lei antitruste, pois quem não o detém não poderá prejudicar a concorrência no mercado, ainda que queira. Portanto, se com esse passo de análise se conclui que a(s) empresa(s) envolvida(s) num caso não possui(em) – individual ou coletivamente – tal poder, o caso deve ser encerrado, pois não é relevante do ponto de vista antitruste. Não existe razão para continuar a investigação, pois será logicamente impossível a existência de prejuízo à concorrência (alvo da ação antitruste) – ainda que se possa estar ferindo outros sistemas legais.

22.3.3 EFICIÊNCIAS E PRINCÍPIO DA RAZOABILIDADE

Como já mencionado, condutas ou atos de concentração restritivos da concorrência podem também promover eficiências compensatórias, que devem, então, ser analisadas caso a caso. O que importa, em qualquer caso, são os efeitos líquidos sobre a eficiência econômica. Essa forma de abordagem decorre *logicamente* dos objetivos da lei antitruste e se fundamenta jurídica e economicamente: a lei busca reprimir o abuso de poder de mercado porque ele é gerador de ineficiências; logo, não deve proibir atos/condutas que gerem ganhos de eficiência líquidos pois, se o fizer, gerará ineficiências tão ou mais significativas do que as que visa combater.

A análise das eficiências deve observar algumas condições:

1. Em primeiro lugar, o exame de eficiências é exigido *apenas se houver* um significativo risco de prejuízo à competição como decorrência do ato ou conduta em questão. Caso se avalie que *sequer existe* um efeito anticompetitivo, a análise será interrompida neste ponto e a identificação e a avaliação de eficiências serão desnecessárias.
2. As eficiências consideradas aptas a compensar os efeitos restritivos devem ser *decorrência necessária daquele ato ou conduta*, para justificar sua autorização por parte das autoridades antitruste (*isto é*, devem ser *críveis* por razões estruturais ou por alguma forma de comprometimento efetivo das empresas envolvidas).
3. Deve-se comprovar, ainda, que as eficiências geradas por atos ou condutas restritivos não poderiam ser alcançadas de outra forma menos lesiva ao processo concorrencial. Só se justifica autorizar uma restrição à concorrência se isso for uma condição necessária aos ganhos de eficiência alegados; caso contrário, tais ganhos poderiam ser alcançados com uma hipótese melhor para o bem-estar.

Em suma, do ponto de vista *lógico*, toda e qualquer análise antitruste de ato ou conduta em que se suponha existir *algum efeito anticompetitivo* real ou potencial envolve, como *último passo conclusivo*, a avaliação dos *efeitos líquidos em termos de eficiência* do ato ou conduta examinado.¹⁴

Tal abordagem – que considera os *efeitos líquidos* e não apenas os restritivos – é hoje conhecida no âmbito antitruste como princípio da razoabilidade (*rule of reason*) e tem sua origem ligada à jurisprudência americana. Nos EUA, o Judiciário faz uma distinção entre duas categorias de práticas e aplica, a cada uma, um método de abordagem: (a) o princípio da razoabilidade, que implica a necessidade de avaliação – caso a caso – dos efeitos das práticas investigadas e de suas eventuais justificativas (em termos de eficiências compensatórias); só é considerada ilícita a prática cujos efeitos líquidos sejam negativos; e (b) a chamada abordagem *per se*, aplicada àquelas práticas restritivas às quais não se associam, normalmente, ganhos de eficiência significativos. Neste caso, não se discutem os *efeitos* da prática investigada; basta provar sua ocorrência para que seja considerada ilícita.

Deve-se notar, contudo, que a abordagem *per se* não é incongruente com a análise econômica desenvolvida em torno da importância das eficiências eventualmente geradas por práticas restritivas; trata-se apenas de um método judicial – de aplicação da lei – que simplifica a análise e economiza custos para julgar tipos de condutas cuja *não razoabilidade* é suposta. Ou seja: o critério traduz a percepção de que é desnecessário avaliar pormenorizadamente os efeitos líquidos em face da

QUADRO 22.5 O PRINCÍPIO DA RAZOABILIDADE E A ABORDAGEM ‘PER SE’ NOS EUA E NO BRASIL

Desde o final do século passado, o Judiciário norte-americano se viu na contingência de julgar certas práticas de empresas, cujas defesas se baseavam no argumento de que as restrições proporcionadas à concorrência eram “razoáveis” e, em alguns casos, constituiriam até meio de garantir a própria concorrência. Como a Lei Sherman desde 1890 proibia práticas que restringissem o comércio, estabeleceu-se intenso debate sobre o significado a ser dado a tais *restrições*. De um lado, colocaram-se aqueles que defendiam caber ao Judiciário a análise – caso a caso – dos efeitos das práticas para só condená-las em casos em que fossem considerados negativos (não razoáveis). Por outro lado, defendia-se que qualquer restrição ao comércio deveria ser proibida *per se* (fosse ou não razoável) em nome da segurança jurídica; opinião que prevaleceu nesse primeiro momento. Por exemplo, num caso de 1897, julgava-se um acordo de preços entre concorrentes e a defesa alegou que os preços fixados eram *razoáveis*; argumentou um juiz no caso: “*se somente se incluir no significado da lei aquelas restrições não razoáveis, a determinação daquilo que é razoável será alcançada com grande incerteza*” (J. Peckham, *in US v. Trans Missouri Freight Association*, 1897).

Em casos julgados já no começo do século passado (anos 1910, aproximadamente), o Judiciário americano começou a admitir certos argumentos de defesa com base na razoabilidade dos efeitos restritivos – critério de análise que ficou conhecido como *rule of reason*; manteve-se, todavia, o entendimento de que certas categorias de práticas seriam presumidamente ilegais – *isto é*, devem ser consideradas ilícitas *per se* (*por exemplo*, costuma-se considerar *acordos horizontais para fixação de preços* uma prática ilegal *per se*).

As definições da jurisprudência americana a respeito do que seja ilegal *per se* têm evoluído no sentido de admitir que não há, praticamente, nenhuma categoria de práticas de mercado que possam produzir efeitos apenas negativos; há só aquelas que, *em sua maioria e na maior parte das vezes*, produzem *mais* efeitos negativos que positivos e que, por uma questão de economia processual/administrativa, não precisam passar por uma apuração e discussão mais detalhada sobre seus efeitos:

“*A regra per se requer que o Tribunal proceda a uma ampla generalização a respeito da utilidade social de certas práticas comerciais... Casos que não se enquadram nessa generalização podem aparecer, mas a regra per se reflete o juízo de que tais casos não são suficientemente comuns ou importantes a ponto de justificar o tempo e os gastos requeridos para identificá-las...*”. In caso *Sylvania* de 1984 (apud Cohen, 1997).

Em suma, uma prática analisada sob o critério *per se* não precisa passar por uma apuração mais detalhada porque os custos que isso implicaria superam os benefícios de tal procedimento. Por outro lado, a análise *per se* propicia uma solução mais rápida para o caso e pode servir de *orientação* para os agentes (no sentido de evitar que sejam cometidas práticas que, na maioria das vezes, têm, de fato, mais efeitos restritivos).

Essa diferenciação, note-se, é jurisprudencial e os tipos de práticas incluídas numa e noutra abordagem mudam ao longo do tempo. Observa-se no Judiciário dos EUA a tendência de aceitar certas justificativas para práticas que antes eram consideradas ilícitas *per se*, aumentando, assim, as categorias de práticas que são analisadas com base no princípio da razoabilidade. Isso é fruto de intenso debate e argumentação econômicos que se travam na literatura e perante o Judiciário.

O sistema antitruste brasileiro é compatível com a abordagem do princípio da razoabilidade, tanto no julgamento de condutas quanto na avaliação de atos de concentração.

Em relação às condutas, a lei caracteriza a ilegalidade pelos efeitos: quaisquer atos, independentemente de sua forma, serão considerados *infração à ordem econômica* se produzirem – efetiva ou potencialmente – efeitos contrários à concorrência. Embora a lei atual não o diga expressamente, está suposto que se devem considerar tais efeitos como *líquidos*. Assim, pode-se dizer que a ilicitude de condutas, na nossa lei antitruste, identificada pela noção de *abuso de posição dominante*, é caracterizada: (a) pela existência de efeitos restritivos da concorrência; e (b) pela ausência de efeitos positivos maiores que compensadores dos primeiros.

Em relação aos atos de concentração, a lei prevê expressamente que a aprovação de atos que apresentem potencial de prejudicar a concorrência ou resultem na dominação de mercados relevantes condiciona-se à análise de eficiências geradas que devem ser aptas a compensar o potencial anticompetitivo.

evidência teórica, estatística e jurisprudencial de que determinadas condutas não geram normalmente benefícios relevantes, de forma que sua condenação não ocasionará injustiças, senão excepcionalmente, permitindo, em contrapartida, substancial economia de custos públicos e privados implicados no processamento de um caso pelo princípio da razoabilidade (ver o Quadro 22.5).

22.4 Padrões da Ação Antitruste

É interessante notar que o devido encadeamento lógico dos conceitos analisados na seção anterior estabelece, como consequência, os passos e procedimentos fundamentais de toda análise de casos sob a lei antitruste – que são, no geral, comuns para julgamento de condutas e para controle dos atos de concentração, como pode ser visto na síntese apresentada no Quadro 22.6.

Em primeiro lugar, identifica-se a existência de poder de mercado e das condições para seu exercício, a partir de um mercado relevante previamente definido; avaliam-se os efeitos anticoncorrenciais – efetivos ou potenciais – da conduta ou do ato de

QUADRO 22.6 PASSOS DA ANÁLISE ANTITRUSTE PARA CASOS DE CONDUTAS E ATOS DE CONCENTRAÇÃO

	Condutas	Atos de concentração
I – Verificação da existência de poder de mercado		
1. Delimitação do mercado relevante	Identificar o(s) mercado(s) relevante(s) de atuação da(s) empresa(s) investigada(s) e/ou o mercado afetado pela conduta (se forem diferentes)	Identificar o(s) mercado(s) relevante(s) afetado(s) pelo ato (basicamente, aqueles mercados em que as duas empresas atuam)
2. Análise da posição das empresas envolvidas no(s) mercado(s) relevante(s)	<ul style="list-style-type: none"> – cálculo do <i>market-share</i> da(s) empresa(s) investigada(s), nos mercados relevantes identificados; – indicadores de concentração do(s) mercado(s). 	<ul style="list-style-type: none"> – cálculo do <i>market-share</i> resultante da operação de concentração no(s) mercado(s) relevante(s) afetado(s); – indicadores de concentração e comparação com a situação anterior (pré-concentração)
3. Análise das condições de exercício do poder de mercado	formas de concorrência e grau de rivalidade no(s) mercado(s) relevante(s); avaliação das barreiras à entrada; possibilidade de concorrência por importações.	formas de concorrência e grau de rivalidade no(s) mercado(s) relevante(s); avaliação das barreiras à entrada; possibilidade de concorrência por importações.
	Conclusão: há poder de mercado e condições para seu exercício?	
não	Os envolvidos não detêm poder de mercado; é logicamente impossível haver danos à concorrência. O caso deve ser encerrado.	Implica que o ato não cria nem reforça poder de mercado, devendo ser, portanto, aprovado sem condições.
sim	É possível que a conduta provoque efeitos restritivos; passa-se à etapa seguinte para identificá-los.	O ato provoca efeitos anticompetitivos <i>potenciais</i> ; passa-se à etapa seguinte para verificar se há eficiências compensatórias.
II – Identificação dos efeitos anticompetitivos	São diversificados e variáveis, conforme o tipo de conduta. Se houver algum efeito restritivo, passa-se à etapa seguinte.	Já está suposto existirem tais efeitos em potencial, como resultado da etapa anterior. É preciso, então, avaliar a <i>intensidade</i> deles, que depende do grau de aumento de concentração e da parcela de mercado em relação à situação anterior
III – Identificação das eficiências geradas pelo ato/conduta	Também são variáveis conforme o tipo de conduta. São típicas as economias de custo de transação.	Eficiências típicas: economias de escala e escopo; economias de custos de transação; melhorias tecnológicas. Avalia-se se as eficiências não poderiam ser atingidas de outra forma.
IV – Conclusão/efeitos líquidos		
se positivos	Não há infração à ordem econômica.	O ato deve ser aprovado integralmente.
se negativos	Há infração: ordem para cessar a prática e penalidades (multas e outras)	O ato deve ser proibido ou, o que é mais comum, podem ser impostas condições para sua aprovação.

concentração sobre o(s) mercado(s) relevante(s) considerado(s) e examinam-se os possíveis ganhos de *eficiência* ou outros benefícios gerados por eles. A análise termina por uma ponderação entre efeitos anticompetitivos e ganhos de eficiência. Caso os primeiros prevaleçam, a conduta será considerada ilícita e terá suas devidas consequências jurídicas; ou o ato de concentração será proibido (ou serão impostas condições para sua aprovação).

Veremos a seguir alguns aspectos principais – econômicos e legais – da análise de condutas e atos.

22.4.1 CONDUTAS ANTICOMPETITIVAS

Costumam-se dividir as condutas anticompetitivas em horizontais e verticais. Horizontais são as que reduzem a concorrência entre empresas *de um mesmo mercado*; verticais, as que ocorrem no âmbito das relações entre empresas que se relacionam como compradoras e vendedoras ao longo da cadeia produtiva.

Práticas *horizontais* consistem, basicamente, ou em acordos entre concorrentes ou em concorrência predatória entre eles. Em ambos os casos, implicam reduzir ou eliminar a concorrência do mercado visando ao aumento de poder de mercado – em conjunto, via acordos, ou individualmente, via preços predatórios; a curto ou a longo prazos, respectivamente (ver o Quadro 22.7).

As condutas *verticais* consistem em restrições praticadas por ofertantes de bens/serviços de um mercado (*de origem*) que afetam outro mercado relacionado verticalmente – *a montante* ou *a jusante* –, chamado de *mercado-alvo*. Pressupõem,

QUADRO 22.7 PRINCIPAIS CONDUTAS TÍPICAS HORIZONTAIS E VERTICAIS¹⁵

Horizontais

Preços predatórios são definidos como “prática deliberada de preços abaixo do custo variável médio, visando eliminar concorrentes para, em momento posterior, poder praticar preços e lucros mais próximos do nível monopolista”; para caracterizar a ocorrência de preço predatório é necessária a análise das condições efetivas de custos e do comportamento dos preços ao longo do tempo, a fim de afastar hipóteses de práticas sazonais normais ou de políticas comerciais da empresa, considerando-se, também, a estrutura de mercado provável resultante da eliminação de concorrentes (para testar a consistência da conduta predatória).

Acordos entre concorrentes podem apresentar-se sob diversos tipos, o que repercute na consideração dos efeitos anticompetitivos prováveis e possíveis eficiências. Temos:

1. *Cartéis* – acordos explícitos ou tácitos entre concorrentes, que afetam parte substancial do mercado relevante, envolvendo o estabelecimento de preços, quotas de produção e distribuição e divisão territorial. Em geral, considera-se que não existem benefícios compensatórios que possam decorrer desta prática.
2. *Outros acordos entre empresas* – costuma-se distinguir dos cartéis aquelas “restrições que envolvem apenas parte do mercado e/ou esforços conjuntos temporários voltados à busca de maior eficiência, especialmente produtiva ou tecnológica”, que exigem uma avaliação mais cuidadosa sobre os possíveis benefícios, pois, normalmente, apresentam efeitos anticompetitivos menores que os cartéis.
3. *Acordos de Associações de Profissionais* – práticas que limitam injustificadamente a concorrência entre profissionais, principalmente pelo tabelamento de preços. Uma possível justificativa, aqui, pode ser a alegação de que preços uniformes visam à garantia de qualidade dos serviços prestados. O CADE tem sistematicamente condenado práticas de tabelamento promovidas por associações e sindicatos de profissionais liberais.

Verticais

1. *Fixação de preços de revenda* – estabelecimento, pelo produtor, dos preços (mínimo, máximo ou rígido) a serem praticados pelos distribuidores/revendedores, garantido por ameaça efetiva de sanções pela não observância da imposição. Em geral, efeitos anticompetitivos são maiores quando os preços fixados têm a função de *mínimos* (ou *rígidos* com função de mínimos).
2. *Restrições territoriais e de base de clientes* – limites, impostos pelo produtor, das áreas de atuação dos distribuidores/revendedores, restringindo a concorrência e a entrada em diferentes regiões do *mercado-alvo*.
3. *Acordos de exclusividade* – compromisso de revendedores/distribuidores no sentido de adquirir bens/serviços de determinado produtor (ou vice-versa) com exclusividade, ficando proibidos de comercializar produtos de rivais. O CADE tem sistematicamente condenado esse tipo de conduta praticada por planos de saúde que impõem obrigação de dedicação exclusiva dos médicos conveniados (que ficam impedidos de atender por outros planos).
4. *Recusa de venda/negociação (ou boicote)* – estabelecimento unilateral, pelo fornecedor, de condições de negócio oferecidas a distribuidores/revendedores, em geral utilizado como forma de retaliação contra distribuidores/revendedores relutantes em aderir a acordos de exclusividade ou fixação de preços de revenda. Pode também ocorrer no sentido contrário, isto é, mediante condições impostas unilateralmente por compradores aos fornecedores. Recusa de venda praticada por detentor de controle sobre *infraestrutura essencial (essential facilities)* requer cuidados especiais de análise. Em geral, tais práticas são objeto de regulação específica.
5. *Venda casada* – subordinação da venda de um produto/serviço à aquisição de outro produto/serviço. Normalmente, está relacionada à tentativa de *alavancagem* de poder de mercado (usa-se o poder de mercado detido no mercado do produto *subordinante* para criar ou aumentar o poder no mercado do produto *subordinado*).
6. *Discriminação de preços* – fixar diferenciadamente preços (ou outras condições de venda) de um mesmo produto/serviço para diferentes compradores. Essa prática não é intrinsecamente anticompetitiva, e frequentemente está associada a eficiências.

sempre, a existência de poder de mercado no mercado relevante *de origem* e seus prováveis efeitos restritivos podem se fazer sentir tanto no *mercado-alvo* quanto no *de origem* (quando a prática reforça posição dominante neste). Por isso, a análise irá requerer sempre atenção especial aos mecanismos de interação entre esses dois mercados. As práticas verticais mais típicas estão descritas no Quadro 22.7.

Os principais efeitos anticoncorrenciais das condutas verticais, e que estão presentes em todos os tipos, são a facilitação de práticas concertadas (cartéis) ou o reforço unilateral de poder de mercado de empresa no mercado relevante *de origem*. Outros efeitos restritivos são variáveis conforme o tipo de conduta, tais como: “fechamento” do *mercado-alvo* para concorrentes efetivos ou potenciais (aumento do nível das barreiras à entrada), inclusive por aumento de custos de rivais; exploração monopolista dos usuários de serviços pós-venda; atenuação da concorrência *entre* ou intramarcas.

Também os prováveis efeitos compensatórios associados às restrições verticais variam conforme o tipo de conduta; os mais comuns são: redução de custos de transação¹⁶ pela detenção de condutas oportunistas, incentivando a qualidade dos serviços nos mercados-*alvo* e protegendo a reputação e os investimentos em ativos específicos de empresas no mercado *de origem*; economias de escala e escopo no *mercado-alvo*; proteção ao desenvolvimento tecnológico no mercado *de origem*. O Quadro 22.8 resume a matéria de lei sobre as condutas anticompetitivas.

QUADRO 22.8 CONDUTAS ANTICOMPETITIVAS NA LEI BRASILEIRA E SUAS CONSEQUÊNCIAS JURÍDICAS

Na lei brasileira, a ilegalidade de uma conduta depende de se verificar se ela produz certos efeitos, que são apresentados nos seguintes termos: “limitar, falsear ou de qualquer forma prejudicar a livre concorrência ou a livre iniciativa; dominar mercado relevante de bens ou serviços; aumentar arbitrariamente os lucros; exercer de forma abusiva posição dominante” (art. 36). A lei estabelece, ainda, um elenco de condutas típicas, mas isso é meramente exemplificativo, no sentido de que não esgota todas as condutas possíveis, nem constituem, em si, condutas ilegais (art. 36, § 3º).

Tais efeitos são descritos na lei de modo muito genérico, ficando a cargo da análise econômica identificá-los nos casos concretos e ponderá-los com eventuais eficiências, como explicado na seção anterior. O que é importante, do ponto de vista jurídico, é que qualquer conduta – prevista ou não prevista expressamente na lei – pode ser considerada ilegal, e que tal ilegalidade dependerá sempre da análise dos efeitos econômicos – considerados em termos líquidos.

Em princípio, submetem-se às disposições da lei antitruste qualquer pessoa, física ou jurídica, pública ou privada, ou qualquer associação ou entidade (art. 31). A definição ampla implica que mesmo entidades sem fins lucrativos devem observar as normas da lei antitruste (já que não se faz referência ao tipo de atividade, nem ao propósito lucrativo), o que é importante para não limitar o escopo de aplicação das regras de defesa da concorrência. Assim, por exemplo, sindicatos patronais ou entidades de classe podem ser processados; isso é relevante porque essas entidades são várias vezes usadas como articuladoras de cartéis. A lei também submete à sua disciplina agentes que “...exerçam atividade sob regime de monopólio legal”. Portanto, a regulação de certas atividades, em princípio, não as retira do escopo da lei antitruste, que se aplica a setores regulados, inclusive serviços públicos (neste caso, coloca-se a questão da articulação institucional entre agências regulatórias – ou poder concedente – com as autoridades antitruste).

Quando uma conduta é julgada como infração à ordem econômica, ela fica sujeita a diversos tipos de consequências jurídicas (no plano administrativo, civil e penal):

Sanções administrativas (aplicadas pelas autoridades antitruste): obrigação de fazer cessar a prática ilícita, por determinação do CADE; multa pela infração, que pode variar de 0,1% a 20% do faturamento bruto da empresa ou grupo, no ramo de atividade em que ocorreu a infração; para entidades não empresas a lei prevê multa no valor de R\$ 50 mil a R\$ 2 bilhões de reais (art. 37).

Essas consequências se aplicam sempre. A lei prevê ainda outras penalidades (art. 38), a serem aplicadas a infrações de maior gravidade: publicação da decisão; proibição de contratar com instituições financeiras oficiais e de participar de licitações; recomendação ao INPI para conceder licença compulsória de patentes; pode-se até impor a cisão de sociedade, transferência de controle acionário, e outras medidas necessárias para a eliminação dos efeitos nocivos à concorrência.

Consequências judiciais civis: os prejudicados podem pedir no Judiciário a cessação da prática e indenização por danos, lembrando que esta última só pode ser reivindicada com base na lei antitruste se ao prejuízo individual se associar também algum prejuízo à concorrência (cf. art. 47 da Lei n. 12.529/2011).

Consequências penais: são estabelecidas na Lei n. 8.137/90, art. 4º, que tipifica crimes contra a ordem econômica, para os quais se prevê pena de dois a cinco anos de reclusão ou multa. Essas penas também só são decididas pelo Judiciário.

22.4.2 ATOS DE CONCENTRAÇÃO

O controle dos atos de concentração tem a finalidade de prevenir a criação ou o reforço de poder de mercado que pode resultar da união de dois ou mais competidores (no caso de concentrações horizontais) ou de parceiros comerciais (no caso de integrações verticais).

As principais fontes de ganhos de eficiências em atos de concentração são: economias de escala e escopo; economias de racionalização e especialização; de utilização e de expansão de capacidade; sinergias e outras formas de interação entre ativos complementares; economias em P&D, tecnologia e eficiências dinâmicas; economias de custos de transação (de negociação de contratos, de prevenção de comportamentos oportunistas, de *governance* de relações contratuais) presentes sobretudo em integrações verticais.

Algumas situações típicas de *atos* que apresentam potencial anticompetitivo – e suas respectivas eficiências – podem ser identificadas como se segue:

1. Concorrentes reais ou potenciais se fundem para melhor aproveitar economias de escala, baixando custos, mas possivelmente aumentando preços e lucros;
2. Concorrentes potenciais se unem em *joint-venture* para desenvolver um novo produto ou eliminar investimentos duplicados em P&D e evitar os custos de uma “corrida para chegar primeiro”; isso pode retardar a introdução da inovação no mercado;
3. Concorrentes multiproduto combinam especializar-se, fornecendo um ao outro os insumos necessários; reduzem, assim, os custos e aumentam a especialização de equipamento e pessoal, mas à custa de eliminar a concorrência em qualidade e preços.

Concentração é termo geralmente empregado para identificar situações em que os participantes perdem sua autonomia (fusão, aquisição, incorporação, formação de grupo). Nossa lei atual, todavia, não delimita os tipos formais de atos que devem ser

QUADRO 22.9 CONTROLE DOS ATOS DE CONCENTRAÇÃO NA LEI BRASILEIRA – CRITÉRIOS PARA SUBMISSÃO

A lei brasileira busca prevenir situações que possam levar ao exercício abusivo de posição dominante obrigando a submissão, ao CADE, de quaisquer atos de concentração em que: (a) uma das partes envolvidas tenha faturamento igual ou maior que R\$ 400 milhões de reais; (b) pelo menos uma outra parte tenha faturamento igual ou maior que R\$ 30 milhões (art. 88). Consideram-se como atos *de concentração* as fusões, aquisições, incorporações, consórcios ou joint-ventures (art. 90)

A concentração submetida ao CADE será proibida se implicar a *eliminação da concorrência em parte substancial do mercado relevante, ou a criação ou reforço de posição dominante*, a não ser que apresente, também ganhos que compensem seu potencial anticompetitivo. Tais condições estão elencadas no art. 88, § 6º, da Lei n. 12.529/2011: (a) que haja ganhos de eficiência (aumento de produtividade ou qualidade, ou eficiência e desenvolvimento tecnológico) e (b) que parte relevante dos benefícios seja repassada aos consumidores)

submetidos à análise e, a rigor, nem faz referência à necessidade de que tais atos consistam em *transferência de propriedade* ou qualquer forma de administração compartilhada. Daí que alguns tipos de *contratos*, apesar de não implicarem mudança na estrutura societária nem transferência de propriedade, podem provocar efeitos estruturais nos mercados e, por isso, podem ser obrigados a passar pelo controle prévio das autoridades antitruste: por exemplo, contratos de cooperação, contratos de distribuição, consórcios societários, redes de franquias etc. O controle alcança também as integrações verticais com potencial anticompetitivo.

A avaliação desses atos, como mostrado no Quadro 22.9, envolve análise de seu impacto sobre o(s) mercado(s) relevante(s) da operação, para verificar em que medida provoca a diminuição do grau de concorrência existente antes da concentração. Se provocar a diminuição da concorrência, o passo seguinte é verificar se seus efeitos restritivos podem ser contrabalançados pelas seguintes condições compensatórias (cumulativamente): (a) que o ato tenha por objetivo incrementar a produtividade ou a qualidade de bens/serviços, ou propiciar a eficiência e o desenvolvimento tecnológico; (b) que seus benefícios sejam repartidos com os consumidores; (c) que não implique a eliminação da concorrência de parte substancial do mercado relevante; e (d) que seus efeitos restritivos se limitem ao estritamente necessário para atingir aos objetivos visados.

A condição (a) corresponde às *eficiências* em sentido estrito (eficiência alocativa); as demais constituem salvaguardas mínimas para a concorrência e buscam minimizar os efeitos anticompetitivos que não sejam estritamente ligados às eficiências. Desse modo, para a lei brasileira o enfoque centrado apenas na eficiência alocativa não é referência absoluta; é preciso também demonstrar que um eventual prejuízo ao mercado será mínimo (pelos condições “c” e “d”) e que os ganhos de eficiência serão, pelo menos em parte, repassados para o consumidor (condição “b”).

Em algumas circunstâncias, a aprovação de tais atos pode ser submetida à assinatura de um *Compromisso de Desempenho*, documento para possibilitar o monitoramento, pelas autoridades antitruste, do cumprimento das condições compensatórias que justificaram a aprovação do ato. Há certo consenso no sentido de que tais compromissos devem restringir-se a medidas de caráter estrutural, e não comportamental; isto é, que não sejam garantidos pela mera ameaça de punição ou ainda por uma “promessa de bom comportamento”, pois isso implicaria altos custos, para as autoridades, de monitorá-los e fiscalizá-los, ou por outro lado poderia constituir uma forma de ingerência na definição de estratégias ou de comportamentos privados.

22.5 Conclusão e Temas para Discussão

Iniciamos este capítulo afirmando que a concorrência pode ser promovida em âmbitos diversos de atuação do Estado e não se resume, portanto, à aplicação da lei antitruste. Deve ser um objetivo mais amplo de política e, como tal, se relaciona com outras políticas públicas que possam ter efeitos pró-competitivos.

Diferentemente da aplicação da lei antitruste, que possui um caráter predominantemente defensivo, existem várias possibilidades de políticas que podem visar tanto à conformação de ambientes competitivos como também ao fortalecimento da competitividade (no sentido de capacidade para concorrer) das empresas. Ambos – ambiente e agentes – são necessários para haver, de fato, a concorrência cujos efeitos são reconhecidamente positivos do ponto de vista do interesse social.

Assim, por exemplo, uma *política industrial*¹⁷ tem estreitas relações com uma política de defesa da concorrência. Pode apresentar uma dimensão pró-competitiva, à medida que se preocupa com o nível de eficiência da indústria e busca criar condições para que as empresas adquiram ou reforcem sua competitividade. Ocorre, porém, que política industrial pode ter outras dimensões que, eventualmente, não se harmonizem com a concorrência (proteção às empresas nacionais contra concorrência externa, por exemplo).

Embora identifiquemos política industrial quase sempre com medidas protetoras *contra* concorrência, essa não necessariamente é uma característica intrínseca a toda política industrial e a discussão sobre *se* e *como* seria possível harmonizá-la com a defesa da concorrência é uma questão ainda em aberto.

Outra questão que coloca desafios à compatibilização com a defesa da concorrência é a *regulação*, que será tratada no Capítulo 23. Em princípio, a existência de regulação setorial implica menor grau de liberdade nas decisões dos agentes econômicos regulados e a imposição de certas obrigações a esses agentes pelo regulador. Mesmo nessas circunstâncias, em que o comportamento dos agentes é controlado pelo próprio poder público, pode haver necessidade de aplicação das normas de defesa da concorrência, particularmente em situações de regulação parcial.¹⁸ Isso, entretanto, gera uma série de problemas de compatibilização entre sistemas legais e decisões de diferentes autoridades do poder público. Ademais, a autoridade reguladora pode ter um importante papel promotor da concorrência e preventivo de práticas anticompetitivas; afinal, concorrência e regulação não são a antítese um do outro, mas meios diferentes de buscar o mesmo objetivo – garantir condições de eficiência nos mercados.

Contudo, aqui nos ressentimos da falta de uma “cultura da concorrência”. A tradição no tratamento dos serviços públicos e outros tipos de regulação de atividades econômicas no Brasil tendeu não só a desconsiderar preocupações concorrenceis (por exemplo, a concessão com exclusividade sempre foi a regra) como também a proteger os participantes contra eventuais possibilidades de competição. Isso parece estar mudando.

Assim, por exemplo, nas mudanças institucionais recentes por que passaram (ou estão passando) importantes setores de infraestrutura – como telecomunicações e energia elétrica – novas agências reguladoras são criadas justamente para implementar um processo de introdução da concorrência; vemos uma série de medidas *preventivas* da concorrência serem tomadas pelas próprias agências reguladoras: normas que garantem o acesso a instalações essenciais, que disciplinam as condições de entrada nos mercados, que limitam as concentrações etc. Tais normas têm um efeito decisivo sobre a conformação de novos ambientes competitivos.

De qualquer modo, a relação entre políticas regulatórias e de defesa da concorrência também enseja temas jurídicos e econômicos para discussão, à medida que, assim como podem ser compatíveis, também podem apresentar dimensões divergentes que irão requerer critérios para solução de conflitos.

Notas

1. No sentido de que o Direito provê a *forma* para conteúdos econômicos. Muitos dos conceitos econômicos que serão utilizados já foram vistos em outros capítulos deste livro; a eles remeteremos o leitor, quando necessário, tratando aqui apenas das especificidades de sua aplicação ao âmbito antitruste.
2. É por isso que a existência de poder de mercado é condição para a aplicação da lei antitruste, como se verá adiante; mas isso não será ilícito quando resultar de processo *natural* decorrente da maior eficiência de um agente em relação a seus competidores (a lei brasileira é clara a esse respeito, cf. art. 36, § 1º).
3. Podem também ser objeto de controle as fusões *conglomeradas*, entre empresas de mercados distintos, mas relacionados estrategicamente – isto é, com proximidade de linha de produtos ou de localização.
4. Isso tem importantes implicações jurídicas, já que se trata de um critério para distinção entre ato lícito e ilícito. A questão será mais bem discutida adiante, quando tratarmos dos efeitos possíveis das práticas.
5. Diz-se que um interesse é *difuso* quando pertence a pessoas indeterminadas, ligadas por circunstâncias de fato (i.e., não se pode determinar quais indivíduos pertencem ao grupo). A concorrência é claramente estabelecida como tal na Lei n. 12.529/2011, que afirma que “*a coletividade é a titular dos bens jurídicos protegidos por esta lei*”.
6. Ver o Capítulo 2.
7. O lado da oferta é incorporado na análise num passo posterior, como veremos adiante.
8. Dependendo do caso também podem ser usados os percentuais de 5% ou de 15%. Nos EUA trabalha-se com a hipótese de um aumento de 5% a partir de um preço no nível competitivo, embora possa ser superior (em geral, até 10%).
9. Vemos, assim, que o conceito de mercado relevante, embora construído totalmente com técnica econômica, é, também, um conceito jurídico, dada a absoluta necessidade de definir previamente a proporção de aumento de preço que se considera abusiva.
10. Por exemplo, uma firma que detenha poder de mercado também pode baixar artificialmente seus preços por tempo suficiente para gerar prejuízos a concorrentes, a fim de excluí-los do mercado. Essa e outras condutas tipicamente anticompetitivas serão discutidas adiante.
11. O HH (índice de *Herfindahl-Hirschman*) é de longe o mais usado na área antitruste, em função de sua simplicidade e qualidades técnicas (é pouco sensível ao *market-share* de empresas de pequena participação, o que permite usar dados incompletos sem acarretar séria imprecisão).
12. Conforme visto também no Capítulo 5.
13. Para análise de barreiras à entrada, ver os Capítulos 7 e 12.
14. Os ganhos típicos de eficiência que podem estar associados a condutas e atos de concentração serão tratados na seção seguinte.
15. Definidas no Anexo da Resolução n. 20/99 do CADE.
16. Sobre custos de transação, ver Capítulo 13.

17. Sobre política industrial, ver o Capítulo 24.
18. Situação na qual se regulam apenas alguns aspectos ou apenas alguns segmentos do setor, deixando espaços livres para o comportamento estratégico dos regulados. Nesses casos, havendo graus de liberdade para as decisões dos agentes, também há espaço para o abuso de poder de mercado, que deve ser, então, reprimido.

Bibliografia

- Bruna, S. V. (1997) “O Poder Econômico e a Conceituação do Abuso em seu Exercício”. São Paulo, *Revista dos Tribunais*.
- Carlton, D. W.; Perloff, J. M. (2000) *Modern Industrial Organization*. Addison-Wesley. 3^a ed.
- Cohen, W. (1997) Per Se Illegality and Truncated Rule of Reason: the search for a foreshortened antitrust analysis. (<http://www.ftc.gov/>).
- Forgioni, P. (1998) *Fundamentos do Antitruste*. São Paulo, Revista dos Tribunais.
- Krattenmaker, T. G. (1996) “Per Se Violations in Antitrust Law: confusing offenses with defenses”, in Gavil, A.I. (org.), *An Antitrust Anthology*, Anderson Publishing.
- Possas, M. (1996) “Os Conceitos de Mercado Relevante e de Poder de Mercado no Âmbito da Defesa da Concorrência”, in *Revista do IBRAC*, n. 1, p. 82-102.
- Schuartz, L. F. (1998) Dogmática Jurídica e Lei 8.884/94, in *Cadernos de Direito Tributário e Finanças Públicas*, ano 6, n. 23, abril/junho.
- Sullivan, E. T.; Harrison, J. L. (1998) *Understanding Antitrust and its Economic Implications*. Nova York, Matthew Bender. 3^a ed.
- Viscusi, W. K.; Vernon, J. M.; Harrington Jr., J. E. (1995) *Economics of Regulation and Antitrust*. MIT Press, 2^a ed.

Sugestões de Leitura

Manuais sobre Economia e direito antitruste nos EUA.

- Hovenkamp, H. (1994) *Federal Antitrust Policy: The Law of Competition and its Practice*. West Publishing.
- Sullivan, E. T.; Harrison, J. L. (1998) *Understanding Antitrust and its Economic Implications*. Nova York, Matthew Bender. 3^a ed.
- Viscusi, W. K.; Vernon, J. M.; Harrington Jr., J. E. (1995) *Economics of Regulation and Antitrust*. MIT Press, 2^a ed.

Manuais de organização industrial, com capítulos voltados para antitruste.

- Carlton, D. W.; Perloff, J. M. (2000) *Modern Industrial Organization*. Addison-Wesley. 3^a ed.
- Scherer, F. M.; Ross, D. (1990) *Industrial Market Structure and Economic Performance*. Boston, Houghton Mifflin. 3^a ed.

Coletânea de textos sobre antitruste.

- Gavil, A. I. (org.) (1996). *An Antitrust Anthology*. Anderson Publishing.

Artigo

- Williamson, O. (1968). “Economies as an Antitrust Defense: the welfare tradeoffs”, in *American Economic Review*, v. 58, março, p. 18-36.

No Brasil

- Mattos, C. (coord.) *A revolução antitruste no Brasil – a teoria econômica aplicada a casos concretos*. São Paulo: Ed. Singular (Volume 1, de 2003; Volume 2, de 2008).

Estes dois volumes são coletâneas de textos de autores diversos – todos especialistas na área antitruste que desenvolvem trabalho acadêmico e de consultoria – que analisam casos já julgados pelas autoridades antitruste no Brasil. São excelentes exemplos de como a análise econômica é fundamental para as decisões jurídicas.

- Farina, E. (1994) “Desregulamentação e o Controle do Abuso de Poder Econômico”, in *Revista de Economia Política*, v. 14, n. 3, p. 78-93.

- Pondé, J.; Fagundes, J.; Possas, M. (1997) “Custos de Transação e Política de Defesa da Concorrência”, in *Revista de Economia Contemporânea*, n. 2, p. 115-135, jul.-dez.

- Possas, M. (1996). “Os Conceitos de Mercado Relevante e de Poder de Mercado no Âmbito da Defesa da Concorrência”, in *Revista do IBRAC*, n. 1, p. 82-102.

- Salgado, L. H. (1997) *A Economia Política da Ação Antitruste*. São Paulo, Ed. Singular.

Sobre direito antitruste no Brasil

- Bruna, S. V. (1997) “O Poder Econômico e a Conceituação do Abuso em seu Exercício”. São Paulo, *Revista dos Tribunais*.

- Ferraz, Jr., T. S. (1995) “Da Abusividade do Poder Econômico”, in *Revista de Direito Econômico*, n. 2123-2130, (out./dez.).

- Schuartz, L. F. (1998) “Dogmática Jurídica e Lei 8.884/94”, in *Cadernos de Direito Tributário e Finanças Públicas*, ano 6, n. 23, abril/junho.

Regulação Econômica

Helder Queiroz Pinto Jr e Ronaldo Fiani

23.1 Introdução

Define-se regulação como qualquer ação do governo no sentido de limitar a liberdade de escolha dos agentes econômicos. Dessa forma, quando um agente regulador (uma agência responsável por algum setor da economia, como eletricidade, telecomunicações etc.) fixa uma tarifa para um determinado serviço, está restringindo a liberdade que uma empresa tem de estabelecer o preço pela sua atividade.

Da definição acima, porém, é possível perceber que o campo da regulação é muito mais extenso do que apenas a regulação de preços (tarifária). Com efeito, ele se estende também à regulação de quantidades (por meio de limites mínimos de produção ou da limitação do número de empresas que podem atuar em determinado setor), regulação de qualidade (garantia da presença de determinadas características no serviço ou produto a ser ofertado), regulação de segurança no trabalho (quando a legislação obriga as empresas no setor de construção civil a equipar seus trabalhadores com determinados equipamentos de segurança, está da mesma forma limitando a liberdade que as empresas possuem de decidir qual equipamento elas devem fornecer aos seus trabalhadores), entre outros. Vamos nos preocupar aqui fundamentalmente com regulação de preço e, em alguma medida, com regulação de qualidade.

Nas abordagens mais convencionais não se espera que haja qualquer necessidade do governo interferir na liberdade de decisão econômica das empresas, uma vez que se supõe a economia em uma situação de concorrência perfeita, quando o mercado fornece estímulos eficientes por meio do mecanismo de preços para a alocação de recursos, como visto no Capítulo 1. Contudo, a teoria econômica, mesmo em suas vertentes mais liberais, reconhece que existem algumas situações em que o mercado não consegue levar a uma alocação eficiente de recursos, entendida como aquela em que os custos de oportunidade são minimizados. Uma dessas situações seria representada pela presença de externalidades. Como é sabido, caso haja externalidades o benefício ou custo sociais superam o benefício ou custo para a empresa que produz o serviço ou bem. Assim, há uma tendência a produzir em quantidade insuficiente o bem ou serviço no caso do benefício social ser superior ao benefício do produtor privado, ou há uma tendência a produzir em quantidade excessiva o bem ou serviço cujo custo social é mais elevado do que o custo do produtor privado.

Mas externalidades não são o único caso em que o mercado não fornece estímulos adequados para a alocação eficiente dos recursos. Uma outra circunstância é a de monopólio natural ou ainda o caso onde o monopólio natural se apresenta em conjunto com a presença de fortes externalidades típicas das indústrias de rede, conforme definida no Capítulo 16.

Neste capítulo vamos discutir, na primeira seção, o conceito do monopólio natural em duas circunstâncias distintas: monopólio natural com um produto e monopólio natural multiproduto, seguida de uma discussão onde se apresentam os fenômenos de monopólio natural nas indústrias de rede. Na segunda seção serão apresentadas as várias formas de regulação dos preços e,

finalmente, na terceira seção será apresentada uma visão histórica da regulação em outros países antes e depois do período de 1980, seguida de uma descrição sobre regulação no Brasil.

23.2 O Conceito do Monopólio Natural

23.2.1 MONOPÓLIO NATURAL COM UM PRODUTO

No caso de um monopólio natural com um único produto, os custos são menores se produzirmos uma dada quantidade x do produto em uma única firma do que em duas. Essa propriedade, conhecida como *subaditividade* da função de custo, foi apresentada no Capítulo 3. Algebricamente, se temos $x^* = x_1 + x_2$, onde x^* , x_1 , x_2 são quantidades de um mesmo produto, teremos um monopólio natural se graças à hipótese de subaditividade da função de custo for verdade que:

$$Ca(x^*) < C_b(x_1) + C_c(x_2) \quad (1)$$

onde C representa a função custo e a , b e c , três empresas distintas. Dessa forma, o que (1) mostra é que é mais barato produzir $x^* = x_1 + x_2$ apenas na firma A do que distribuir essa produção entre as firmas B e C . A questão que se coloca agora é: sob quais condições pode-se assegurar que a proposição (1) é verdadeira? A condição necessária e suficiente para que a proposição (1) seja verdadeira é que haja *economias de escala em toda a amplitude relevante de produção*. Como veremos adiante, economias de escala são condição suficiente e necessária somente no caso de um único produto.

Ocorre porém que, se há economias de escala em toda a amplitude relevante de produção, a firma se encontra na situação descrita no Gráfico 23.1 a seguir. Na situação do Gráfico 23.1, há o problema típico de um monopólio natural: se fosse estabelecido que a firma monopolista deve produzir com o preço ótimo do ponto de vista do bem-estar social, p_c , o qual é igual ao custo marginal (representado pela curva CMg) segue-se que à quantidade demandada a esse preço (q^* , definida pela curva D , o preço que permitiria à firma obter um lucro normal sobre o capital investido é p_e , que cobre os custos médios e é superior a p_c . Isso se deve ao fato de que como o custo médio (CMe) é declinante, dada a presença de economias de escala, o custo marginal é inferior ao médio. Assim, a solução ótima, de que o preço deve ser estabelecido igual ao custo marginal, não é mais possível.

As soluções para o preço de um monopolista natural nesse caso serão discutidas mais adiante. É necessário agora, contudo, considerar o caso de um monopólio natural multiproduto.

23.2.2 MONOPÓLIO NATURAL MULTIPRODUTO

A condição para o monopólio natural multiproduto continua sendo a mesma do monopólio natural com um único produto: subaditividade de custos. Ocorre que a subaditividade de custos em monopólios naturais multiprodutos para dois produtos X e Y exige que:

$$C_a(Q_x, Q_y) < C_b(Q_x, 0) + C_c(0, Q_y) \quad (2)$$

onde C representa a função custo, Q_x e Q_y as quantidades de X e Y e a , b e c , três empresas distintas.

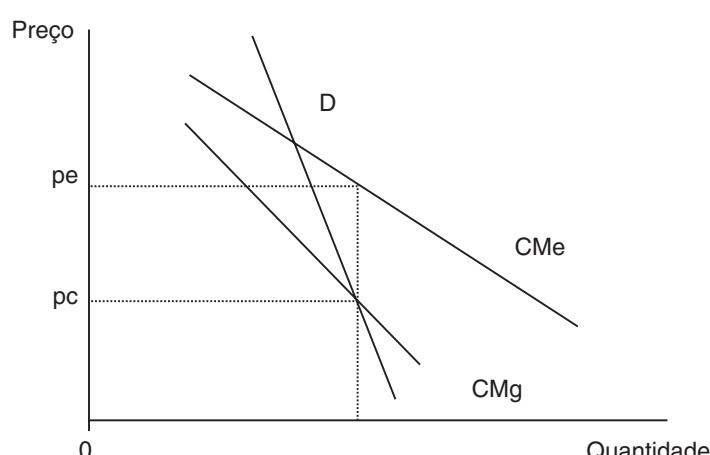


Gráfico 23.1 Monopólio natural.

Note-se que agora, na proposição (2), é mais barato produzir uma dada quantidade de cada produto (Q_x, Q_y) em uma única empresa A do que a mesma quantidade dos dois produtos, cada um em uma firma diferente. Assim, é mais barato fornecer acesso à Internet e ligações convencionais de voz pela mesma empresa, usando a mesma rede, do que por duas empresas diferentes.

O que é importante perceber é que, nesse caso, a presença de economias de escala não é condição nem necessária, nem suficiente, como antes era no caso de um monopolista com um produto, para a subaditividade de custos. Isto porque agora é importante a presença de *economias de escopo*. Caso não haja economias de escopo, *ainda que haja economias de escala em cada produto considerado isoladamente*, a proposição (2) não será verdadeira. Por outro lado, com significativas economias de escopo, mesmo que não haja economias de escala em cada produto, a proposição (2) será verdadeira.

23.2.3 INDÚSTRIAS DE REDE

Indústrias de rede, como definido no Capítulo 16, são um caso especial de monopólio natural. As indústrias de rede exploram a multiplicidade de relações transacionais entre os agentes econômicos situados em diferentes nós da rede, envolvendo um princípio de organização espacial e territorial. A partir desta definição, é fácil compreender por que as indústrias voltadas para a provisão de infraestrutura econômica (eletricidade, gás, telecomunicações, transportes, água e saneamento básico) são caracterizadas como indústrias de rede.

As indústrias de rede são marcadas por três elementos que contribuem para a formação de um modo de organização industrial particular. São eles:

1. A existência de externalidades.
2. A importância das economias de escala.
3. A articulação em torno da infraestrutura propriamente dita – a infraestrutura de base que comporta os serviços de transporte/transmissão do fluxo do produto – dos diferentes tipos de serviços finais e do serviço de coordenação da rede.

As indústrias de rede possuem como característica distintiva o fato de gerarem as *externalidades de rede*: o benefício de um usuário depende do número de usuários ligados à rede. O benefício de um consumidor que dispõe de uma linha telefônica depende diretamente do número de pessoas que estão conectadas, e com as quais ele pode se comunicar.

Como acontece em todos os casos de externalidades, trata-se de um problema de falha de mercado: como não há forma de um consumidor remunerar outro pelo benefício da adesão desse segundo a rede, há a possibilidade de não se alcançar uma expansão eficiente da rede. Essa necessidade de garantir um nível adequado de interconexão pode demandar a regulação do setor.

Além disso, o fato de que essas redes em geral apresentam economias de escala demanda regulação da entrada para se evitar duplicação ineficiente de infraestrutura, com a consequente elevação de custos e perda de bem-estar.

Por último, o fato de essas redes envolverem conexão direta com os consumidores gera um poder de mercado significativo para as empresas que administram essas redes, o que, associado com as vantagens derivadas de sua localização, que geram rendas extraordinárias, acabam por completar o conjunto de argumentos favoráveis à regulação.

As indústrias de redes eram tradicionalmente consideradas como monopólios naturais, com pesadas barreiras à entrada e à saída que justificaram a presença ativa do Estado na regulação dessas atividades.

23.3 Formas de Regulação dos Preços

Nesta seção iremos apresentar as várias formas de regulação que podem ser adotadas e algumas formas de regular os preços conhecidas.

23.3.1 REGULAÇÃO POR TAXA DE RETORNO

No processo de regulação por taxa de retorno, o regulador arbitra um vetor tarifário (tarifas para cada tipo de produto ou serviço da firma regulada), visando garantir para a firma regulada uma taxa de retorno considerada adequada ao prosseguimento de suas atividades. A proposição (3) a seguir descreve esse tipo de procedimento:

$$\sum p_i q_i = CV(q_1, q_2, \dots, q_n) + \pi(K), \quad (i = 1, \dots, n) \quad (3)$$

onde p_i representa a tarifa para um dado serviço, ou produto, produzido na quantidade q_i pela firma regulada, CV é o custo variável total resultante da produção dos vários bens e serviços nas quantidades q_i , π é a taxa de lucro bruto (incluindo depreciação) sobre o capital investido da firma K . Assim, o que a equação (3) diz é que a receita total, derivada dos vários produtos ou serviços da firma regulada, deve ser igual ao custo operacional, mais a depreciação do capital, e ainda deixar uma taxa de lucro normal sobre o capital investido.

A determinação prática da equação (3) envolve uma série de dificuldades. A primeira delas diz respeito à dimensão de K . Quais os ativos que devem ser computados no valor do capital da empresa? A inclusão indiscriminada de qualquer ativo incentivaria a empresa a investir excessivamente não só em ativos de pouca utilidade para suas tarefas produtivas, como até mesmo a manter ativos que, embora úteis do ponto de vista de sua atividade, estariam superdimensionados em relação à demanda, isto é, seriam mantidos ociosos. Cabe ao regulador admitir para o cálculo do capital da empresa regulada apenas os ativos *úteis e utilizados*.

O segundo problema diz respeito à taxa de retorno da empresa (π). Qual é a taxa de retorno adequada para a empresa regulada? A solução mais óbvia pareceria ser o custo de capital da empresa: desde que a empresa obtivesse uma taxa de retorno igual ao custo que possui para levantar seu capital, não haveria obstáculos à continuidade de suas atividades, assim como não haveria, em princípio, motivo para supor que a firma estaria obtendo lucros extraordinários. Ocorre, porém, que a determinação do custo de capital da empresa não é tarefa simples. Conforme visto no Capítulo 21, embora seja relativamente mais fácil determinar o custo de capital sobre os títulos emitidos pela empresa, e até mesmo em suas ações preferenciais, isso não ocorre com suas ações ordinárias, cuja taxa de retorno, exigida pelos acionistas, está sujeita a diferentes métodos de cálculo, que acabam por dar diferentes resultados.

O problema seguinte é a determinação do custo variável (CV). Isso exige não só conhecer a fundo a contabilidade da empresa regulada, mas, frequentemente, também seu processo produtivo. Essa necessidade requer do regulador um corpo técnico altamente especializado.

Por último, temos o problema da determinação da receita da empresa. De fato, embora a proposição (3) não explice, as próprias tarifas p_i acabam influenciando a quantidade demandada, por meio das funções de demanda inversas $q_i(p_1, p_2, \dots, p_n)$. Assim, ao estabelecer as tarifas, o regulador indiretamente estabelece as quantidades vendidas que, junto com aquelas tarifas, vão compor a receita total. E mais, indiretamente, dada a quantidade vendida, as tarifas também irão afetar o custo variável total CV . É necessário, então, conhecimento preciso da demanda, para evitar desequilíbrios entre receitas e custos.

Há ainda uma outra dificuldade com este método de regulação tarifária, que diz respeito às características gerais do método, e não à aplicação da proposição (3) em particular. O método de regulação por taxa de retorno, dada a complexidade de sua operação, é um método adequado apenas quando as condições de custos e demanda não variam de forma significativa em períodos relativamente curtos de tempo, isto é, quando custos e demanda são relativamente estáveis. Quando custos e demanda mudam com rapidez, devido a mudanças tecnológicas ou a variações nos hábitos ou na renda dos consumidores, este processo se torna muito lento para dar conta dessas mudanças. Mais grave ainda é que o método de regulação por taxa de retorno, na medida em que busca garantir uma taxa de lucro mínima, não fornece estímulos adequados ao aumento da eficiência, tão importante em fases de mudanças tecnológicas aceleradas.

Por fim, aponta-se ainda como outro problema da regulação por taxa de retorno o chamado *efeito Averch-Johnson*. Este efeito adviria do fato de que a agência reguladora, ao estabelecer a taxa de retorno adequada para a operação da firma regulada, que irá servir de base para o cálculo das tarifas, tende a estabelecer essa taxa de retorno acima do valor de mercado (pois se estabelecesse abaixo a firma regulada teria problemas para operar, por não estar obtendo o custo de oportunidade sobre o capital investido).

Ao estabelecer a taxa de retorno acima do valor de mercado, a agência torna o capital para a empresa regulada mais barato do que ele efetivamente é. Por exemplo, se a taxa de retorno for estabelecida em 10% ao ano para a empresa regulada, e a taxa de retorno média da economia é de 5% ao ano, o regulado recebe pelo capital investido o dobro de custo de levantar esse mesmo capital no mercado.

Assim, ele tenderá a substituir trabalho por capital (pois o capital para ele é subsidiado). Supondo-se a substitutibilidade perfeita entre esses dois fatores, a consequência será a de que a empresa regulada empregará uma quantidade excessiva de capital, a qual, se obtivesse por ela o mesmo retorno das demais atividades da economia, somente empregaria se a sua produção fosse bem maior do que é. Isto resulta em alocação ineficiente de recursos.

Contudo, esse mesmo *efeito Averch-Johnson* pode apresentar um lado positivo: maior intensidade de capital significa frequentemente não apenas melhor qualidade, como também tecnologias mais modernas. Em função desses problemas de caráter mais geral, desenvolveu-se um método alternativo, conhecido como preço-teto, que discutiremos a seguir.

23.3.2 PREÇO-TETO (PRICE CAP)

Uma das inovações em regulação econômica de mais rápida difusão foi o critério do preço-teto (*price cap*), na sua versão inglesa batizado como *Índice de Preço de Varejo Menos X* (IPV-X; em inglês *RPI-X – Retail Price Index Minus X*). Inicialmente aplicado para a *British Telecom*, em 1984, acabou por se expandir para outras empresas e setores na Inglaterra (*British Gas*, *British Airports Authority*, companhias regionais de fornecimento de água e na distribuição de energia elétrica), assim como para outros países (sendo o caso mais notório o do setor de telecomunicações nos EUA).

Basicamente, o sistema consiste em estabelecer um limite superior para a indústria regulada aumentar seus preços, limite este que pode ser estabelecido para cada preço individualmente ou para a média de preços dos serviços fornecidos pela indústria regulada. No caso do *IPV-X*, o teto do reajuste é estabelecido como sendo um índice geral de preços menos um valor *X* a título de aumento de produtividade. Esse teto de reajuste vale entre os períodos de revisão tarifária, quando a tarifa que serve como base do reajuste é reavaliada.

Assim, se fosse escolhido um índice de preço *I* para reajuste entre os períodos de revisão tarifária, e se fosse almejado um crescimento de produtividade de 2% ao ano, teríamos um fator *X* de 2%, e nosso índice de reajuste anual ficaria:

$$(I) - 2\%.$$

Portanto, em um dado ano, se o índice de preços *I* fosse de 5%, a firma regulada sujeita a essa regra de reajuste teria direito então a $5\% - 2\% = 3\%$ de aumento em sua tarifa.

Os defensores do *IPV-X* apontam as seguintes vantagens em relação aos métodos até então empregados de regulação, especialmente o com base na taxa de retorno:

1. É um método que atinge exclusivamente os serviços em que a empresa regulada atua como monopolista. Assim, supondo uma empresa multiproduto, que atue também em mercados competitivos, o *IPV-X* incidiria apenas naqueles mercados em que a empresa efetivamente atua como monopolista. Nos demais mercados, não há necessidade de regulação. Note a diferença em relação à regulação por taxa de retorno, em que é necessário estabelecer todas as tarifas, de modo que a receita global da empresa gere a taxa de retorno adequada.
2. Como toda redução de custos é apropriada pela empresa, espera-se que o *IPV-X* estimule a eficiência produtiva e promova a inovação. Assim, se os custos da empresa subiram apenas 1%, em função de ganhos de eficiência, ao passo que, como no exemplo anterior, o índice de preços *I* aumentou 5% e ela tem o direito a um reajuste de 3% nas tarifas, a empresa regulada se apropria integralmente do ganho resultante da diferença entre o reajuste a que tem direito, em função do aumento do índice de preços *I*, e o crescimento inferior de seus custos.
3. O custo do aparato da regulação econômica seria baixo, uma vez que este se resumiria ao cálculo de índices de preços, sem envolver o levantamento de dados contábeis a respeito da empresa regulada (quase sempre sujeitos a problemas de alocação de custos fixos e avaliação de valor de ativos), exceto no momento de revisão tarifária.
4. Dada a simplificação do processo regulatório, este se encontra menos sujeito ao risco de ser manipulado pela empresa regulada, com informações falsas sobre demanda e custos, ou seja, está menos sujeito ao que se conhece como “risco de captura”.

As vantagens citadas merecem tratamento em maior detalhe. No que diz respeito à vantagem de número (1), é importante salientar suas virtudes como instrumento de defesa da concorrência: a imposição de um preço-teto nos segmentos em que a empresa atua como monopolista permite evitar que ela, por meio de uma política de subsídios cruzados, subsidie os preços nos segmentos competitivos com os lucros extraordinários obtidos no segmento monopolista para praticar preços predatórios contra seus competidores.

A vantagem de número (2) é o resultado da substituição de um esquema de incentivos pouco poderoso como é a regulação por taxa mínima de retorno. Nesta última, tendo a empresa uma taxa de retorno mínima garantida, os incentivos à redução de custos são pouco eficientes. É verdade que o modelo de regulação anteriormente adotado tentava corrigir este problema por meio do conceito de ativos “úteis e utilizados”, que tentava distinguir do capital contabilizado pelo agente regulado aqueles elementos que realmente contribuíam de forma efetiva para a produção. Assim, na escolha dos ativos úteis, seria possível privilegiar aqueles itens mais modernos. Obviamente, dada a assimetria de informações entre a empresa regulada e a agência

reguladora, esta distinção entre os ativos produtivos envolve problemas conceituais significativos, além de substanciais custos de transação para as partes envolvidas.

Porém o *IPV-X* também apresenta problemas, o mais sério deles dizendo respeito ao investimento. Uma forma de aumentar a taxa de lucro quando há um teto de receita é reduzir a base de capital sobre a qual esta taxa é calculada. O sistema de preço-teto tem, portanto, como resultado indesejável promover o subinvestimento, com efeitos negativos não apenas sobre o crescimento da oferta da empresa regulada, mas também sobre a qualidade dos serviços prestados ou sobre os estímulos para a inovação. Para minimizar este problema é exigido da agência reguladora um esforço adicional para controlar os planos de investimento e a qualidade dos serviços prestados pela firma regulada, o que invalida em grande medida a vantagem (3) descrita anteriormente.

Outro elemento que também contribui por colocar em dúvida a vantagem (3) e acaba por afetar igualmente a vantagem (4) é o fato de que, na prática, o valor de *X* não pode ser estabelecido sem levar em consideração elementos tais como taxa de retorno da firma regulada, valor de seus ativos, custo do capital, taxas esperadas de crescimento da produtividade e da demanda etc., o que torna o processo de regulação tão complexo e vulnerável às assimetrias de informação quanto o método convencional com base no estabelecimento de uma taxa mínima de retorno para a firma regulada.

Ainda assim, permanece apenas a vantagem operacional (1) e, salvo as restrições anteriores e *ceteris paribus*, espera-se que um teto para o preço atue favoravelmente principalmente nos setores sujeitos a processos de rápida inovação tecnológica, onde os incentivos à modernização e eficiência são mais fortes.

23.3.3 REGRA DO COMPONENTE DE PREÇO EFICIENTE

A regra do componente de preço eficiente (*efficient component-pricing rule*, ou simplesmente RCPE) é citada na literatura para problemas de interconexão, ou quando existe a necessidade de uma empresa utilizar a infraestrutura de uma rival. Suponha que, como acontece, uma empresa de telefonia local seja proprietária de uma prestadora de serviço de acesso à Internet. Qual a tarifa que esta empresa de telefonia local deve cobrar de outras prestadoras de acesso à Internet, que também utilizam sua rede? A RCPE procura estabelecer um critério para essa tarifa, e casos desse gênero.

A Figura 23.1 representa uma situação em que só existe uma forma de ligar X a W, isto é, passando por *c*, ao passo que há duas maneiras de ligar W a Y: ou por meio de *a*, ou por meio de *b*. O leitor pode imaginar que o trecho XW é a linha telefônica que possui, sendo *c* a empresa de telefonia local, enquanto Y seria o acesso à Internet que tanto pode ser feito pelo provedor *a*, de propriedade da mesma empresa *c*, quanto pelo provedor independente *b*. O ponto a ser destacado aqui é que *b* precisa do acesso a *c* se quiser fornecer seus serviços. Por isso, esse tipo de problema em regulação é conhecido como problema do *preço de acesso*.

Vamos supor agora que o custo de utilizar *c* seja constante e igual a R\$ 4. Por sua vez, o custo de *a* (o custo de acessar a Internet) do provedor de propriedade da empresa de telefonia local seja R\$ 3. E que a receita total recebida pela empresa de telefonia fixa e seu provedor por um acesso a Internet seja R\$ 10 (um preço que, a princípio, vamos supor equivalente a uma situação de concorrência). Qual deve ser o preço cobrado pelo acesso a *c*?

Uma primeira hipótese seria cobrar um preço igual a R\$ 4, remunerando a empresa de telefonia local apenas pelo custo de utilizar sua rede. Na verdade, esse tipo de solução não é adequado porque quando fornece acesso à sua rede, a empresa de telefonia local não tem apenas o ônus do custo de utilização da rede, mas também o ônus do lucro que perde, pois deixa de fornecer, pela empresa provedora que possui, o acesso à Internet.

Assim, o mais adequado, na verdade, é remunerar a firma de telefonia local pelo custo de *c* *mais os lucros sacrificados por deixar de vender um acesso à rede pelo provedor que possui*. O lucro quando ela própria vende o acesso à rede é: R\$10 – R\$4 – R\$3=R\$3, ou seja, a receita de um acesso, menos o custo da rede de telefonia local, menos o custo do provedor da

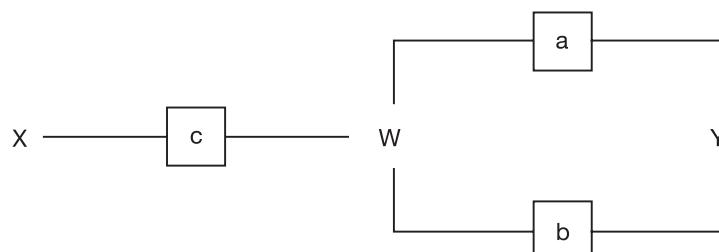


Figura 23.1 Diferentes formas de acesso.

empresa de telefonia fixa. Dessa maneira, o valor que deve ser cobrado do provedor concorrente é: R\$4 + R\$3 = R\$7. Essa é uma solução eficiente?

Para avaliar se essa é uma solução eficiente, imaginemos que o provedor independente é mais eficiente, de tal forma que seu custo de acesso é menor. Vamos supor R\$2. Seu custo total seria então: R\$2 + R\$7 = R\$9. Note que, dado o preço final de R\$10, o provedor independente tem um lucro extraordinário de R\$1. Isto lhe permite cobrar um preço ligeiramente inferior ao da firma de telefonia local com seu provedor, digamos, R\$9,50, e capturar o mercado. A firma de telefonia local fecharia o seu provedor e passaria apenas a alugar o serviço de acesso à rede de telefonia fixa para o provedor independente. Apesar disso, ela estaria tão bem quanto antes, pois continuaria obtendo o mesmo lucro de R\$3.

Vamos supor agora que o provedor independente é menos eficiente. Vamos supor que seu custo é de R\$4. Nesse caso, seu custo total é R\$4 + R\$7 = R\$11. Como o preço é R\$10, o provedor independente, por ser menos eficiente, acaba por ter prejuízo e se retira do mercado.

É nesse sentido, então, que essa é uma regra do preço eficiente: uma vez aplicada, ela garante que apenas as firmas mais eficientes permanecerão no mercado. Mas é importante notar que isso apenas é verdade se o preço final tiver sido estabelecido em bases competitivas. Suponhamos que o preço tenha sido estabelecido em R\$11, um valor acima daquele que seria realizado em condições de concorrência. O provedor independente seria obrigado a pagar R\$4 + R\$4 = R\$8. Se após a sua entrada o preço se reduzisse (em virtude da concorrência) para R\$10, o provedor independente penalizado com o pagamento de R\$8 apenas permaneceria no mercado se seus custos fossem iguais ou inferiores a 2/3 dos custos da empresa verticalmente integrada, supondo que o provedor da empresa integrada tenha custo de R\$3 (isto é, se fossem iguais a R\$2).

Assim, a eficiência dessa regra depende significativamente da fixação do preço final em níveis concorrenenciais.

23.3.4 REGULAÇÃO DE MONOPÓLIO MULTIPRODUTO: A REGRa DE RAMSEY

A regra de Ramsey é uma solução para monopólios multiprodutos, onde os preços dos produtos ou serviços são estabelecidos para minimizar as perdas dos consumidores, resultantes da necessidade do monopolista de cobrir seus custos totais e, portanto, dada a situação de monopólio natural, não poder igualar os preços aos custos marginais. Trata-se de uma solução em preços lineares (as despesas do consumidor variam na mesma proporção da quantidade consumida). Um monopólio multiproduto seria o caso de uma estrada de ferro que transportasse diferentes tipos de cargas (*containers* e granel ou grãos), ou ainda carga e passageiros.

Uma derivação um pouco mais formal da regra de Ramsey, ou regra de Ramsey-Boiteux, como também é conhecida, é apresentada a seguir. Considere uma firma monopolista produzindo n produtos nas quantidades $q^* = (q_1, q_2, \dots, q_n)$. Sua função custo é dada por $C(q^*)$. As demandas para os n produtos são representadas por funções de demanda inversa do tipo: $p_i(q_i)$, $i = 1, 2, \dots, n$. Supõe-se que as funções de custo e de demanda são diferenciáveis, assim como as elasticidades cruzadas das demandas são nulas.

Como de costume, o excedente líquido do consumidor, V_i , para o i -ésimo produto seria dado por:

$$V_i = \int_0^{q_i} p_i(q_i) dq_i - p_i(q_i) q_i \quad (4)$$

O leitor deve reconhecer que a integral em (4) representa a área total sob a curva de demanda, ao passo que o segundo termo representa a despesa total do consumidor com o bem i . Assim, (4) nada mais é do que o excedente líquido do consumidor.

Se derivarmos (4) em função de q_i , o resultado dará a variação no excedente do consumidor em função da quantidade:

$$\frac{\partial V_i}{\partial q_i} = p_i(q_i) - p_i'(q_i) - p_i'(q_i) q_i = -p_i'(q_i) q_i \quad (5)$$

Por outro lado, o lucro total π da firma monopolista é dado por:

$$\pi = \sum_{i=1}^n p_i(q_i) q_i - C(q^*) \quad (6)$$

Se considerarmos como medida do bem-estar da sociedade (W) a soma dos excedentes líquidos dos consumidores para todos os produtos com o lucro da firma, teremos:

$$W = \sum V_i + \pi \quad (7)$$

Temos agora de encontrar uma solução que maximize o bem-estar da sociedade dada a restrição de que o lucro do monopolista seja igual a zero (lucro econômico nulo), o que no caso do monopólio significa preço igual a custo médio. Assim trabalhamos com a restrição de que $\pi^* = 0$. Esse tipo de problema se resolve da forma habitual, formando o lagrangeano, L, a partir do multiplicador de Lagrange λ :

$$\max_q L = W + \lambda(\pi - \pi^*) \quad (8)$$

A condição de primeira ordem para um dado produto q_i nos dá:

$$\frac{\partial L}{\partial q_i} = p_i(q_i) - CMg_i + \lambda[p_i(q_i) + q_i p'_i(q_i) - CMg_i] = 0 \quad (9)$$

Note-se que o custo marginal do produto i , CMg_i , é calculado supondo-se constante a quantidade produzida dos demais produtos.

Rearrumando os termos de (9) obtemos:

$$[p_i(q_i) - CMg_i](1 + \lambda) = -q_i p'_i(q_i) \lambda \quad (9.a)$$

Dividindo ambos os lados de (9.a) por $p_i(q_i)$ e $(1 + \lambda)$, e lembrando que a elasticidade-preço é, por definição, $(dq/dp) \times (p/q)$, chega-se a:

$$\frac{p_i(q_i) - CMg_i}{p_i(q_i)} = \frac{-\lambda}{1 + \lambda} \cdot \frac{1}{e_i} \quad (10)$$

Onde e_i é a elasticidade-preço da demanda por i .

O preço calculado em (10) é o preço de Ramsey para o produto i . Nessa equação vemos que o preço do produto i é inversamente proporcional à sua elasticidade: quanto maior ela for, menor será o preço. A intuição para esse resultado é que quanto maior a elasticidade-preço, maior a redução na quantidade consumida de um bem em função de uma elevação no seu preço e, consequentemente, maior a perda de bem-estar dos consumidores. Logo, para minimizar essas perdas, sujeitas à necessidade da empresa de cobrir os seus custos (lucro zero), é necessário que os preços dos vários produtos sejam estabelecidos na proporção inversa de suas elasticidades.

23.3.5 TARIFA EM DUAS PARTES

Outra regra para o cálculo do requerimento de receita da empresa para a tarifação pode ser feita pela utilização de preços não lineares.¹

A tarifa em duas partes (*two-part tariff*) é calculada por uma taxa fixa para o rendimento total, que é independente da venda do produto ou serviço, e um preço por unidade do serviço efetivamente usado. Os resultados de Ramsey para o monopólio multiproduto podem ser utilizados neste caso. Para isto, basta considerar que, além do produto que está sendo transportado pela rede, existe um outro serviço: o acesso à rede.

Assim, esse tipo de tarifa é particularmente útil na definição de um regime tarifário para os segmentos de transporte e/ou transmissão nas indústrias de redes. Ela pode ser definida da seguinte maneira:

$$T(q) = A + P_i$$

Ou seja:

Tarifa em duas partes = componente fixo (acesso) + preço por unidade consumida.

É importante observar que a solução de *Ramsey* permite uma boa ilustração do problema de tarifa em duas partes. No cálculo da tarifa em duas partes, considerando os elementos que a compõem como produtos distintos, o produto pelo qual a demanda é mais inelástica (acesso) deve ser remunerado com um *mark-up* maior sobre o custo marginal ($A > CMg$). Por sua vez, o componente de utilização corrente deverá ter seus preços iguais ao custo marginal, gerando, portanto, uma combinação ótima.

Porém, se o preço de acesso é muito alto, alguns consumidores podem considerar que o benefício de preço ao custo marginal para utilização corrente é menor do que o custo de acesso, levando-os a não adquirirem o direito de acesso. Isso faz com que o

preço de acesso se torne elástico. Neste caso, a solução de *Ramsey* implica o aumento em ambos os componentes dos preços, produzindo como resultado: $A > CMg$ e $P > CMg$.

QUADRO 23.1 TARIFA EM DUAS PARTES: MÉTODOS PARA ESTIMAÇÃO DE CUSTOS

Este tipo de regime vem sendo progressivamente adotado nas indústrias de rede, em particular para as linhas de transmissão de eletricidade e para a rede de transporte de gás (gasodutos). Vários exemplos concretos ilustram o uso da tarifa de transporte nas indústrias de rede.

Quatro métodos distintos foram utilizados na classificação de custos para a determinação deste tipo de tarifa de transporte:

Straight Fixed-Variable (SFV): quanto maior o fator de carga² da empresa que utiliza as linhas de transporte e transmissão (denominadas *carriers* – carregadoras), menores serão seus custos unitários para o serviço de transporte. O custo de transportar volumes incrementais é baixo pelo fato de se ter uma pequena parcela de custos fixos na movimentação adicional.³ À medida que as empresas carregadoras podem vender volumes adicionais no mercado *spot* a baixo custo, elas têm incentivo a operar com maior fator de carga possível, o que os leva a maximizar a capacidade de transporte.

Seabord: garantia que os consumidores “interrupíveis”⁴ fora dos períodos de pico,⁵ também arcassem com parte dos custos fixos, incidindo uma maior porcentagem sobre a parcela de movimentação.

United: neste critério, além dos custos variáveis, 75% dos custos fixos seriam repassados à parcela de movimentação. Sua utilização teve como objetivo reverter a situação declinante de utilização da capacidade de transporte instalada, ocorrida nos Estados Unidos em 1973, em consequência do baixo suprimento de gás no mercado. Neste caso, os carregadores com baixo fator de carga não seriam tão penalizados como no caso de SFV.

Modified Fixed – Variable (MFV): vigorou nos Estados Unidos entre os anos de 1983 e 1992, quando a FERC (órgão regulador federal de energia) o substituiu pelo método SFV. Os custos fixos relacionados ao Imposto de Renda e à remuneração do capital, juntamente com os custos variáveis, seriam repassados à parcela de movimentação. Este critério visava à adequada utilização dos gasodutos no mercado a preços competitivos em relação aos combustíveis alternativos.

23.4 A Regulação na Prática

23.4.1 ANTECEDENTES DE REGULAÇÃO ECONÔMICA: MODELOS BÁSICOS NORTE-AMERICANO E EUROPEU

No início do século XX, com o surgimento das primeiras empresas de telefonia, água, eletricidade e gás, o processo efetivo de constituição de redes beneficiou-se da adoção de inovações tecnológicas, que permitiram às empresas trocarem de escala geográfica de operação: inicialmente organizadas em torno do fornecimento local dos serviços, as inovações tecnológicas associadas à transmissão/transporte do fluxo de serviços permitiram a conexão de consumidores mais distantes, favorecendo a otimização da capacidade instalada e o aproveitamento dos ganhos de escala. Como resultado, monopólios territoriais e integração vertical das diferentes etapas da cadeia produtiva tornaram-se o modo básico de organização dessa indústria.

A importância da infraestrutura para o desenvolvimento econômico, pelo fato de ser portadora de externalidades positivas, associada às características básicas do modo de organização mencionado anteriormente, justificou a necessidade da intervenção estatal nessas indústrias.

Apesar dos traços comuns desse modo de organização, a intervenção estatal e os instrumentos de regulação utilizados variaram muito nos países industrializados. Podemos, de forma esquemática, destacar duas formas de regulação das indústrias de rede até o início dos anos 1980. A primeira, desenvolvida essencialmente nos EUA, tem como objetivo básico a defesa do interesse público. Ela é centrada no controle dos monopólios privados das indústrias de rede, tendo como característica básica o arcabouço jurídico-institucional norte-americano, apoiado nas tradições e com forte primado da jurisprudência para arbitragens dos conflitos entre diferentes agentes.

Nessas circunstâncias e em função da dimensão continental, um grande poder regulatório foi accordado às comissões públicas estaduais a partir dos anos 1930.⁶ Vale observar que na tradição da legislação americana de regulação da concorrência baseada nas leis antitruste, o monopólio é a exceção; a concorrência é a norma.

Nos EUA, por exemplo, a necessidade de supervisionar a concentração (política antitruste) é que fornece o fundamento básico da intervenção estatal para garantir o interesse público nos setores onde o abuso de posições dominantes demanda uma vigilância particular.⁷ Esta abordagem do serviço público está fortemente vinculada à concepção de liberalismo político, que por meio da instrumentalização da esfera do direito público busca proteger os consumidores do poder de monopólio da operadora das indústrias de rede.

Para entender melhor este argumento é indispensável examinar alguns dos atributos da concepção de serviço público de infraestrutura. É possível identificar a presença de dois atributos nos chamados serviços públicos:

1. É essencial para a maior parte da população e para os diferentes setores da atividade econômica;
2. O mercado é incapaz de fornecê-lo na quantidade demandada e com a mesma qualidade, pois existe assimetria na relação entre usuários e produtores; ou seja, em presença de falhas de mercado, os produtores podem se beneficiar de poder de monopólio.

Nos EUA, esse encaminhamento institucional não objetiva, como em vários países europeus, atribuir uma legitimidade ao Estado para que este seja o agente promotor – como operador, financiador e gestor do desenvolvimento das indústrias de rede. Do ponto de vista econômico, essas indústrias comportavam as características de rendimentos crescentes e externalidades que justificavam as situações de monopólio natural e a estrutura verticalizada das empresas operadoras. Porém, é indispensável notar que a grande parte das empresas de utilidade pública (*utilities*) americanas é privada.

A definição de interesse público, nos EUA, repousa sobre a ideia de arbitragem de conflitos, isto é, o interesse público é resultante do processo de confrontação de interesses individuais. Por essa razão, as instituições de direito público ocupam um lugar de destaque para a regulação de diferentes indústrias de rede, estruturando-se no nível estadual, por meio de diferentes comissões públicas estaduais, que funcionam com elevado grau de autonomia com relação à administração federal. Os instrumentos de regulação, em particular os mecanismos tarifários, garantiam, por um lado, a operação das indústrias de rede articuladas em torno da estrutura monopolista; por outro, eles limitavam a extensão geográfica das suas atividades.

Na Europa, a partir da década de 1940, o Estado assumiu as responsabilidades, na maioria dos países europeus, de planejamento, operação, coordenação e gestão da infraestrutura econômica. Do ponto de vista jurídico-institucional, a atenção foi voltada para a definição das propriedades de serviço público (continuidade, neutralidade e capacidade de adaptação).

Do ponto de vista da estrutura industrial, o modelo das indústrias de redes europeias caracterizou-se pela constituição de grandes empresas estatais dispostas de monopólios territoriais e integradas verticalmente.

O principal modo de organização das indústrias de redes dos países em desenvolvimento foi fundamentado em duas razões principais. A primeira está ligada à percepção da infraestrutura econômica como vetor das transformações estruturais necessárias (passagem de uma economia de base agrícola para uma economia de base industrial) e, consequentemente, como fonte de externalidades positivas para a economia como um todo. A segunda razão está relacionada com os fluxos de financiamento necessários para a expansão acelerada das redes. Dada a limitação da capacidade de investimento, os créditos provenientes dos bancos internacionais e os aportes do Banco Mundial foram de extrema importância nesse período. Conclui-se que as condições de financiamento induziram os países em desenvolvimento a adotarem o modo de organização industrial já difundido em outros países.

23.4.2 AS REFORMAS DOS ANOS 1980 E A NOVA FASE DA REGULAÇÃO

Os traços gerais das reformas, muito embora sejam subordinados ao ambiente institucional de cada país, podem ser ilustrados pela implementação total ou parcial das seguintes medidas:

1. Desverticalização dos diferentes segmentos de atividade da cadeia produtiva dos serviços de infraestrutura.
2. Introdução da concorrência em diferentes segmentos de atividade das indústrias de rede.
3. Abertura do acesso de terceiros às redes.
4. Estabelecimento de novas formas contratuais.
5. Privatização das empresas públicas.
6. Implementação de novos mecanismos de regulação e criação de novos órgãos reguladores.

O processo de reestruturação das indústrias de rede vem modificando profundamente o modo de organização industrial e as formas de regulação que governavam as atividades econômicas e os serviços prestados pelas empresas operadoras. Esse processo engendra novas oportunidades de negócio e uma redefinição das estratégias tradicionais das empresas, além de reservar ao Estado novas atribuições especialmente em matéria de regulação e de defesa da concorrência.

Esse processo, muitas vezes apelidado de forma imprecisa de desregulamentação, reúne um conjunto de medidas de política econômica, visando à introdução da concorrência, à desintegração vertical e, em muitos casos, à privatização das empresas públicas.

Note-se que com a entrada de novas empresas operadoras, desloca-se sensivelmente o foco central da regulação, antes fundado na supervisão de empresas estatais. A nova forma de regulação setorial implica uma maior complexidade institucional e uma nova forma de intervenção do Estado.

Ao contrário do que o termo sugere, a desregulamentação não é sinônimo de ausência de regulamentação. Ao contrário, o processo de desmantelamento das estruturas de mercado verticalizadas e monopolistas tem reservado um papel central para as tarefas de regulação. Desse modo, a desregulamentação deve ser entendida como uma política econômica, conduzida pelos governos, visando à remoção de barreiras à entrada. Dessa forma, objetiva-se introduzir pressões competitivas nas indústrias de infraestrutura com o intuito de incrementar a eficiência econômica.

Sob esta ótica, vários países criaram novas agências de regulação econômica responsáveis pela transição para novos modos de organização, contemplando maior grau de concorrência em determinados segmentos de atividade.

Observa-se que, na maior parte das atividades em rede, surge uma separação cada vez mais clara entre, de um lado, a rede de infraestrutura de base que transmite sinais, energia, água etc. e, de outro, a rede teleinformática de comando que presta o serviço de coordenação do uso da infraestrutura. As funções de comercialização e de coordenação tornam-se muito mais importantes.

É precisamente esta dualidade funcional entre a infraestrutura de base e a rede de comando que representa o motor do movimento de abertura de diversas indústrias de rede (livre acesso ou *open access*), por meio do incremento e da diversificação da oferta de serviços, desestabilizando o modelo tradicional de organização industrial das indústrias de rede.

Não obstante a diversidade de novos modos de organização industrial que emergem em diferentes países e a importância relativa das funções de transmissão/transporte e de distribuição,⁸ o traço comum que orienta a reforma das indústrias de rede é sustentado por dois princípios básicos: por um lado, o serviço de comando das redes pode ser oferecido por um mercado aberto a diferentes empresas, com regras de planejamento flexíveis para que a segurança da infraestrutura de base seja garantida e as necessidades dos diferentes clientes possam ser atendidas; por outro lado, a gestão das infraestruturas de base permaneceria em regime de monopólio natural, baseada na justificativa da existência de economias de escala importantes, com possibilidade, contudo, de introdução de mecanismos quase concorrenenciais. A articulação desse conjunto de tarefas implica a construção de um novo marco regulatório e a criação de instrumentos de coordenação diferentes daqueles utilizados no modelo de planejamento centralizado.

É interessante observar que do ponto de vista da abordagem de Economia Industrial nasce uma problemática nova: no passado, a firma gozava de condições monopolísticas em todos os segmentos das indústrias de rede; atualmente, as possibilidades oferecidas pelas inovações tecnológicas e as pressões competitivas que favorecem a entrada de novos atores fazem com que diversas indústrias de rede passem a combinar atividades/serviços concorrentes e não concorrentes, abrindo campo, por um lado, para a separação, ao menos do ponto de vista contábil e da gestão, do controle/propriedade da infraestrutura de base e dos demais serviços. Por outro lado, reforça-se a necessidade de coordenação da utilização da infraestrutura de base, já que a eficiência das indústrias de rede permanece vinculada à importância da interconexão que, na prática, materializa economias de coordenação.⁹

As firmas presentes nessas indústrias podem efetivamente competir com outros concorrentes em um determinado segmento das indústrias de rede e, ao mesmo tempo, dispor de um monopólio em um segmento diferente. Em outras palavras: atualmente, no âmbito de uma indústria de rede, diferentes estruturas de mercado podem ser identificadas, comportando um número variável de empresas operadoras. A própria tarefa de regulação assume também outra dimensão: sem abrir mão da necessidade de coordenação sistêmica, a tarefa de regulação das indústrias de rede, hoje, requer instrumentos que permitam regular, simultaneamente, segmentos dessas indústrias onde algum grau de concorrência é possível de ser observado e outros que preservam características de monopólio natural.

O papel dos novos órgãos de regulação setorial torna-se muito mais complexo. No passado, regular a indústria era sinônimo de regular tanto a conduta como o desempenho de uma única empresa monopolista e verticalizada. A complexidade de mudanças simultâneas no modo de organização industrial, das formas de regulação e nos direitos de propriedade numa indústria de rede requer a construção de um novo ambiente institucional, no qual as ações dos órgãos reguladores reúnem um leque de missões de regulação, entre as quais se destacam:

1. Supervisionar o poder de mercado dos operadores e evitar práticas anticompetitivas.
2. Organizar a entrada de novos operadores e promover a competição.
3. Zelar pela implementação de um novo modo de organização industrial.
4. Defender e interpretar as regras, arbitrando os eventuais conflitos entre atores.
5. Complementar o processo de regulamentação.
6. Estimular a eficiência e a inovação, estimulando a repartição dos ganhos de produtividade registrados na indústria com os consumidores.
7. Zelar pelas condições de operação coordenada das redes.
8. Assegurar o cumprimento das missões de serviço público.

23.4.3 A REGULAÇÃO NO BRASIL

No Brasil, as experiências internacionais influenciaram a agenda de reestruturação dos setores de infraestrutura, sustentada pelos mesmos princípios. Aqui, como já mencionamos, o argumento central estava articulado com a falta de capacidade de financiamento das empresas estatais. Assim, caberia aos capitais privados dos novos operadores a missão de recuperar o nível de investimentos em infraestrutura, eliminando os gargalos de crescimento dos demais setores da economia.

Atualmente o escopo de atuação das agências reguladoras se expandiu. Após a criação das agências de eletricidade (ANEEL), telecomunicações (ANATEL) e de petróleo e gás (ANP), foram criadas novas agências reguladoras que regulam outros setores da economia, não necessariamente de infraestrutura, tais como: a Agência Nacional de Saúde Suplementar (ANS), criada pela Lei n. 9.961, de 28 de janeiro de 2000 para regulamentar o setor de saúde suplementar – que compreende os planos de saúde; a Agência Nacional do Cinema (ANCINE), criada em 6 de setembro de 2001 pela Medida Provisória 2.228-1 visando ao fomento, à regulação e à fiscalização do mercado do cinema e do audiovisual no país; e a Agência Nacional de Aviação Civil (ANAC), criada pela Lei n. 11.182, de 27 de setembro de 2005 para regular e fiscalizar as atividades de aviação civil e de infraestrutura aeronáutica e aeroportuária.

23.4.3.1 Regulação da Indústria Elétrica

O processo de reestruturação da indústria elétrica brasileira tem como marco fundamental a Lei de Concessões (fevereiro/1995) que estabelece regras para tarifas, licitação, autorização e permissão de concessão, bem como determina as cláusulas presentes nos contratos de concessão. A Lei de Concessões, que abrange todos os serviços públicos, representa um marco importante para o setor elétrico. Ela viabiliza a abertura industrial ao capital privado, sinalizando o setor com o fim da integração vertical dividindo-o nos segmentos de geração, transmissão e distribuição.

Um dos dispositivos institucionais mais importantes no processo de reestruturação da indústria elétrica consiste na promulgação da Lei n. 9.427 (dezembro/96). Essa lei institui a ANEEL (Agência Nacional de Energia Elétrica), substituindo o DNAEE (Departamento Nacional de Águas e Energia Elétrica) e disciplinando o regime das concessões de serviços públicos de energia elétrica, onde são estabelecidos princípios básicos para o processo de descentralização e delegação de atividades para os estados da federação.

A ANEEL, como agência federal de regulação do setor, tem como função a regulação e a fiscalização das atividades de geração, transmissão, distribuição e comercialização de energia elétrica. Na qualidade de autarquia, a agência tem autonomia financeira e decisória, sendo dirigida por um diretor-geral e mais quatro diretores, sendo o processo decisório caracterizada pela forma de colegiado. Os mandatos dos diretores são de quatro anos, sendo autorizada uma recondução. Dentre as várias atribuições da ANEEL, destacam-se:

1. Resolver conflitos e divergências entre concessionárias, permissionárias, autorizadas, produtores independentes e autoprodutores, bem como entre esses agentes e seus consumidores.
2. Assegurar a entrada de novos agentes no mercado. Nesse sentido, por meio da homologação das regras do mercado atacadista de energia, pretende-se estimular a concorrência na geração de eletricidade.
3. Definição e aplicação dos novos princípios de regulação de tarifas, substituindo a tarifação ao custo de serviço pelo regime preço-teto (segmento de distribuição).
4. Defesa das condições de concorrência, estabelecendo regras que limitam o poder de mercado das empresas concessionárias.
5. Definição dos padrões técnicos e normativos de qualidade e desempenho das empresas.

Além disso, foi estabelecida, à imagem do que foi observado nos Estados Unidos, a garantia do livre acesso à rede básica de transmissão por meio da criação do ONS (Operador Nacional do Sistema Elétrico), ao qual será delegado o direito de coordenar o uso das redes de transmissão. O ONS é uma entidade de direito privado, sem fins lucrativos, estruturada sob a forma de associação civil, da qual participam geradores, empresas de transmissão e distribuição, consumidores livres, comercializadores, importadores e exportadores de eletricidade.

As principais tarefas do ONS são: (1) o despacho das centrais, visando ao menor custo para o sistema; (2) a garantia da qualidade do suprimento elétrico na rede de transmissão; e (3) a garantia de acesso equitativo à rede de transmissão para todos os agentes do mercado elétrico. Neste contexto, a ANEEL estabeleceria as condições gerais do acesso ao sistema de transmissão e de distribuição e regularia as tarifas correspondentes.

Em 2004, o sistema regulador do setor de energia elétrica passaria por profundas transformações, sendo ampliado com a criação da EPE (Empresa de Pesquisa Energética, criada pela Lei n. 10.847, de 15 de Março de 2004), responsável pelo planejamento do setor de longo prazo; o CMSE (Comitê de Monitoramento do Setor Elétrico, criado pela Lei n. 10.848, de

2004), responsável pela avaliação da segurança do suprimento de energia elétrica; e a CCEE (Câmara de Comercialização de Energia Elétrica, criada pela Lei n. 10.848, de 15 de março de 2004), que assumiu as tarefas relativas à comercialização de energia elétrica no sistema interligado, dando continuidade às funções do antigo MAE (Mercado Atacadista de Energia).

23.4.3.2 Regulação de Telecomunicações

A regulação da indústria de telecomunicações no Brasil está a cargo da ANATEL, Agência Nacional de Telecomunicações. Sua criação se deu por meio da Lei Geral de Telecomunicações (Lei n. 9.472 de 16/07/1997) no seu artigo 8º, tendo sido posteriormente regulamentada pelo Decreto 2.238/97.

Trata-se de uma autarquia especial ligada ao Ministério das Telecomunicações, tendo como característica a sua autonomia e independência: os cinco conselheiros-diretores da ANATEL são indicados pelo presidente da República e aprovados pelo Senado, com mandatos fixos de cinco anos, sendo impedida a sua recondução. A perda do mandato só poderá ocorrer nos casos de renúncia, condenação judicial (quando transitada em julgado) ou ainda por meio de processo administrativo disciplinar.

A autonomia orçamentária é estabelecida por meio de receitas próprias a partir do Fundo de Fiscalização dos Serviços de Telecomunicações (*Fistel*), que é composto a partir das taxas cobradas nas outorgas das licitações, de taxas de fiscalização, multas, entre outros.

A Anatel regula os serviços de telecomunicações, que incluem a telefonia fixa comutada e celular. Na telefonia fixa comutada atuam as empresas concessionárias e suas concorrentes (empresas-espelho), distribuídas em três áreas no país para telefonia regional e uma área (correspondendo à totalidade do território nacional) para ligações nacionais e internacionais. Já a telefonia celular foi dividida em dez áreas, onde atuam empresas na banda A (empresas estatais que foram privatizadas) e banda B (empresas que foram admitidas no setor para concorrer com as empresas estatais privatizadas). No momento a ANATEL vem encaminhando o processo de licitação da chamada quarta geração de telefonia móvel pessoal (4G), dez vezes mais rápida que a banda larga móvel disponível em 2012. Planeja-se que todos os municípios com mais de 100 mil habitantes terão cobertura 4G até 31 de dezembro de 2016. Planeja-se também que as sedes e subsedes da Copa do Mundo tenham o serviço 4G até 31 de dezembro de 2014.

As tarifas de telefonia também estão sujeitas ao regime de preço-teto, que consiste na aplicação do IGP-DI, menos um redutor, para a cesta de preços da firma. Cumpre ressaltar que a Anatel pode adotar a liberdade tarifária caso constate haver competição efetiva entre as empresas do setor. Mesmo depois de suspenso, contudo, o controle tarifário pode voltar a ser adotado, caso a Anatel identifique práticas anticompetitivas no setor.

O regime de preço-teto no caso brasileiro se caracteriza por um fator de transferência atualmente baseado na produtividade total dos fatores. A produtividade total dos fatores examina – como o próprio nome indica – não apenas a produtividade do fator trabalho, mas a totalidade dos fatores de produção empregados pela empresa. Essa produtividade será calculada de acordo com as projeções de ganhos de produtividade da empresa, substituindo o valor arbitrado de 1% que foi aplicado entre 2000 e 2005.

23.4.3.3 Regulação do Petróleo e Gás

A Lei n. 9.478, de 06/08/1997, que institui o Conselho Nacional de Política Energética-CNPE, como órgão vinculado à Presidência da República e à Agência Nacional do Petróleo-ANP, como autarquia especial vinculada ao Ministério de Minas e Energia regulamenta a atuação de todas as empresas operadoras, inclusive a da estatal Petrobras, no mercado brasileiro de petróleo e de gás natural, atendendo ao dispositivo da reforma constitucional, de 1995, que retirou da estatal a competência de executora do monopólio.

A ANP também tem autonomia financeira e decisória, sendo dirigida por um diretor-geral e mais quatro diretores, com mandatos de quatro anos, sendo autorizada uma recondução. Assim como as demais agências, a ANP acumula as responsabilidades de poder concedente e de regulação.

Por se tratar da regulação de uma indústria de recursos naturais não renováveis, os objetivos de introdução de pressões competitivas e atrair capitais privados foram compatibilizados com a manutenção da titularidade da União dos direitos de propriedade das reservas de petróleo e de gás natural.

Nesse sentido, a lei ratificou os direitos de propriedade da Petrobras das áreas de produção, bem como das refinarias e dos seus equipamentos de transporte dutoviário/marítimo e do seu complexo portuário/armazenagem.

Cabe observar que as atribuições da ANP compreendem um conjunto de segmentos de atividade econômica, comportando estruturas de mercado e características técnico-econômicas bastante distintas. A indústria de petróleo e derivados é tradicionalmente analisada a partir de três etapas produtivas: *upstream* (fases de exploração e produção), *middlestream* (transporte e refino) e *downstream* (distribuição e revenda). A possibilidade de importação de petróleo bruto e derivados e as características descentralizadas de distribuição e revenda fazem com que a indústria de petróleo e derivados não seja caracterizada como uma

indústria de rede. Entretanto, a indústria de gás é uma ilustração típica das indústrias de rede, especialmente pelas características de monopólio natural dos segmentos de transporte e de distribuição.

Essas características revelam que a ANP tem um papel regulador que é bastante peculiar e que, por ora, não encontra paralelo em outros países.

A partir da Lei n. 9.478/1997, as principais atribuições de regulação sob responsabilidade da ANP são:

1. Implementar a política nacional de petróleo e gás natural.
2. Fiscalizar diretamente ou mediante convênios as atividades integrantes da indústria do petróleo.
3. Promoção de licitações de blocos petrolíferos, a fim de consolidar o processo de entrada de novas empresas.
4. Estrutura e controle dos *royalties* e demais participações governamentais.
5. Estabelecer os critérios para a movimentação e comercialização do petróleo, derivados e gás natural.
6. Estabelecer a regulação do acesso aos dutos.
7. Fazer cumprir as boas práticas de conservação e uso racional do petróleo e seus derivados e do gás natural e de preservação do meio ambiente.
8. Manter base de dados e difusão das informações geológicas das bacias sedimentares brasileiras.
9. Garantir o suprimento de derivados em todo o território nacional.
10. *Proteger o interesse dos consumidores quanto a preço, qualidade e oferta dos produtos.*

Como nos casos das agências de telecomunicações e eletricidade, uma tarefa fundamental nos primeiros anos de vida das agências de regulação é a edição de normas e portarias regulamentando aspectos gerais estabelecidos nas leis das agências.

Por fim, importa salientar que a ANP, ao contrário da Aneel e Anatel, não exerce a regulação de preços.

Na indústria do gás natural, as tarifas de transporte são negociadas entre a empresa proprietária dos ativos de transporte e as demais empresas que queiram se valer do acesso aos gasodutos. Se não houver entendimento quanto aos termos contratuais, a ANP é solicitada para atuar como mediadora do conflito, sendo sua atribuição apresentar uma solução a ser adotada pelas partes.¹⁰ Além disso, a regulação das tarifas de distribuição, segmento que apresenta características de monopólio natural, também não é competência da ANP. Isto é explicado pelo fato de que, de acordo com a Constituição, o poder concedente e de regulação da distribuição de gás é dos estados da Federação. Assim, as tarefas de regulação são atribuição das também recém-criadas agências reguladoras estaduais.

Notas

1. Vale recordar que um preço linear é aquele em que a receita iguala-se ao produto entre o volume vendido e o seu preço ($R = \text{preço} \times \text{volume transportado}$).
2. Fator de carga corresponde ao percentual da capacidade instalada total que é utilizada.
3. A noção de movimentação se refere aos volumes das mercadorias (commodities) efetivamente transportados.
4. Consumidores interruptíveis são aqueles que assinaram contratos de compra de um produto ou serviço com uma cláusula que facilita ao transportador a interrupção do fornecimento do produto. Note que, neste caso, o comprador do produto deve buscar diversificar as fontes de suprimento.
5. Período de pico ou de ponta corresponde aos horários de máxima utilização da capacidade instalada. A demanda por serviços de infraestrutura possui essa componente de sazonalidade. Em determinados momentos do dia, a demanda é particularmente mais elevada. Por exemplo, os horários de pico do fornecimento de eletricidade acontecem, em geral, à noite entre 19h e 22h, quando uma grande quantidade de equipamentos elétricos é utilizada ao mesmo tempo.
6. Por meio da promulgação do PUHCA (*Public Utility Holding Act Commission*).
7. Vale recordar que, apesar da enorme influência política econômica que exerciam, o presidente Roosevelt colocou os “barões” das estradas de ferro, da telefonia e da energia elétrica sob o controle das leis antitruste.
8. Diversas indústrias de rede são articuladas em torno da infraestrutura de base de distribuição local, como o serviço postal. Outras indústrias de rede não dispõem de uma infraestrutura de base de distribuição, como, por exemplo, as redes ferroviárias e aeroviárias.
9. As economias de coordenação eram viabilizadas no antigo modelo pela concentração em torno das estruturas públicas; elas podem ser identificadas em três níveis distintos: operacional (regras de acesso à rede, encaminhamento dos fluxos, gestão das externalidades); comercial (regras tarifárias e contratuais, eventual perequação) e estratégica (segurança e expansão da rede a longo prazo).
10. Em caso de não concordância de uma das partes, é evidentemente assegurado o direito de apelação ao Poder Judiciário.

Bibliografia

- Armstrong, M.; Cowan, S.; Vickers, J. (1994) *Regulatory Reform: Economic Analysis and British Experience*. Cambridge, Mass., The MIT Press.
- Baumol, W. J.; Sidak, J. G. (1995) *Transmission Pricing and Stranded Costs in the Electric Power Industry*. Washington: The AEI Press.
- Berg, S.; Tschirhart, J. (1988) *Natural Monopoly Regulation: Principles and Practice*. Cambridge: Cambridge University Press.
- Kansky, K. J. (1989) “Measures of Networks Structure”, *Flux*, número especial.
- Newbery, D. M. (2000) *Privatization, Restructuring, and Regulation of Network Utilities*. Cambridge, Mass., The MIT Press.
- Viscusi, W. K.; Vernon, J. M.; Harrington Jr., J. E. (1995) *Economics of Regulation and Antitrust*. Cambridge, Mass., The MIT Press.

Política Industrial

João Carlos Ferraz, Germano Mendes de Paula e David Kupfer

24.1 Introdução

A participação do Estado na promoção das atividades produtivas de um país é uma questão plena de controvérsias entre os economistas. O tema é defendido ou atacado a partir de suas diferentes bases teórico-analíticas. São três as correntes principais: ortodoxa, desenvolvimentista e evolucionista. A perspectiva ortodoxa coloca em questão as fronteiras de atuação do Estado e do mercado na promoção de atividades econômicas. A ótica desenvolvimentista prioriza o poder econômico e produtivo das nações no contexto internacional. Na perspectiva evolucionista o foco está na competência dos agentes econômicos em promoverem inovações que transformem o sistema produtivo. As diferentes correntes estão associadas a diferentes quadros analíticos e posições normativas. Diante deste quadro, por muitas vezes, o debate escapa aos limites da discussão econômica, para enveredar por questões tipicamente ideológicas, dificultando a compreensão adequada do tema. Na medida do possível, tentaremos evitar essa via do debate.

Esvaziada de juízos de valores, o objetivo mais tradicional pretendido pela política industrial é a promoção da atividade produtiva, na direção de estágios de desenvolvimento superiores aos preexistentes em um determinado espaço nacional. Do ponto de vista conceitual, política industrial deve ser entendida como *o conjunto de incentivos e regulações associadas a ações públicas, que podem afetar a alocação inter e intraindustrial de recursos, influenciando a estrutura produtiva e patrimonial, a conduta e o desempenho dos agentes econômicos em um determinando espaço nacional*.

Este capítulo está organizado da seguinte forma. Na próxima seção, discutiremos a relação entre Estado e mercado. Nas três seções seguintes, abordaremos os principais tópicos da literatura sobre política industrial, mediante o exame das três correntes mencionadas: que associam política industrial às falhas de mercado, ao desenvolvimento de nações e à evolução das competências de agentes econômicos. Na sexta seção iremos nos concentrar na relação entre política industrial e outras políticas de Estado, tais como a política macroeconômica, de comércio exterior, regulação de infraestruturas e de ciência e tecnologia. A seguir, a experiência brasileira será revista, assim como a prática recente dos países da Organização para Cooperação e Desenvolvimento Econômico (OCDE).¹ À luz da discussão anterior, ao final do capítulo, serão resumidas as principais questões relativas à intervenção do Estado na promoção do desenvolvimento industrial.

24.2 As Relações entre Estado e Mercado

Em uma retrospectiva histórica, as ideias sobre a relação entre Estado e mercado oscilam como um pêndulo. O primeiro pensamento sistemático vem dos mercantilistas, que advogavam a intervenção do Estado no comércio e na indústria. Com Adam Smith, surgiu e se difundiu a noção da mão invisível do mercado, que aloca recursos de modo eficiente, culminando com o liberalismo e o padrão ouro do fim do século passado. A partir da primeira metade do século XX, o pêndulo se moveu novamente: da hegemonia do mercado em direção a uma forte intervenção do Estado, para atingir metas nacionais e sob distintos formatos ideológicos, como o keynesianismo, sob o *New Deal*, nos Estados Unidos, ou o marxismo-leninismo na União Soviética. Após a Segunda Guerra Mundial, as diferenças do papel do Estado entre as diversas nações cresceram: algumas seguiram trajetórias mais intervencionistas ao passo que outras, caminhos mais liberalizantes.

A partir dos anos 1980, o liberalismo voltou a ganhar espaço entre as nações, a partir das políticas praticadas pelos governos Thatcher, na Inglaterra, e Reagan, nos Estados Unidos. Os estados nacionais passam a concentrar atenção na manutenção da estabilidade macroeconômica e na liberalização dos mercados, em detrimento da política industrial.

Esse movimento pendular reflete-se nos países em desenvolvimento ao longo de todo o século XX, primeiro, com a primazia das políticas de substituição de importações e, mais recentemente, com a aplicação do rol de políticas associadas ao “Consenso de Washington”, expressão cunhada para resumir um decálogo de medidas de política econômica consensuais entre as agências norte-americanas e agências internacionais localizadas na capital norte-americana (ver o Quadro 24.1).

Na atualidade, estamos observando esforços teóricos para ancorar as análises sobre as relações entre Estado e mercado em bases mais substantivas. Os avanços nesse campo tentam incorporar, principalmente, três constatações. Primeiro, o sucesso do desenvolvimento asiático na década de 1980 forçou os economistas a incluírem as instituições públicas nos modelos de desenvolvimento econômico. Segundo, a importância do progresso técnico e do aprendizado como fontes de eficiência, já demonstrada por economistas consagrados, foi finalmente formalizada com as contribuições da chamada “Nova Teoria do Crescimento” sobre retornos crescentes de escala associados ao avanço do progresso técnico, que implicam espaços economicamente justificáveis para a ação do Estado. Terceiro, os economistas passaram a discutir as restrições do funcionamento do mercado sob as quais se dá a atuação do Estado: a existência de racionalidade limitada, de informação imperfeita e de interesses múltiplos, que implicam a necessidade de maiores esforços para entender e melhorar a qualidade da ação pública.

Esses avanços indicam que a polarização do debate entre Estado e mercado não tem sentido, fortalecendo a visão de cientistas políticos que enfatizam a responsabilidade histórica dos Estados no processo de transformação econômica das sociedades. Isto porque cada Estado está imerso em um conjunto concreto de relações sociais, que define espaços, canais e modo de negociação entre administração pública e agentes econômicos de forma dinâmica e associada ao estágio de desenvolvimento de suas nações. Assim, nas palavras de Peter Evans, “(...) o debate estéril sobre o ‘quanto’ o Estado intervém deve ser substituído por argumentos sobre os diferentes tipos de envolvimento do Estado na sociedade e seus efeitos.... o envolvimento do

QUADRO 24.1 “O CONSENSO DE WASHINGTON”

1. *Disciplina fiscal*, caracterizada por um expressivo superávit primário e por déficits operacionais de não mais de 2% do PIB.
2. *Priorização dos gastos públicos*, mediante o seu redirecionamento de áreas politicamente sensíveis, que recebem mais recursos do que seria economicamente justificável – como a manutenção da máquina administrativa, a defesa ou os gastos com subsídios indiscriminados –, para setores com maior retorno econômico e/ou com potencial para melhorar a distribuição de renda, tais como saúde, educação e infraestrutura.
3. *Reforma fiscal*, baseada na ampliação da base tributária e na redução de alíquotas marginais consideradas excessivamente elevadas.
4. *Liberalização do financiamento*, com vistas à determinação da taxa de juros pelo mercado, ou como objetivo intermediário mais realista e até mesmo mais conveniente no curto e médio prazos – para evitar taxas muito elevadas –, procurando o fim de juros privilegiados e visando à obtenção de uma taxa de juros real positiva e moderada.
5. *Unificação da taxa de câmbio* em níveis competitivos, com o fim de eliminar sistemas de taxas de câmbio múltiplos e assegurar o rápido crescimento das exportações.
6. *Liberalização comercial*, por meio da substituição de restrições quantitativas por tarifas de importação que, por sua vez, deveriam ser reduzidas para um nível baixo, “... de 10% ou, no máximo, perto de 20%”.
7. *Abolição de barreiras ao investimento externo direto*.
8. *Privatização*.
9. *Desregulamentação*.
10. *Garantia do direito de propriedade*, por meio da melhoria do sistema judiciário.

Estado é dado. A questão adequada não se refere a ‘quanto’ o Estado intervém, mas sim ‘que tipo’ de intervenção ocorre e quais suas consequências”.

Este tipo de visão recupera as contribuições de autores clássicos, como F. List e A. Gerschenkron, que buscaram demonstrar, com base em análises históricas dos processos de industrialização de países europeus, a importância do papel do Estado como agente do desenvolvimento industrial. Na Europa do século XIX, o tipo de intervenção do Estado variou em grande medida de acordo com o nível de desenvolvimento das forças produtivas das economias nacionais. A disseminação da ideologia do livre comércio à época atendia aos interesses da economia britânica, então dominante, e o Estado era visto como protetor dos interesses dos capitalistas nacionais, em outros países europeus. Para List, a posse da capacidade de aumentar a riqueza nacional era mais importante do que a riqueza em si mesma, como defendida por Adam Smith. Em condições de atraso relativo, cabia aos estados nacionais – em seu caso, o alemão –, empreender esforços para aumentar a quantidade e qualidade do capital humano e para acessar a melhor tecnologia disponível, o que demandaria políticas industriais ativas e de longo prazo, por meio da combinação de mecanismos de proteção e incentivo, especialmente tarifas de importação e crédito de longo prazo em condições favoráveis. Para esse autor, liberalização ou proteção comercial, por exemplo, são somente meios para determinados fins: o desenvolvimento do poder produtivo. Qual política seria eficaz e adequada dependeria, em grande medida, do estágio de desenvolvimento de uma nação em relação aos líderes internacionais.

Em síntese, a discussão sobre as relações entre Estado e mercado surge quando se questiona a efetividade do sistema privado em alocar recursos econômicos escassos para os fins desejados de uma sociedade, em momentos específicos do tempo. Essa discussão também deve ser pautada pelas necessidades de desenvolvimento de uma nação, associando-se desenvolvimento à geração e distribuição de riqueza, eficiência e conhecimento. A estratégia e o modo de intervenção do Estado devem ser coerentes com o estágio de desenvolvimento dos agentes produtivos de um país, sempre tendo como referência os avanços da fronteira internacional.

24.3 Política Industrial pela Ótica das Falhas de Mercado

Como visto no Capítulo 1, na visão neoclássica o mercado competitivo é o alocador eficiente dos recursos. Supondo informação perfeita e reversibilidade sem ônus das decisões, os agentes racionais realizam escolhas que maximizam não apenas o seu bem-estar individual, mas também o coletivo. A livre mobilidade dos fatores e o atomismo dos agentes levam a que o mecanismo de demanda e oferta determine preços de equilíbrio ótimos do ponto de vista social, significando que qualquer aumento extra na utilidade de um grupo específico somente pode ser atingido às expensas da utilidade de outro grupo. Portanto, se todos os mercados forem competitivos, a política industrial é não somente desnecessária como, principalmente, indesejável. Analogamente, a intervenção governamental seria necessária se, e somente se, o mecanismo de preços não capturasse todos os benefícios e custos de oportunidade associados à produção e ao consumo de bens. Nessas situações ocorrem o que os economistas neoclássicos denominam falhas de mercado.

Pela ótica das falhas de mercado, a intervenção pública é passiva, mirando somente aquelas condições de funcionamento subótimo da alocação via mercados competitivos. Somente nesses casos o custo de intervenção pública poderia ser inferior aos benefícios potenciais a serem produzidos.

Não é simples a passagem de uma justificativa teórica baseada em uma assertiva genérica – as falhas de mercado justificam a intervenção – para sua aferição em um contexto específico e, consequentemente, a derivação concreta de implicações de política. O primeiro passo seria a classificação, sob algum tipo de critério geral suficientemente robusto, do que constitui uma falha de mercado. Embora a própria literatura econômica sobre o tema não seja muito clara, é possível enumerar cinco tipos principais de falhas de mercado:

1. Estruturas de mercado ou condutas não competitivas (oligopólios e monopólios)
2. Externalidades
3. Bens públicos
4. Direitos de propriedade comuns
5. Diferenças entre as taxas de preferências intertemporais sociais e privadas

Estruturas oligopolizadas ou monopolizadas decorrem, muitas vezes, da existência de economias de escala, isto é, os custos unitários de produção reduzem-se à medida que a produção se eleva. Se a empresa se depara com custos médios de longo prazo decrescentes ou, pelo menos, não crescentes, a situação do monopólio natural poderá ter lugar (veja a exposição desse tema no Capítulo 3). Para a sociedade, fica estabelecido um dilema: a minimização de custos irá significar uma estrutura monopólica, com a empresa dotada de poder de mercado para fixar preços além do nível competitivo.

Uma vez que os mercados, na atualidade, são preponderantemente oligopolizados, não chega a ser surpreendente que a política industrial seja um tema tão relevante. Nesse caso, os principais mecanismos da política industrial são a política de concorrência com vistas à redução do poder de mercado das grandes empresas, tanto do ponto de vista de possíveis condutas anticompetitivas, quanto da concentração do mercado por intermédio de fusões e aquisições. Enfim, de acordo com a perspectiva de falhas de mercado, as políticas de regulação de concorrência buscam evitar que o exercício do poder de mercado resulte em perda de bem-estar para o consumidor, tendo como parâmetro desejável o desempenho de mercados competitivos, conforme explicado no Capítulo 22.

Externalidades ocorrem quando as decisões de um agente econômico influenciam, positivamente ou negativamente, outros agentes. O exemplo clássico de externalidade negativa diz respeito à poluição. Considerem-se duas empresas, sendo uma refinaria de produtos petrolíferos e uma fábrica de pescado. A refinaria, obviamente, produz não somente derivados de petróleo, mas também poluição marinha. Assim, a poluição é um custo para a fábrica de pescado, mas não para a refinaria, que define sua produção em função dos custos e preços dos seus produtos. A consequência desta externalidade é a ineficiência alocativa, pois haverá excesso de produção de derivados de petróleo e subprodução de pescado. Analogamente, uma externalidade positiva (ou sinergia), como um distrito industrial, não seria captada na contabilidade individual de cada agente, levando a um subinvestimento em cada uma das atividades.

Reforçando o argumento, a política industrial sob a ótica das falhas de mercado requer medidas corretivas, complementares à atuação de agentes privados racionais. São em número de três as formas de correção do problema introduzido pela externalidade: (1) fusão entre as empresas envolvidas, porque, nesse caso, a externalidade negativa ou positiva passaria a ser contabilizada, respectivamente, como um custo ou como uma receita, pelo tomador de decisão; (2) criação de um imposto ou subsídio para corrigir os preços dos produtos e aproximá-los de seus custos de oportunidade ou (3) atribuição de direitos de propriedade de forma que crie um mercado para externalidade. Naturalmente, mecanismos de política industrial para correção de falhas de mercado devido a externalidades devem ser aplicados cuidadosamente para não gerar nenhuma outra distorção derivada.

Outra falha de mercado ocorre no *fornecimento de bens públicos*. Bens públicos possuem duas características principais: a não exclusividade e a não rivalidade. A primeira significa que a propriedade do bem não pode ser atribuída como um direito exclusivo de um único agente econômico. Portanto, ele não pode ser possuído, comprado ou vendido. Por exemplo, o benefício da iluminação de uma avenida não pode ser privilégio apenas de um indivíduo ou de um subconjunto dos usuários dessa avenida, pois ninguém pode ser excluído desse benefício. A não rivalidade quer dizer que a agregação de novos consumidores não altera o custo dos bens consumidos. No exemplo citado, mesmo com o incremento de população, esse custo não aumentaria.

Em função dessas características, o fornecimento de bens públicos dá margem a um tipo de comportamento oportunista conhecido como o problema do “carona” (*free rider*): a possibilidade de usar sem pagar. O mercado tende a falhar no fornecimento de bens públicos pois os incentivos para o ofertante são insuficientes. Como solução, esses bens devem ser ofertados diretamente pelo Estado ou por meio de concessões.

Bens de propriedade comum (ou difusa) também tendem a estimular comportamentos incompatíveis com o ótimo social. Isto é, quando direitos de propriedade não são apropriáveis individualmente, podem existir poucos incentivos para conservar ou melhorar propriedades comuns, como, por exemplo, uma área de produção de pescado. Um determinado agente, mesmo sabendo que certas práticas são predatórias e poderão comprometer a viabilidade futura da atividade, não seria incentivado a mudar sua conduta pelo receio de que os demais agentes não farão o mesmo. Novamente, caberá ao Estado industrial disciplinar a taxa de exploração dessas reservas.

Finalmente, o mercado irá falhar se existem *diferenças quanto às preferências intersetoriais sociais e privadas*. Elas existem quando agentes privados e a sociedade divergem quanto à preferência entre consumo corrente e futuro de um bem ou serviço. A pesquisa básica é um bom exemplo dessa situação. Como apresenta uma taxa de retorno muito demorada, investimentos nesse campo, embora de interesse para a sociedade como um todo, tendem a ser subalocados pelos agentes privados. Neste caso, a intervenção do Estado se justificaria.

Em resumo, se os mercados falham em prover o uso eficiente de recursos existiriam, em tese, espaços para a intervenção pública, na forma de políticas industriais. Assim, segundo uma perspectiva ortodoxa, a política industrial teria finalidade essencialmente corretiva, no sentido de amenizar os impactos negativos das falhas de mercado.

A teoria neoclássica, no entanto, não parece ser uma base apropriada para prescrições de política. Isto porque, devido aos pressupostos de informação e rationalidades perfeitas típicos da visão ortodoxa, a possibilidade de falhas de mercado é restrita a poucas situações. Porém, se os agentes econômicos falham em assimilar e interpretar quantidades ilimitadas de informação, isto é, se prevalecem a rationalidade limitada e a informação imperfeita, geradoras de incertezas, a percepção de futuro varia de modo considerável. Isto cria oportunidades para que certos agentes tenham vantagens sobre outros ou que alguns possam cometer equívocos. Incertezas, em suma, são geradoras de falhas de mercado e podem demandar intervenções públicas, uma impossibilidade dentro do campo conceitual neoclássico.

24.4 Política Industrial pela Ótica Desenvolvimentista

A corrente desenvolvimentista deve ser apreciada levando-se em consideração três conjuntos de condicionantes: o contexto específico, isto é, as características intrínsecas da nação onde a intervenção está sendo empreendida; o tempo histórico, a saber, em que estágio de desenvolvimento de um país estão sendo implementadas políticas ativas; e, finalmente, o contexto internacional. Ou seja, em diferentes momentos do tempo, o contexto internacional ajuda a definir se políticas intervencionistas são aceitas ou rechaçadas, facilitando ou dificultando a ação de um Estado específico. Por exemplo, o período que parte da Grande Depressão, do final dos anos 1920, até os anos 1980, é rico em experiências nacionais desenvolvimentistas bem-sucedidas.

Essa corrente comprehende a atuação do Estado como um elemento ativo, e não apenas corretivo, como sugerido pelos defensores da abordagem de falhas de mercado. O conceito “desenvolvimentista” (*developmental*) caracteriza um Estado que estabelece como princípio de legitimidade a capacidade de promover e sustentar desenvolvimento, entendido como a combinação de taxas de crescimento econômico altas e sustentadas e mudança estrutural no sistema produtivo.

A argumentação da intervenção parte da noção do *apoio e proteção à indústria nascente*. Inicialmente defendido por John Stuart Mill e levado ao extremo por Friederich List, a premissa principal é de que a indústria que está se constituindo em um determinado país terá, muito provavelmente, custos mais elevados que os vigentes em países onde a atividade já se encontra estabelecida. Portanto, a não intervenção governamental tenderia a perpetuar uma determinada divisão internacional do trabalho. Logo, a ótica desenvolvimentista guarda estreita relação com o estágio de desenvolvimento das forças produtivas de um determinado país e é mais “intervencionista”, quanto mais tardio for o processo de industrialização.

A proteção à indústria nascente possui duas premissas básicas. Primeiro, os custos de produção, apesar de relativamente altos inicialmente, tendem a se reduzir de modo significativo, à medida que os fabricantes se aproveitam das economias de aprendizagem. Os custos unitários de produção cairiam em função da experiência industrial acumulada ao longo do tempo. Com isso, o hiato de eficiência entre as nações tenderia a se reduzir no tempo. Segundo, tendo em vista a diminuição da desvantagem inicial, a proteção deveria ser temporária. Caso contrário, os consumidores domésticos estariam sempre adquirindo bens a preços acima dos preços praticados no mercado internacional, reduzindo não apenas o seu bem-estar, mas também perpetuando a ineficiência produtiva. Na agenda de temas que devem integrar a definição de uma estratégia industrial ativa por parte do Estado destacam-se a relevância da manufatura como setor estratégico, capaz de dinamizar a economia como um todo; a importância central do capital intelectual e do *learning by doing* e a necessidade de importar as tecnologias estrangeiras mais avançadas.

Para A. Gerschenkron, os países avançados representam o horizonte de possibilidades para o futuro para os países em desenvolvimento. Portanto, não se trata de inovar, mas de copiar um mapa produtivo existente e fazer uma nação crescer a taxas superiores aos líderes internacionais, em um processo de emparelhamento (*catching-up*). A existência desse horizonte produtivo facilitaria a identificação das metas a serem atingidas e minimizaria as falhas das administrações públicas.

Em um contexto desenvolvimentista, todos os instrumentos de política econômica (cambial, monetária e fiscal; de comércio exterior, de regulação da concorrência e da propriedade etc.) são colocados a serviço do objetivo industrializante. Em um Estado desenvolvimentista, as políticas beneficiam o setor privado, a empresa nacional, e priorizam o crescimento, a rivalidade e a produtividade, tendo como referência a melhor prática internacional, para emular experiências. O Estado tem legitimidade política e um corpo técnico capacitado, dotado dos instrumentos de intervenção necessários. Nesse sentido, o *Estado-desenvolvimento lidera o mercado*, pois as autoridades tomam iniciativas sobre que produtos e tecnologias deveriam ser encorajados, mobilizando os necessários instrumentos de incentivo e regulação.

Mas, em geral, subsistem diferenças em termos de taxa de crescimento, estrutura e composição da indústria, instrumentos efetivamente utilizados em cada país e ideologia por trás das políticas empreendidas em cada nação, principalmente quando o processo de industrialização envolve um período longo do tempo. Por exemplo, a Coreia do Sul diferencia-se da maioria dos países de industrialização tardia pela disciplina que o Estado exerceu sobre o setor privado. O sucesso desse país dependeu, em grande medida, da capacidade de o Estado criar distorções de preço para guiar as atividades econômicas na direção do investimento. Assim, o Estado sul-coreano privilegiou a grande empresa nacional, proibiu as atividades de empresas estrangeiras, facilitou a importação de tecnologia e, notavelmente, subsidiou o investimento de um grupo selecionado de setores e firmas, para estimular um conjunto de indústrias específicas. Em contrapartida, foram impostos padrões de desempenho extremamente rígidos, em geral associados ao sucesso exportador. Ao mesmo tempo, no Estado sul-coreano, assim como em Taiwan, as políticas focalizaram alvos em andamento, isto é, privilegiaram, sucessivamente, indústrias produtoras de bens com crescente elasticidade – renda da demanda.

Na América Latina, o Estado também participou ativamente do processo de industrialização, focalizando indústrias de bens de consumo não duráveis e duráveis e ainda de bens intermediários, independentemente da natureza do capital das empresas. Em grande medida, a extensão dos setores promovidos esteve associada ao tamanho das economias de cada país. Sob a argumentação de Raul Prebisch, quanto à perversidade das relações desiguais de troca entre países centrais e periféricos, que perpetuariam

a restrição externa ao crescimento, a industrialização era vista como forma de promover o desenvolvimento das sociedades latino-americanas. O que se popularizou como estratégia de substituição de importações, na verdade, constituiu um grande esforço dos países latino-americanos em criar capacidade produtiva local, com a contribuição de empresas estatais, nacionais e estrangeiras. Para tanto, foram introduzidos incentivos ao investimento e regulações contra importações, bem como criadas empresas estatais principalmente para atuar na área de infraestrutura. Mas, ao contrário do padrão asiático, a meta e, portanto, o controle das políticas não era definido pelo desempenho nos mercados e sim pela construção de capacidade produtiva. Para os asiáticos, o indicador de controle era o sucesso exportador; para os latino-americanos, o grau de nacionalização da produção.

A despeito das diferenças de sucesso destas duas regiões – nos países do leste asiático o dinamismo industrial após os anos 1980 é nitidamente superior ao dos países da América Latina –, uma avaliação de suas estruturas industriais revela que, sob a ótica desenvolvimentista, há sucesso quando a estratégia e as políticas dos estados nacionais são coerentes com o estágio de desenvolvimento das nações.

Os principais desafios para a corrente desenvolvimentista estão associados à capacidade do Estado evoluir em sua forma de intervenção, adaptando-se às mudanças na indústria. Note-se que ao longo do tempo, com a maturação da competência industrial de um país, a intervenção reduz-se para abrir mais espaços para as decisões privadas.

24.5 A Política Industrial pela Ótica da Competência para Inovar

A política industrial pela ótica das competências para inovar destaca as relações entre estrutura de mercado, estratégia empresarial e progresso técnico. Este ponto de partida é de fundamental importância, pois aí se ressalta não apenas como a estrutura de mercado influencia as estratégias empresariais – uma causalidade que já tinha sido amplamente discutida pelo paradigma denominado “Estrutura-Conduta-Desempenho” –, mas também como as estratégias empresariais possuem a capacidade de alterar as próprias estruturas de mercado (ver Capítulo 4). É claro que este processo é mais visível em setores com ritmo elevado de mudança tecnológica. Assim, inspirada em Schumpeter, esta abordagem enfatiza que as inovações constituem-se no motor do desenvolvimento do capitalismo.

Como discutimos nos Capítulos 8 e 18, os economistas associados a esta corrente rejeitam os pressupostos de equilíbrio nos mercados, informação perfeita e racionalidade dos agentes. A racionalidade é limitada, persistem assimetrias de informação e as externalidades não são meros determinantes de falhas de mercado que devem ser corrigidas para convergir para o caminho do equilíbrio. Assimetrias e externalidades são a razão de ser do processo de acumulação e crescimento de uma economia de mercado. A competição é um processo dinâmico, o mercado é o *locus* das interações estratégicas, da rivalidade entre as empresas, é este seu principal agente. As empresas investem na formação de competências para criar assimetrias competitivas, diferenciar produtos e ganhar posição no mercado; esse é o ânimo que move as empresas a crescer à frente de seus concorrentes.

Quatro aspectos são chaves para o entendimento desta abordagem:

1. *Concorrência por inovação tecnológica*: Diversamente da visão das falhas de mercados, que é centrada na visão convencional de concorrência via preços, aqui se enfatiza a competição por meio de inovações e seus efeitos dinâmicos.
2. *Inter-relações entre agentes econômicos*: Existem vantagens advindas da cooperação entre empresas e dessas com universidades, centros de pesquisa e mesmo consumidores, o que se expressa no conceito de aprendizado por interação.
3. *Estratégia, capacitação e desempenho*: As empresas avaliam seu ambiente competitivo, definem os caminhos a seguir, sob a restrição do nível da capacitação existente, alocam recursos para o fortalecimento da capacitação tecnológica que, uma vez posta em marcha, irá definir parâmetros de eficiência produtiva e diferenciação de produtos para cada uma.
4. *Importância do ambiente e processo seletivo*: Fortalece a importância econômica das tecnologias superiores ao longo do tempo, de modo que melhores práticas são repetidamente introduzidas e tornam-se referências móveis e constantes para a conduta dos agentes econômicos.

Assim, o que os economistas ortodoxos chamam de falhas de mercado, na verdade, constituem a força do crescimento e da mudança estrutural do capitalismo, que podem e devem ser induzidas mediante regimes específicos de regulação e incentivos. A sequência lógica da argumentação de origem schumpeteriana parte do processo de concorrência pela inovação; o investimento em inovações, por sua vez, é pleno de incertezas, abrindo espaços para a intervenção pública orientada para induzir as empresas a experimentar, descobrir e introduzir produtos, serviços e processos superiores aos existentes em um mercado específico. O mercado, por sua vez, é relevante como espaço de seleção entre agentes e não como um mecanismo de alocação. Embora o protagonista seja a empresa e o espaço de atuação o mercado, o Estado tem um papel relevante a desempenhar, seja ampliando a intensidade do processo seletivo, seja criando instituições facilitadoras do processo de geração e difusão de novas tecnologias.

Nesse campo, a política industrial e a política tecnológica superpõem-se, dando lugar ao que se denomina hoje política de inovação. O âmbito da política industrial pela ótica da competência para inovar deve ser o estímulo a um ambiente econômico competitivo. Mais do que conferir prioridade à política antitruste, o Estado volta-se para estimular as empresas a desenvolverem novas capacitações, uma vez que o progresso tecnológico baseia-se em conhecimentos tácitos e específicos, de difícil transferência. Além disso, o Estado deve estimular a articulação e as alianças estratégicas para que essas aumentem as capacitações empresariais num ambiente competitivo mais seletivo. Em outras palavras, a cooperação entre empresas pode representar um aspecto da própria concorrência, e, portanto, pró-competitiva e não uma conduta antagônica a ela.

A intervenção pública deve focalizar tanto o lado da demanda quanto a capacidade de oferta de novas tecnologias. O objetivo constante é lograr que qualquer posição de mercado possa ser disputada por firmas inovativas. Pelo lado da demanda, são úteis os subsídios para agentes interessados na difusão de novas tecnologias, principalmente aquelas associadas a retornos crescentes de adoção. Pelo lado da oferta, é necessário apoiar as iniciativas de construção de capacitação tecnológica.

Para promover um ambiente indutor de condutas tecnológicas proativas, é necessário mesclar instrumentos genéricos, que afetam o conjunto de agentes econômicos, com medidas seletivas, focalizadas em um grupo específico de empresas. Os instrumentos de intervenção constituem em: subvenção a projetos de alta densidade tecnológica, incentivos fiscais à pesquisa e desenvolvimento, financiamento em condições preferenciais para a inovação, compras do setor público, e a disponibilidade de capital de risco para novos empreendimentos, além de medidas orientadas a garantir a apropiabilidade privada do investimento tecnológico (patentes) e manter padrões técnicos (metrologia, padronização e qualidade). Adicionalmente, são extremamente importantes os investimentos e ações orientadas ao fortalecimento das instituições que compõem um sistema de inovação local ou nacional, especialmente institutos dedicados à investigação científica e tecnológica e à formação de recursos humanos de alto nível.

Todavia, não podemos esquecer que os recursos para “fazer política” são limitados e os processos decisórios também carregados de racionalidade limitada. Por lidar com a inovação tecnológica, a política industrial pela ótica da competência para inovar opera, ela própria, sob incerteza, fazendo da existência da uma institucionalidade pública altamente capacitada um importante requisito para o seu sucesso.

24.6 Política Industrial na Prática: Instrumentos Horizontais e Verticais

A discussão da política industrial na prática exige como passo prévio demarcar as relações existentes entre esta e a política macroeconômica. De um lado, a política macroeconômica afeta a política industrial ao: (a) determinar os preços relativos de produtos transacionáveis e não transacionáveis por meio da taxa de câmbio; (b) influenciar o nível de investimentos via taxa de juros; (c) sinalizar, mediante a estabilidade macroeconômica e a capacidade fiscal do Estado, quanto à possibilidade de implantar políticas de incentivo e de investimento em infraestrutura, educação, ciência e tecnologia. Diferentes indústrias são afetadas de forma igualmente diferente pela política macroeconômica. Um exemplo claro é a política cambial: indústrias exportadoras e importadoras têm, naturalmente, suas receitas afetadas de forma contrária, caso a taxa de câmbio esteja sobre ou subvalorizada. Do outro lado, uma política industrial bem-sucedida pode facilitar a gestão da política macroeconômica por meio do aumento da eficiência e da produtividade na economia.

Um ponto essencial da política industrial refere-se ao caráter deliberado da atuação estatal. É bem verdade que existe uma diferença entre a política industrial e tecnológica explícita e implícita. A primeira se expressa em leis, regulações, órgãos, planos de desenvolvimento e declarações governamentais. Já a política industrial implícita é mais difícil de identificar. Não é incomum que determinadas ações de política industrial não sejam implementadas como anunciadas ou mesmo não atinjam plenamente seus objetivos. Analogamente, um país pode estar adotando uma determinada política industrial, sem que essa ação esteja explicitamente contida em documentos formais.

A política industrial pode ser descrita e avaliada de acordo com a natureza do instrumento e o alvo pretendido. Existem dois alvos prioritários da política industrial. De um lado, estão as chamadas políticas horizontais (ou funcionais), pautadas em medidas de alcance global. De outro lado, estão as políticas verticais (ou seletivas), desenhadas para fomentar indústrias, cadeias produtivas ou grupos específicos de empresas.

Os instrumentos de política industrial podem ser agrupados de acordo com a sua natureza. Um primeiro grupo congrega o *regime de regulação*: a arbitragem do processo concorrencial, englobando a política antitruste e a comercial, assim como regulações referidas à propriedade intelectual, consumidor e meio ambiente. Em última instância, procura-se ampliar a pressão competitiva sobre as empresas, por meio, por exemplo, de uma política antitruste mais vigorosa, ou a

redução dessa pressão, mediante o aumento da proteção tarifária. Um segundo grupo relaciona-se ao *regime de incentivos*, por intermédio de medidas fiscais e financeiras, como os incentivos fiscais à pesquisa e desenvolvimento e créditos e estímulos à exportação.

As *políticas industriais horizontais* são aquelas que buscam melhorar o desempenho da economia na sua totalidade, sem privilegiar alguma indústria específica. Retornando à definição de política industrial, esse tipo de política busca alterar o mecanismo geral de alocação de recursos na produção.

Do ponto de vista da regulação, vários instrumentos encaixam-se nesse tipo de política, com destaque para:

1. Concorrência: repressão de condutas anticompetitivas (vendas casadas, acordos para fixação de preços, discriminação de preços etc.) e controle dos atos de concentração (fusões, aquisições e *joint-ventures*).
2. Infraestrutura: políticas de concessões (privatizações) e controle administrativo de preços (mecanismos de reajuste de tarifas de serviços, como energia elétrica e telecomunicações).
3. Comércio exterior: política tarifária e não tarifária, prevenção de concorrência desleal (*anti-dumping*, diretos compensatórios e salvaguardas).
4. Propriedade intelectual: patentes, marcas e transferência de tecnologia.

Do ponto de vista dos incentivos, vários instrumentos são mobilizados pelo Estado, sendo os principais:

1. Inovação: incentivos ao gastos com pesquisa e desenvolvimento, fomento à difusão de tecnologias e informações.
2. Capital: crédito e financiamento a longo prazo, estímulos às exportações (crédito e seguro de crédito), financiamento às importações.
3. Incentivos fiscais: deduções fiscais em âmbito nacional, estadual ou municipal para promoção de atividades industriais.
4. Compras de governo: mecanismos preferenciais para produtores locais.

Também são alvos da política industrial horizontal políticas genéricas associadas ao desenvolvimento do entorno onde operam as empresas. Dentre elas, destacam-se: a política de infraestrutura: geração e distribuição de energia elétrica, transporte, portos e telecomunicações; a política de recursos humanos: educação, qualificação da mão de obra; e a política de ciência e tecnologia: investimentos e subsídios aos institutos de pesquisa e universidades de um país.

Já as *políticas industriais verticais* privilegiam deliberadamente uma indústria específica. Ou seja, a partir de decisões estratégicas, o Estado mobiliza parte dos instrumentos anteriormente descritos, focalizando e privilegiando um conjunto de empresas, indústrias ou cadeias produtivas. Retornando, mais uma vez, à definição de política industrial, este tipo de política visa a modificar as regras de alocação entre setores. Por esse motivo, elas também são denominadas de seletivas ou de *targeting*.

Quatro são os principais argumentos que justificariam a proeminência de algumas indústrias comparativamente a outras:

1. Indústrias com maior valor agregado: supondo tudo mais constante, uma maior proporção de trabalhadores localizados em indústrias com maior valor agregado resultaria em maior renda *per capita*.
2. Indústrias com grande poder de encadeamento: esses setores apresentariam grande efeito multiplicador ao longo da cadeia produtiva, com efeitos “para frente” (*forward linkage*) ou “para trás” (*backward linkage*). Historicamente, esse foi o argumento utilizado para estimular indústrias produtoras de insumos básicos, como a siderurgia (grandes efeitos de encadeamento para frente) ou indústrias produtoras de bens de consumo de alto valor agregado, como a indústria automobilística (grande efeito de encadeamento para trás).
3. Indústrias com grande dinamismo potencial: tendo em vista que o crescimento da renda agregada seria maior, consequentemente, haveria um incremento da renda *per capita*.
4. Indústrias nascentes ou com retornos crescentes de escala: o Estado deveria estimular o desenvolvimento de novas indústrias, que, inicialmente, apresentariam custos mais elevados do que os verificados em países já produtores.

É importante indicar que, apesar de ênfases distintas, os países, geralmente, adotam políticas de cunho horizontal e vertical, simultaneamente, embora a importância delas tenda a se alterar ao longo do tempo. A partir da década de 1990, foi se configurando um padrão de intervenção estatal no âmbito dos países industrializados, nos quais a política industrial recorre predominantemente a instrumentos de cunho horizontal. Ao mesmo tempo, instrumentos de cunho vertical são aplicados especialmente na promoção de indústrias nascentes e em declínio, nas quais a reestruturação industrial se mostra particularmente relevante. Ou seja, tanto indústrias de alto ritmo tecnológico, quanto indústrias muito maduras, que tendem a empregar um elevado contingente de pessoas, acabam sendo priorizadas em termos de política industrial. Naturalmente, os instrumentos são divergentes, utilizando-se estímulos financeiros para pesquisa e desenvolvimento no primeiro caso, e restrições comerciais de caráter não tarifário no segundo.

24.7 Experiências Nacionais: Os Países da OCDE e o Brasil

Ao contrário do discurso político e da postura das nações desenvolvidas em fóruns internacionais, em que se destaca a defesa da liberalização econômica como melhor forma de se promover o crescimento econômico e o bem-estar social, a maioria dos países implementa, de fato, políticas industriais. Estas buscam fortalecer a capacidade competitiva de suas empresas, em um ambiente internacional crescentemente aberto aos fluxos de capital, bens, serviços e tecnologia. Na virada do século, as ações dos países da OCDE podem ser divididas em medidas de apoio à capacidade de concorrência externa e medidas de apoio às atividades tecnológicas.

1. Medidas de apoio à capacidade de concorrência externa: auxílio às exportações (financiamento de clientes estrangeiros de fabricantes nacionais, reforço direto à capacidade exportadora); medidas setoriais (redução de setores em declínio como os setores siderúrgico, têxtil, naval; reestruturação de setores expostos a acirrada concorrência internacional, como o automobilístico, o aeroespacial e o eletrônico); medidas de apoio a empresas em dificuldades (apoio à consolidação financeira via fusões e aquisições).
2. Medidas de apoio às atividades de P&D e à difusão tecnológica: redução dos custos de P&D para as empresas, sobretudo mediante vantagens fiscais, com ênfase em projetos cooperativos; apoio a tecnologias específicas e estratégicas, principalmente informática, biotecnologia e novas formas de energia; apoio à difusão tecnológica: reforço aos investimentos em P&D das empresas.
3. Salvaguarda do tecido industrial: políticas industriais regionais (principalmente regiões subdesenvolvidas ou com especialização em setores em declínio); apoio a pequenas e médias empresas por intermédio de tratamento fiscal privilegiado e estímulos à criação de empregos ou estímulos à modernização (P&D, automação, consultoria em gestão); apoio ao emprego e à formação profissional.
4. Apoio ao investimento fixo, por meio do financiamento em condições favoráveis a planos correntes de investimento.

No Brasil, a política industrial ativa correspondeu ao processo de substituição de importações, que se iniciou nos anos 1930 com os bens não duráveis de consumo, aprofundou-se nas décadas de 1950 e 1960 com os bens duráveis de consumo e, nas de 1970 e 1980, com os bens intermediários e parte da indústria de bens de capital. Conforme mencionado anteriormente, o objetivo da política industrial era de criar capacidade produtiva local, porém, deixando em segundo plano algumas questões associadas à eficiência produtiva. As principais características desse longo período, que foi marcado por uma política industrial ativa e com instrumentos de cunho vertical, foram:

1. Estado-empresário: verificou-se a proliferação de empresas estatais tanto nos setores de infraestrutura (como telecomunicações e energia elétrica), quanto na indústria de transformação (siderurgia, petroquímica, fertilizantes);
2. Protecionismo: a indústria nascente sempre contou com uma política comercial baseada especialmente em barreiras não tarifárias. Além disso, mesmo indústrias já maduras continuaram a contar com esse tipo de benefício;
3. Investimento estrangeiro: atração de um elevado contingente de empresas transnacionais para instalarem filiais no país. Porém, a atuação em determinados setores, considerados de segurança nacional, foi desestimulada e, em outros, forçaram-se *joint-ventures* entre empresas estrangeiras e o capital local (privado ou estatal);
4. Incentivos fiscais setoriais e regionais: alguns setores considerados prioritários receberam elevados incentivos fiscais (como empréstimos a taxas de juros subsidiadas). Estes também tiveram uma lógica de tentativa de diminuir a disparidade entre regiões, estimulando a constituição de empresas nas regiões norte, nordeste e centro-oeste.

A direção da política industrial brasileira foi fortemente alterada no governo Collor (1990-92), quando se passaram a privilegiar instrumentos horizontais. Neste sentido, pode ser destacado o Programa Brasileiro de Qualidade e Produtividade (PBQP), que buscava disseminar novas técnicas organizacionais com vistas ao aumento da produtividade. Nesse período, instrumentos de cunho vertical foram utilizados apenas para a indústria de informática. Os principais instrumentos utilizados foram: (a) o início da abertura comercial, eliminando barreiras não tarifárias e promovendo uma reforma tarifária, que resultou num crescimento das importações; (b) o começo do programa de privatização, com a venda do controle acionário em empresas siderúrgicas e de fertilizantes; (c) desregulamentação, revertendo a herança anterior em que o papel diretor do Estado era muito acentuado. Nessa época, também foram introduzidos incentivos fiscais para pesquisa e desenvolvimento, em especial para a indústria de informática.

A partir de 1994, em função do Plano Real, a política industrial subordinou-se de forma veemente à prioridade de estabilização econômica, perdendo ainda mais espaço em relação à política macroeconômica. Como consequência da redução das tarifas e valorização cambial, as importações incrementaram-se substancialmente e foram utilizadas como mecanismo de pressão de disciplina dos preços para os bens produzidos domesticamente. O processo de privatização foi expandido, incluindo setores de infraestrutura, como telecomunicações e distribuição de energia elétrica. A política de concorrência ampliou o seu leque de atuação, pois, até 1994, somente disciplinava práticas anticompetitivas, passando, a partir de então, a atuar em

atos de concentração (fusões, aquisições e *joint-ventures*). Por outro lado, a indústria automobilística foi contemplada com redução de tributos e tarifas de importação superiores à média nacional, como forma de garantir empregos aos metalúrgicos e, especialmente, expandir a produção de automóveis com motores de baixas cilindradas. Da mesma forma, foram mantidos os incentivos à produção em Manaus (Zona Franca) e para a indústria de informática. Cresceram, também, as iniciativas dos estados em promover isenções fiscais para empreendimentos industriais, em particular para a indústria automobilística e para as indústrias têxtil, de confecções e de calçados, estas, principalmente em estados do Nordeste, dando início a um processo de guerra fiscal, nem sempre favorável ao desenvolvimento da indústria brasileira.

24.8 Conclusão

Política industrial deve ser entendida como o conjunto de incentivos e regulações associadas a ações públicas, que podem afetar a alocação inter e intrassetorial de recursos, influenciando a estrutura produtiva e patrimonial, a conduta e o desempenho dos agentes econômicos, em um determinado espaço nacional. Pretende-se aumentar a capacidade produtiva e competitiva das empresas e setores; para países em desenvolvimento, busca-se alcançar os líderes internacionais; para países avançados, o objetivo é manter a posição de liderança ou avançar sobre competidores de outras nações.

A discussão anterior pode ser resumida em quatro pontos principais. Primeiro, não é relevante discutir a intensidade de intervenção, mas os propósitos desta. Em geral, em estágios de desenvolvimento iniciais, as políticas são muito ativas; em estágios superiores, a intervenção torna-se indireta, e os espaços para o capital privado se ampliam. Assim, o estilo da intervenção deve evoluir ao longo do tempo, *pari passu* com a evolução das competências empresariais. Se bem-sucedido, o Estado transita na direção de políticas de coordenação, utilizando instrumentos cada vez mais sofisticados. Ao longo do tempo, a criação de capacidade produtiva cede espaço para medidas que promovam a rivalidade concorrencial, a inserção externa competitiva e o desenvolvimento tecnológico.

Segundo, é importante destacar as relações próximas entre política industrial e outras políticas econômicas, principalmente aquelas de caráter macroeconômico. Estas afetam a conduta e o desempenho dos agentes econômicos ao passo que as políticas industriais, se bem-sucedidas, podem fortalecer a estabilidade e a capacidade de crescimento econômico.

Terceiro, a literatura econômica diverge, profundamente, em termos da necessidade de ações estatais, do tipo de política industrial e quanto a quais devem ser seus objetivos. Em grande medida, essas divergências se devem aos pressupostos teóricos adotados pelas diversas correntes. Mas o avanço do conhecimento econômico indica a existência de espaços legítimos para a intervenção pública, em particular em áreas sujeitas a retornos crescentes de escala, ou, ainda, a presença de externalidades positivas ou negativas. Nesse sentido, é incontestável a necessidade de medidas públicas para promover a concorrência, o desenvolvimento tecnológico e para minimizar efeitos negativos da atividade industrial – concorrência predatória, poluição etc.

Quarto, a análise concreta de experiências nacionais, regionais ou setoriais demanda uma avaliação criteriosa da competência do Estado em exercer um papel proativo na sociedade. Para minimizar as imperfeições da intervenção, o Estado deve ter agências, pessoal qualificado, informação e instrumentos adequados. O poder dessas agências em relação a outras agências públicas deve ser suficiente para empreender sua missão. Mas a real capacidade do Estado de empreender políticas para a indústria deve estar ancorada na legitimidade pública para exercer sua tarefa.

Finalmente, a prática atual dos países industrializados está eliminando as fronteiras entre as políticas industriais, tecnológicas e de comércio exterior. O que diferencia significativamente esses países são os arranjos institucionais e a filosofia subjacente à intervenção.

Nota

1. A OCDE – Organização para Cooperação e Desenvolvimento Econômico – formada pelos países desenvolvidos, é um dos mais ativos organismos mundiais na definição de tendências de política industrial na atualidade.

Bibliografia

- Amsden, A. (1989) *Asia's next giant: South Korea and late industrialization*, Oxford, Oxford University Press.
Amsden, A. (1994) "Why isn't the whole world experimenting with the East Asian model to develop?", *World Development*, v. 22, n. 4.
Aoki, M. (1997) "Unintended fit; organizational evolution and government design of institutions in Japan", in Aoki, M.; Hyung-Ki, K.; Okuno-Fujiwara, M. *The role of government in East Asian economic development*, Oxford, Clarendon Press.

- Arrow, K. J. (1962) "The economics implications of learning by doing", *Review of Economic Studies*, v. 29, junho.
- Correa, P. G.; Villela A. (1995) Política Industrial: fundamentos teóricos com referência ao caso brasileiro. BNDES, *Textos para Discussão*, v. 24, Rio de Janeiro.
- Dosi, G.; Pavitt, K.; Soete, L. (1990) *The economics of technical change and international trade*. Nova York, New York University Press.
- Evans, P. (1995) *Embedded Autonomy: States and Industrial Transformation*, Nova Jersey, Princeton University Press.
- Ferguson, P.R.; Ferguson, G. J. (1994) *Industrial Economics; Issues and Perspectives*, Nova York, New York University Press, segunda edição.
- Freeman, C. (1974) *The economics of industrial innovation*, Londres, Penguin Books.
- Gerschenkron, A. (1962) *Economic Backwardness in Historical Perspective*, Cambridge, Harvard University Press.
- Hay, D. A.; Morris, D. J. (1991) *Industrial Economics and Organization: Theory and Evidence*, Nova York, Oxford University Press, segunda edição.
- Herrera, A. (1973) "Social Determinants of Science Policy in Latin America: Explicit Science Policy and Implicit Sciency Policy", Cooper, C. (org.). *Science, Technology and Development: the political economy of technical advance in underdeveloped countries*. Londres, Frank Cass.
- Johnson, C. (1982) *MITI and the Japanese Miracle: the growth of industrial policy, 1925-1975*, Stanford, Stanford University Press.
- Jorge, I. (1994) Privatização; experiências internacionais, aspectos teóricos e resultados. Faculdade de Ciências Econômicas, Dissertação de Mestrado, Porto Alegre.
- Katz, J. (1999) *Pasado y presente del comportamiento tecnológico de América Latina*, Santiago, Cepal, mimeo.
- Lipsey, R.; Carlaw, K. (1998) *Assessing Innovation Policies: Taking Schumpeter Seriously on Technology Policy*, Industry Canada Working paper 25, Ottawa.
- List, F. (1841) *The national system of political economy*, edição inglesa, Londres, Longman.
- Metcalfe, S. (1995) "The Economic Foundations of Technology Policy: Equilibrium and Evolutionary Perspectives", Stoneman, P. (org.) *Handbook of Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Mowery, D. (1995) "The Practice of Technology Policy", Stoneman, P. (org.) *Handbook of Economics of Innovation and Technological Change*, Blackwell, Oxford.
- Page, J. (1994) "The East Asian miracle, an introduction", *World Development*, v. 22, n. 4, 1994.
- Possas, M. (1995) *Competitividade Sistêmica e Política Industrial: Implicações para o Brasil*, IE-UFRJ, Rio de Janeiro, mimeo.
- Rodrik, D. (1997) "The paradoxes of the successful State", *European Economic Review*, v. 41, p. 411-442.
- Romer, P. M. (1990) "Endogenous technological change", *Journal of Political Economy*, v. 98, n. 5, outubro.
- Solow, R. M. (1957) "Technical change and the aggregate production function", *Review of Economics and Statistics*, v. 39, agosto.
- Suzigan, W.; Villela A. V. (1997) *Industrial Policy in Brazil*. Campinas, Instituto de Economia da Unicamp.
- Williamson, J. (1993) "Democracy and the Washington Consensus", *World Development*, v. 21, n. 8, p. 1329-1336.

Política Comercial

Jorge Chami Batista e Marta Lemme

25.1 Introdução

A política comercial compreende um conjunto de instrumentos capazes de modificar diretamente os preços relativos locais entre bens exportados e importados de um país, mantida a taxa de câmbio constante. Dessa forma, por alterar os incentivos à produção e ao consumo dos diferentes setores da economia, a política comercial pode ser vista como parte importante da política industrial.

Os economistas tendem a concordar que há ganhos com o comércio internacional. Um país que não faz comércio com o resto do mundo (país em autarquia) é forçado a produzir todos os bens e serviços a serem consumidos no país. Ao fazer comércio, o país pode se tornar mais especializado, exportando os bens e serviços que faz melhor e mais barato e importando os bens e serviços que o resto do mundo faz melhor e mais barato. Assim, o país e o resto do mundo tornam-se mais eficientes. O comércio também tende a ampliar as oportunidades de aprendizagem, estimulando imitação e inovação (Grossman e Helpman, 1991).

Na ausência de políticas comerciais, diz-se que o país está em livre-comércio. Entre a autarquia e o livre-comércio há muitos diferentes graus de intervenção possíveis por meio de políticas comerciais. O conjunto de modelos teóricos e de evidências empíricas existentes na literatura não oferece recomendações claras e consistentes para a implementação de políticas comerciais (Krugman, 1987). Dessa forma, descreveremos neste capítulo os principais argumentos a favor e contra o uso de políticas comerciais, deixando ao leitor ponderá-los. No entanto, é relativamente consensual que, ao se decidir pela implementação de políticas comerciais, estas devem ser temporárias e seletivas.

É um fato bem documentado que países avançados como Estados Unidos, Alemanha e Japão iniciaram seus processos de industrialização com o apoio de políticas comerciais. Mais recentemente, países como República da Coreia, Taiwan e Cingapura também se industrializaram com o uso de políticas industriais e comerciais, obtendo grande expansão de suas exportações e alcançando altos níveis de renda. Contudo, o nexo causal entre as diferentes políticas adotadas e o sucesso econômico desses países não é evidente. Para alguns analistas, essas políticas tiveram papel positivo (Stiglitz, 1996) ao passo que, para outros, elas não foram relevantes (Krugman, 1994).

O Brasil adotou políticas fortemente protecionistas em quase todo o período do pós-guerra, especialmente entre o primeiro choque do petróleo e o final dos anos 1980, sendo um dos últimos países emergentes a iniciar seu processo de abertura comercial. Nos anos 1990, a abertura unilateral brasileira reduziu significativamente as tarifas aduaneiras e eliminou a quase totalidade das barreiras não tarifárias. O Brasil também assinou o Tratado de Assunção, dando início ao Mercosul, uma união aduaneira que pretendia tornar inteiramente livre o comércio das mercadorias entre Argentina, Brasil, Uruguai e Paraguai, e estabelecer uma Tarifa Externa Comum (TEC) para o comércio entre esses países e o resto do mundo.

Além desta introdução, este capítulo está organizado da seguinte forma: a segunda seção descreve os instrumentos da política comercial. As medidas do grau de proteção de um país são apresentadas na terceira seção, ao passo que os efeitos da política comercial são examinados na quarta seção. A quinta seção trata das principais regras do comércio internacional e a sexta seção faz uma breve referência aos acordos regionais. Ao longo de todo o capítulo, procura-se ilustrar os pontos abordados com as práticas no Brasil.

25.2 Instrumentos da Política Comercial

Existe um leque muito amplo de instrumentos de política comercial. Neste item, serão abordados alguns desses instrumentos, à exceção das medidas antidumping e compensatórias, que serão tratadas na próxima seção.

25.2.1 TARIFAS DE IMPORTAÇÃO E SUBSÍDIOS ÀS EXPORTAÇÕES

Uma tarifa de importação é um imposto cobrado sobre bens importados. A tarifa pode ser *ad valorem* ou específica. A primeira é a mais frequentemente utilizada pelos países e é cobrada por um percentual aplicado ao valor do bem importado. Por exemplo, 10% de cada 100 dólares de importação geram dez dólares de imposto.¹ A segunda é cobrada por um valor fixo por unidade do bem importado. Por exemplo, dez centavos de dólar por quilo.

Dessa forma, mantido o volume (quantidade) de importação constante, o imposto arrecadado com uma tarifa *ad valorem* tende a aumentar quando o preço do bem importado aumenta, porém não se altera quando a tarifa é específica. Por esta razão, a tarifa específica tende a ser aplicada sobre produtos primários, cujos preços são mais voláteis no mercado internacional. Os Estados Unidos aplicam tarifas específicas, por exemplo, sobre suas importações de açúcar, fumo, e suco de laranja, produtos em que o Brasil é competitivo no mercado internacional. O Brasil não utiliza tarifas específicas, tendo em vista o risco de violar os limites máximos das tarifas *ad valorem* acordados na OMC e no Mercosul. No passado, o Brasil utilizou alguns instrumentos de políticas protecionistas, como os leilões cambiais e o depósito prévio à importação, que são equivalentes a um acréscimo na tarifa de importação.

Embora os países tipicamente imponham impostos (tarifas) apenas sobre suas importações, impostos podem igualmente ser aplicados sobre as exportações. Os impostos sobre exportações, quando ocorrem, são comumente aplicados com a finalidade de garantir o abastecimento interno a preços inferiores. O Brasil, por exemplo, aplica um imposto sobre as exportações de couro desde dezembro de 2000. Em países com grande participação nas exportações mundiais de determinados produtos primários, os impostos de exportação podem aumentar os preços internacionais dessas mercadorias e, dessa forma, melhorar os termos de troca dos países exportadores.

Um subsídio à exportação é análogo a uma tarifa de importação, podendo também ser *ad valorem* (proporcional ao valor exportado) ou específico (fixo por unidade exportada). Para todos os efeitos, um subsídio é um imposto com o sinal negativo. Subsídios também podem ser dados ao crédito de exportação. Nesse caso, o governo reduz as taxas de juros pagas nos financiamentos às exportações relativamente às taxas de mercado para financiamentos com prazos e riscos equivalentes.

25.2.2 COTAS DE IMPORTAÇÃO E RESTRIÇÕES VOLUNTÁRIAS ÀS EXPORTAÇÕES

As cotas de importação fixam a quantidade ou o valor máximo que pode ser importado em um determinado período de tempo, geralmente um ano. Para implementar uma cota de importação, o governo normalmente emite licenças de importação para pessoas físicas ou jurídicas residentes no país ou no exterior. Ao emitir um número determinado de licenças, cada uma autorizando a importação de certa quantidade ou valor do produto, o governo garante um teto para o volume ou valor total a ser importado. Se a cota impõe uma restrição ao volume ou valor que seria importado na sua ausência, então o preço local de importação do bem aumenta e os detentores de licenças de importação obtêm lucros conhecidos como renda da cota.

Assim, as cotas de importação, quando restritivas, da mesma forma que as tarifas, tendem a aumentar o preço pago domesticamente pelo bem importado e reduzem as quantidades importadas. Porém, ao contrário das tarifas, o governo nada arrecada como receita direta da importação, se ele simplesmente emite e distribui as licenças. Esse procedimento pode ainda ter custos ao país, se envolver corrupção por parte da burocracia estatal responsável pela concessão das licenças. Alternativamente, o governo pode leiloar as licenças de importação, quando há um grande número de potenciais compradores, passando a receber a renda da cota. Como veremos, os acordos multilaterais restringiram severamente o uso desse instrumento protecionista. O Brasil eliminou praticamente todas as cotas e proibições de importações de bens em 1990.

Outra prática comum é estabelecer uma cota de importação com tarifa reduzida ou livre de tarifa para determinados países exportadores (cotas tarifárias). As importações acima da cota sofrem a incidência de tarifa superior. Nesse caso, não há limite fixo para o total importado, mas a renda da cota obtida pelos detentores de licenças depende da diferença entre as tarifas dentro e fora da cota. Um exemplo é a chamada cota Hilton de importações de carnes nobres pela União Europeia oferecida ao Brasil, Argentina, Uruguai, Austrália, Nova Zelândia, EUA e Canadá. Esta cota foi negociada nos início dos anos 1980, durante a Rodada Uruguai, como uma compensação da União Europeia a alguns países que exportavam carnes nobres para os hotéis da cadeia Hilton desde os anos 1950. Cotas tarifárias também foram recentemente introduzidas no comércio intra-Mercosul (pêssegos, pneus remoldados), nesta instância com o objetivo de restringir o livre-comércio e reduzir as importações dos países-membros.

Em alguns casos, cotas de importação foram impostas por meio de acordos voluntários de exportação. Essa foi frequentemente uma maneira de driblar restrições impostas por acordos multilaterais. Um dos exemplos mais famosos foi a restrição quantitativa aceita pelo Japão às suas exportações de automóveis para os Estados Unidos a partir de 1981. Essas restrições são forçadas pelo país importador ao país exportador sob a ameaça de impor unilateralmente restrições ainda mais severas. O país exportador aceita a restrição se a ameaça é crível. Como as licenças de importação são distribuídas aos países exportadores, eles ao menos se beneficiam da renda da cota. Portanto, esse instrumento tem um custo elevado para o país importador, já que a quantidade importada diminui, o preço local do bem importado aumenta, e o governo não obtém qualquer receita direta.² Restrições voluntárias têm sido mais recentemente negociadas entre o Brasil e os demais membros do Mercosul e entre o Brasil e a China.

25.2.3 NECESSIDADE DE REQUISITOS LOCAIS E AQUISIÇÃO DE BENS

Outro instrumento protecionista são as normas legais que exigem uma fração mínima de conteúdo doméstico nos bens vendidos localmente. Esse instrumento foi utilizado por vários países em desenvolvimento que estabeleciam uma parcela mínima para o valor adicionado local no preço do bem. As compras de governo, de empresas estatais ou de empresas fortemente regulamentadas também podem ser utilizadas como instrumentos da política comercial, se são forçadas a privilegiar fornecedores domésticos, mesmo quando seus preços são superiores aos dos produtos importados.³

O Brasil manteve requisitos locais até 1990. As importações de equipamentos eram reguladas, permitindo-se apenas a importação de bens sem similar nacional.⁴ Os exames de similaridade e dos índices de nacionalização eram feitos por empresas estatais ou pelas associações empresariais. Mais recentemente, o Brasil passou a adotar o conceito de Processo Produtivo Básico (PPB) como uma das contrapartidas para a obtenção de benefícios fiscais por parte das empresas situadas na Zona Franca de Manaus (ZFM) e de fabricantes de bens de informática e automação no país. O Processo Produtivo Básico (PPB) foi criado pela Lei n. 8.387, de 30 de dezembro de 2001, e consiste de etapas fabris mínimas necessárias que as empresas devem cumprir para fabricar determinado produto e se beneficiar dos incentivos fiscais estabelecidos pela ZFM e pela Lei de Informática.⁵

25.2.4 BARREIRAS TÉCNICAS E BURECRÁTICAS

Diferenças de regulamentação entre os países com respeito aos seus imperativos de: segurança nacional; prevenção de práticas enganosas; proteção da saúde ou segurança humana, da saúde ou vida animal ou vegetal, ou do meio ambiente; podem significar enormes obstáculos ao comércio internacional. Considera-se que as normas e os regulamentos constituem barreira técnica ao comércio quando são muito pouco transparentes, ou não se baseiam em normas internacionalmente aceitas, ou implicam procedimentos de avaliação de conformidade não transparentes e/ou dispendiosos, ou, ainda, inspeções excessivamente rigorosas (CNI/INMETRO, 2002). Os produtos importados de origem vegetal e animal são, particularmente, sujeitos a normas fitosanitárias e a procedimentos mais exigentes e menos transparentes do que os aplicados aos produtos domésticos. Procedimentos burocráticos podem elevar os custos de comércio de alguns bens de tal forma que, no limite, acabam por inviabilizar o próprio comércio do bem.

25.3 Medidas do Grau de Proteção

O cálculo do grau de proteção de uma economia que possa ser comparado ao de outras economias não é nada trivial. Um dos indicadores mais utilizados é a tarifa média simples (média aritmética). Esse indicador nada informa sobre a distribuição dos valores das tarifas por produtos. Outras estatísticas descritivas, tais como valor máximo, mínimo, moda, mediana, e

desvio-padrão podem ajudar nessa tarefa. No entanto, alguns setores ou produtos podem ter maior peso que outros na pauta de importações. Para isto, costuma-se calcular a tarifa média ponderada. O problema com esse indicador é que uma alta tarifa pode restringir as importações de tal forma, que a multiplicação da alta tarifa pelo valor importado pode resultar em um valor próximo de zero. Dessa forma, subestimar-se-ia o grau de proteção. As barreiras não tarifárias são mensuradas tipicamente pela frequência em que ocorrem no conjunto de itens tarifários, ou por meio de uma estimativa do seu valor equivalente a uma tarifa *ad valorem*. Na prática, essas estimativas são bastante imprecisas. As tarifas nominais de importação também nada informam sobre o resultado líquido da proteção quando, além do bem final, os insumos, partes e componentes também sofrem a incidência de imposto de importação. Para isso, calcula-se a taxa (tarifa) de proteção efetiva, considerando a tarifa sobre o bem final, as tarifas sobre os insumos importados, e o coeficiente direto e indireto desses insumos na produção do bem.⁶

A tarifa nominal média brasileira decresceu de 57,5% em 1987 para 13,5% em 1993 e para 12% em 1995, mas manteve-se pouco acima de 15% entre 1997 e 2000, voltando a decrescer gradativamente a partir de então até atingir 10,7% em 2004 (Kume et al., 2003). A tarifa efetiva caiu de 45% em 1990 para cerca de 14% em 1994, mas manteve-se relativamente elevada em 25% de 1997 a 2000,⁷ sem considerar o reaparecimento de barreiras não tarifárias. Segundo o indicador de restrições às importações do Banco Mundial, que inclui barreiras tarifárias e não tarifárias, o Brasil mantém-se como uma economia mais fechada que a grande maioria dos países em desenvolvimento.⁸ Embora tanto o desvio-padrão quanto a amplitude tenham se reduzido substancialmente entre o final dos anos 1980 e o final dos anos 1990, a estrutura da tarifa nominal sofreu pouca alteração. Isto sugere uma certa endogeneidade da estrutura tarifária no Brasil.⁹ Assim, as indústrias de veículos automotores, equipamentos eletrônicos e vestuário mantiveram-se entre os setores com as maiores tarifas nominais.

25.4 Efeitos da política comercial

A política comercial pode ser usada com diferentes finalidades: (1) melhorar os termos de intercâmbio; (2) corrigir as distorções causadas por falhas de mercado; (3) obter uma distribuição de renda ótima; e (4) alcançar metas exógenas de comércio, produção e consumo doméstico. Trataremos no que se segue apenas das duas primeiras finalidades.¹⁰

25.4.1 EFEITOS ESTÁTICOS EM COMPETIÇÃO PERFEITA

Ao adotar uma medida de proteção, o governo cria uma cunha entre o mercado interno e o externo, implicando, em geral, aumento do preço doméstico em relação ao preço internacional, redução do consumo doméstico e das importações e aumento da produção. Se o país for grande, isto é, relevante no mercado internacional como demandante do produto, também será observada redução do preço internacional, implicando ganhos dos termos de troca do país que adotou a tarifa.

Com base na mensuração das variações dos excedentes do produtor e do consumidor, é possível quantificar o efeito da tarifa sobre o bem-estar, conforme ilustrado pelo Quadro 25.1.

A aplicação desse mesmo método de análise a um subsídio às exportações revela um efeito negativo sobre o bem-estar, uma vez que haverá perdas de eficiência na produção e no consumo e o efeito nos termos de troca, se houver, também será negativo. As demais medidas protecionistas não tarifárias tendem a ter um efeito estático sobre o bem-estar pior que o da tarifa,¹¹ seja porque não geram receita para o governo (exceto as medidas de defesa comercial: direito antidumping e direito compensatório), seja por permitir poder de mercado aos ofertantes locais.

As estimativas desse efeito estático das políticas comerciais sobre o bem-estar de algumas economias tendem a ser negativas. A Tabela 25.1 revela que esse efeito é pequeno nos países desenvolvidos com baixa proteção, porém já foi significativo nos países em desenvolvimento com alta proteção.

25.4.2 IMPERFEIÇÕES DE MERCADO

Os efeitos sobre o bem-estar tornam-se mais complexos quando se admite que as economias estão sujeitas a falhas de mercado.¹² Nesse caso, o conceito de variação no excedente do produtor não é uma boa aproximação para os custos e benefícios da política comercial sobre a oferta. O argumento tem como fundamento a teoria da segunda melhor alocação de recursos. Essa teoria estabelece que a política de não intervenção é desejável em qualquer mercado somente se todos os outros mercados estiverem funcionando de modo adequado. Caso contrário, uma intervenção pode aumentar o bem-estar, mesmo que distorça os incentivos em um mercado específico.¹³

QUADRO 25.1 EFEITOS DA TARIFA SOBRE BEM-ESTAR

Gráfico 1.a

País Grande

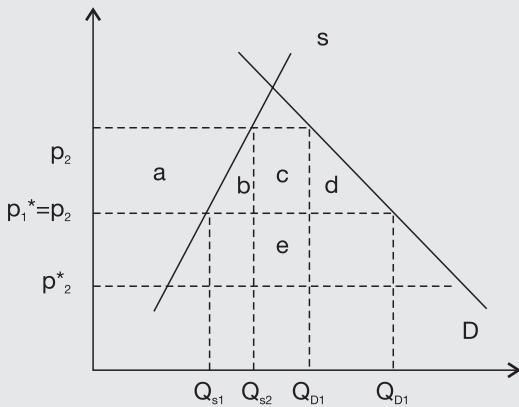
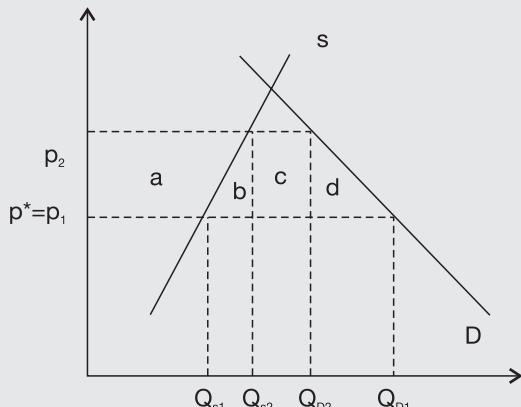


Gráfico 1.b

País Pequeno



p_1^* – preço internacional antes da tarifa; p_1 – preço doméstico; p_2^* – preço internacional após tarifa; p_2 – preço doméstico após tarifa.; $(QD_1 - Qs_1)$ – importação antes da tarifa; $(QD_2 - Qs_2)$ – importação após tarifa

Perdas e Ganhos

Redução do Excedente do Consumidor: $a + b + c + d$

Aumento do Excedente do Produtor: a

Ganho do Governo (receita com tarifa): $c + e$

Resultado Líquido

- País Grande: $e - (b + d)$;

se $e > b + d$, há aumento do bem-estar e a tarifa corresponde à “tarifa ótima”;

se $e < b + d$, redução do bem-estar

- País Pequeno: Redução do bem-estar equivalente a $(b + d)$

Significado de e, b e d

- e – ganhos dos termos de troca (só ocorre no caso de país grande)

- b – perda de eficiência na produção; e

- d – perda de eficiência no consumo

TABELA 25.1 Custo Estimado da Proteção (% da renda)

País	Ano	Custo
Brasil	1966	9,5
Turquia	1978	5,4
Filipinas	1978	5,2
Estados Unidos	1983	0,26

Fonte: Krugman e Obstfeld (1994).

Mas isto não significa que qualquer intervenção deverá aumentar o bem-estar em países em desenvolvimento. Da mesma forma que políticas comerciais podem trazer benefícios sociais marginais não computados na análise de mercado, podem também introduzir novos custos. Argumenta-se que além das perdas de eficiência alocativa (distorções na produção e consumo), políticas protecionistas geram perdas de eficiência técnica e perdas pela menor diversidade de produtos e insumos. As perdas de eficiência técnica ocorrem quando a menor concorrência com produtos importados produz menor empenho gerencial das firmas locais por redução de custos e melhoria de qualidade (Moreira e Correa, 1996). A menor diversidade de produtos produz perda de bem-estar, se o consumidor valoriza a diversidade, e pela queda na produtividade pela falta de oferta de alguns bens de capital e insumos especializados. O Quadro 25.2 apresenta argumentos pró e contra à política comercial na presença de

QUADRO 25.2 ARGUMENTOS PRÓ E CONTRA POLÍTICA COMERCIAL NA PRESENÇA DE ECONOMIAS INTERNAS E EXTERNAS DE ESCALA

Pró**Contra**

1. Economias Internas de Escala: Necessidade de Investimento Inicial Elevado e Mercado de Capitais Inexistente ou Ineficiente

Proteção permite sobre preço que viabilizaria autofinanciamento das empresas

Proteção implica entrada excessiva de firmas, com operação abaixo da escala mínima competitiva.¹⁷

Melhor política: Medidas para atrair capital estrangeiro de longo prazo e desenvolvimento de mercado local de capitais de longo prazo.

2. Economias Externas Estáticas: País Anão produz, porém, detém Vantagens Comparativas Potenciais (baixos custos). Há produção no País B, com firmas competitivas internacionalmente

Uma firma que quisesse se instalar no País A apresentaria custo superior àquelas empresas que atuam no País B.

Melhor política: subsidiar a produção de forma decrescente em relação ao número de firmas é mais adequado para sanar o problema da a propriedade (firmas não conseguem ser integralmente compensadas pelos benefícios que produzem).

Proteção viabilizaria instalação de empresas no País A (desde que mercado suficientemente grande para abrigar número de firmas necessário para alcance de custo competitivo).

3. Economias Externas Dinâmicas: País A não produz, porém, detém Vantagens Comparativas Potenciais (baixos custos). Há produção em países desenvolvidos (pioneiros)

Produção nos países pioneiros apresenta baixo custo unitário, pois firmas já acumularam conhecimento, descendo boa parte da curva de aprendizagem.

Melhor política: subsídio à produção decrescente em relação à produção acumulada (indicador de aprendizagem).¹⁸

No país A, ausência de experiência implica custo unitário inicial muito elevado, constituindo obstáculo à implantação da indústria.

Proteção temporária como forma de viabilizar instalação da indústria no País A (Argumento da Indústria Nascente)

economias de escala interna e externa, exemplos de imperfeições de mercado, cuja principal implicação é que, na ausência de políticas para sua correção, as firmas produzirão e investirão aquém ou além do socialmente ótimo.

Os argumentos contrários à proteção se baseiam no princípio geral de que é sempre preferível tratar diretamente as falhas de mercado, evitando-se assim distorções não intencionais e custos adicionais decorrentes da política comercial. Há ainda duas linhas de argumentação contrárias às políticas comerciais: a primeira argumenta que não há métodos de análise, suficientemente precisos, para diagnosticar as falhas de mercado,¹⁴ de modo que se prescrevam políticas que assegurem maior bem-estar; a segunda argumenta que as políticas comerciais sofrem interferências de natureza política, que introduzem custos adicionais significativos (falhas de governo).¹⁵ Por fim, políticas comerciais podem gerar retaliação de outros países que se sintam prejudicados, gerando ainda novos custos.¹⁶

■ 25.5 Regulação Internacional – O Acordo da OMC e as Principais Regras sobre Comércio Internacional

O desejo de criar uma instituição que tratasse do comércio internacional vem desde o final da II Guerra Mundial. Porém, naquela época, não foi possível a sua criação, tendo sido firmado tão somente o Acordo Geral sobre Comércio e Tarifas (GATT, sigla em inglês). O GATT serviu de base para a realização de oito rodadas de negociação, possibilitando a redução do imposto de importação, notadamente dos países desenvolvidos. À medida que esses países reduziam as alíquotas do imposto de importação, intensificavam a utilização de medidas não tarifárias, gerando a necessidade de um maior detalhamento sobre as regras relativas a essas medidas. Adicionalmente, setores considerados como sensíveis pelos países desenvolvidos, tais como agricultura, têxteis e vestuário, foram de fato excluídos do âmbito das regras multilaterais, passando a ser protegidos por meio de restrições quantitativas e, no caso de produtos agrícolas, beneficiados por subsídios às exportações.

Ademais, as mudanças sofridas pela economia mundial fizeram com que outros temas, até então não cobertos pelo Acordo Geral (investimentos, propriedade intelectual e serviços) ganhassem relevância na agenda negociadora.

Assim, em 1986, foi lançada a Rodada Uruguai, considerada a mais importante das rodadas de negociação e da qual resultou a criação da Organização Mundial do Comércio. A Rodada Uruguai aprimorou as regras relativas ao comércio de bens tangíveis, resgatou os produtos da cadeia agrícola e da têxtil para o âmbito das regras multilaterais, e estabeleceu regras relativas aos novos temas (serviços, propriedade intelectual e investimentos).¹⁹ Destaca-se também, dentre os resultados dessa Rodada, o aprimoramento do procedimento de solução de controvérsias, que tornou mais eficaz o mecanismo para a resolução de disputas entre os países.²⁰

Ao se criar a OMC, estabeleceu-se que uma de suas funções²¹ seria servir de fórum de negociações entre os Membros, buscando-se um desenho no qual as negociações fossem realizadas, inicialmente, no âmbito dos Comitês que compõem aquela instituição, evitando-se, assim, a necessidade de realização de grandes rodadas de negociações. Este desenho, no entanto, não se mostrou eficaz, visto que gerava a polarização das negociações, impedindo ofertas cruzadas entre os diferentes temas (por exemplo, agricultura e serviços). Assim, após a tentativa frustrada de lançamento da Rodada do Milênio, deu-se início, em novembro de 2001, à Rodada de Doha, com prazo até janeiro de 2005 para sua conclusão. Entretanto, após quase oito anos, a Rodada permanece inconclusa e sem perspectiva, devido aos interesses divergentes diante de uma abrangente agenda negociadora.

A análise da evolução das Rodadas negociadoras mostra que se, por um lado, os países consideram importante o estabelecimento de regras que impeçam uma escalada protecionista e a transformação do comércio em um jogo de soma negativa; por outro, revela a crescente resistência dos países, à medida que boa parte da proteção foi sendo eliminada, em restringir a capacidade para atender interesses econômicos e políticos locais, ao limitar seus instrumentos remanescentes de política comercial.

A seguir, apresentam-se as principais regras atualmente vigentes relacionadas ao comércio de bens tangíveis.

25.5.1 PRINCÍPIOS FUNDAMENTAIS

Os princípios fundamentais do Acordo, presentes desde o GATT 1947, se referem à não discriminação. O princípio do “Tratamento Nacional” estabelece que um produto estrangeiro, após a sua internalização no país importador (isto é, após o pagamento do imposto de importação e taxas específicas), deverá receber o mesmo tratamento em relação a impostos e taxas, leis e regulamentos domésticos que as mercadorias produzidas internamente. Busca-se evitar, assim, a discriminação contra o produto estrangeiro, o que invalidaria concessões negociadas com os demais países signatários do Acordo.

O outro princípio fundamental refere-se à “Cláusula da Nação Mais Favorecida”, a qual estabelece que, quando um país dá a algum outro Membro um tratamento favorável, o mesmo tratamento deverá ser estendido aos demais países signatários do Acordo. Busca-se assegurar a igualdade de tratamento para todos os países Membros da OMC, independentemente de seu peso econômico. Esta cláusula, no entanto, comporta algumas exceções, dentre as quais se destacam: a possibilidade de formação de áreas de livre comércio e uniões aduaneiras e tratamento diferenciado para países em desenvolvimento (Sistema Geral de Preferências, por exemplo).

25.5.2 PRINCIPAIS REGRAS RELATIVAS AO COMÉRCIO DE BENS TANGÍVEIS

25.5.2.1 Imposto de Importação e Cotas Tarifárias

Conforme já mencionado, o acordo multilateral não implica qualquer compromisso de eliminação da proteção tarifária, nem existe uma forma predeterminada para a realização das reduções tarifárias. Os resultados das concessões (reduções tarifárias) de cada país são registrados em uma lista individual de compromissos. As tarifas registradas nessa lista são denominadas de “tarifas consolidadas” e representam um teto ao imposto de importação a ser cobrado pelo país.²²

Até a Rodada Uruguai, as consolidações de tarifa foram feitas principalmente pelos países desenvolvidos e se referiam basicamente a produtos não agrícolas. Um dos resultados alcançados pela Rodada Uruguai foi uma ampliação significativa dos compromissos assumidos pelos países, em termos de tarifas (Figura 25.1), em especial pelos países em desenvolvimento.

Em relação aos produtos agrícolas,²³ antes da Rodada Uruguai, muitos desses produtos eram protegidos pelos países desenvolvidos por meio de restrições quantitativas (cotas). Na Rodada Uruguai, houve a conversão dessas cotas para tarifas, o que ficou conhecido como “processo de tarificação”. Essas tarifas foram consolidadas e, em sua maioria, foram estabelecidas na forma de alíquotas específicas. No entanto, as tarifas resultantes desse processo eram tão elevadas que inviabilizavam as importações desses produtos. Assim, para garantir um acesso mínimo aos mercados desses países, foram estabelecidas cotas tarifárias.

É interessante notar que o imposto de importação de fato cobrado nos países desenvolvidos para a maioria dos produtos corresponde à alíquota consolidada. Já, no caso dos países em desenvolvimento, o imposto efetivamente aplicado, para a maioria das linhas tarifárias é inferior ao consolidado (Tabela 25.2). Assim, os países em desenvolvimento podem, a qualquer momento, aumentar o nível de proteção tarifária até o nível consolidado.

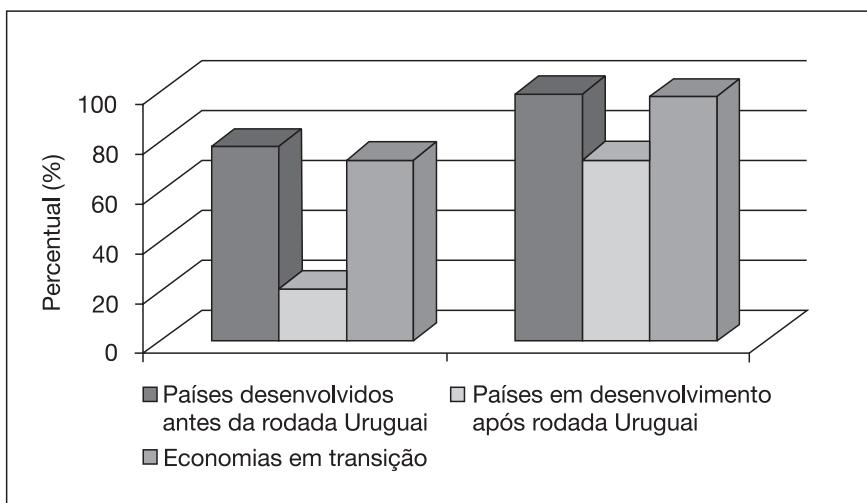


Figura 25.1 Linhas Tarifárias Consolidadas. Fonte: www.wto.org/apud “The Results of the Uruguay Round of Multilateral Trade Negotiations: Market Access for Goods and Services – Overview of the Results”, Geneva, 1994.

TABELA 25.2 Alíquotas Consolidadas e Aplicadas (Países Selecionados)

País / Bloco	Alíquota Média Simples					
	Todos Produtos		Produtos Agrícolas		Produtos não Agrícolas	
	Consolidada	Aplicada (2007)	Consolidada	Aplicada (2007)	Consolidada	Aplicada (2007)
Em Desenvolvimento						
Africa do Sul	19,1	7,8	40,8	9,2	15,7	7,6
Brasil	31,4	12,2	35,5	10,3	30,8	12,5
Chile	25,1	6,0	26,0	6,0	25,0	6,0
China	10,0	9,9	15,8	15,8	9,1	9,0
Índia	50,2	14,5	114,2	34,4	36,2	11,5
Desenvolvidos						
CE	5,4	5,2	15,1	15,0	3,9	3,8
Japão	5,1	5,1	22,7	21,8	2,4	2,6
Estados Unidos	3,5	3,5	5,0	5,5	3,3	3,2

Fonte: Elaboração Própria, com base em informações disponíveis em WTO & ITC (UNCTAD/WTO) (2008).

25.5.2.2 Salvaguardas

Para viabilizar o processo de liberalização, o GATT contém algumas cláusulas de salvaguarda, que se referem à possibilidade de aplicação de uma medida emergencial de proteção, de caráter temporário. As salvaguardas previstas no Art. XIX do GATT foram regulamentadas pelo Acordo sobre Salvaguardas (AS), apenas na Rodada Uruguai, e podem ser aplicadas se a produção doméstica de um bem sofrer prejuízo grave (ou ameaça de sofrê-lo) se, em decorrência de evolução imprevista e das obrigações assumidas no âmbito do Acordo da OMC, houve um surto de importações.²⁴

Não existe definição no Acordo sobre a forma de aplicação dessa medida. No entanto, as medidas até hoje aplicadas corresponderam a aumento do imposto de importação e aplicação de restrição quantitativa. Cabe ainda registrar que o Acordo proibiu a realização de “acordos de restrição voluntária” e “acordos de organização de mercado”. No entanto, conforme já mencionado, esses acordos continuam sendo “voluntariamente” estabelecidos.

A aplicação da medida de salvaguarda, regulamentada no AS, reflete um reconhecimento de que os produtores domésticos perderam competitividade e, por isso, não têm condições de conviver com o nível de proteção tarifária consolidada.²⁵ Assim, a aplicação dessa medida objetiva a preservação de uma indústria local e pode envolver a adoção de medidas por parte do governo para viabilizar a recuperação da produção doméstica.

A despeito do caráter emergencial da medida e das condições bastante restritas para sua aplicação, observa-se que a sua utilização tem sido relativamente intensa,²⁶ ainda que, para evitar o risco de retaliação por parte dos parceiros comerciais, a maioria aplicou a medida por um período não superior a três anos.

25.5.2.3 *Medidas Antidumping*

As medidas antidumping têm como objetivo proteger os produtores domésticos de importações objeto de dumping, originárias de um ou de alguns países.²⁷ A definição de dumping, estabelecida no Acordo da OMC, refere-se à mera discriminação de preços praticados no mercado interno (denominado de “valor normal”)²⁸ do país exportador e o preço de exportação, devidamente ajustados para se garantir comparação justa. Assim, existirá dumping se o valor normal for superior ao preço de exportação.²⁹ O cálculo da margem de dumping é realizado por empresa produtora/exportadora do(s) país(es) investigado(s) e a margem apurada determinará o limite da medida a ser aplicada. Para esse fim, são enviados questionários aos produtores/exportadores estrangeiros, com a intenção de apurar informações que permitam o cálculo individual da margem.

A determinação de dumping não é condição suficiente para a aplicação de medida de proteção, há que ser demonstrado, por meio da análise da evolução da situação dos produtores domésticos do produto similar ao importado (“indústria doméstica”), com base em diversos indicadores, que as importações investigadas estão causando problemas (perdas) para os produtores domésticos ou ameaçam causá-los.

Assim, pelo menos teoricamente, a medida antidumping teria um efeito menos distorcido sobre o comércio do que as tarifas. Em primeiro lugar, seria justificada pelo fato de evitar que os produtores domésticos de um país que promoveu a abertura de seu mercado sejam prejudicados pelo excesso de proteção conferido por outro país. Isto porque, quanto maior a proteção, maior a possibilidade de diferenciação de preços. Ademais, seria um instrumento seletivo – isto é, aplicado apenas a determinadas origens e limitada ao montante de dumping apurado por exportador. E, ainda, no caso de uma empresa estrangeira ser competitiva, a aplicação da medida, ainda que reduzisse a vantagem de preço que a empresa tem ao exportar, não seria necessariamente suficiente para impedir a importação.

No entanto, na prática, a realidade é bastante distinta, uma vez que o Acordo Antidumping ainda contém inúmeras ambiguidades e imprecisões que permitem que os governos dos países importadores atuem com bastante discricionariedade ao calcular o dumping e avaliar o eventual dano causado pelas importações investigadas.

Neste sentido, tendo em vista à existência de barreiras ao comércio, naturais e derivadas da proteção, que restringem a concorrência no mercado interno, é relativamente frequente a prática de dumping, tal como definido no Acordo, nas exportações. Quanto à determinação do dano, a análise, em virtude dos diversos indicadores a serem considerados, contempla um elevado grau de subjetividade. Ademais, observa-se, com certa frequência, a aplicação de medidas antidumping sobre importações de um produto, originárias de diversos países, fechando, assim, de fato, a economia às importações de um determinado bem.

Adicionalmente, ainda que as medidas tenham prazo de vigência de cinco anos, a introdução no Acordo da possibilidade explícita de sua prorrogação por igual período sem restrições, condicionada à realização de revisão, permite que uma medida antidumping seja mantida indefinidamente.

À medida que o processo de redução tarifária foi avançando, sob a égide do GATT, observou-se a intensificação da utilização das medidas antidumping. Os principais aplicadores deste instrumento, no período pré-OMC, foram os países desenvolvidos, visto serem estes os que estavam de fato implementando cortes dos níveis tarifários. Este cenário se modificou após a Rodada Uruguai, com os países em desenvolvimento passando a ter um papel de destaque na aplicação dessas medidas. A intensificação da utilização das medidas antidumping pelos países em desenvolvimento (Figura 25.2) ocorreu não apenas em função dos compromissos por eles assumidos em termos de consolidação tarifária e delimitação da utilização de restrições quantitativas, como também pelo fato de muitos desses países terem liberalizado as importações (inclusive, conforme já comentado, com tarifas aplicadas inferiores aos níveis consolidados).³⁰

O Brasil, de 1995 a 2008, aplicou 86 medidas, ocupando a nona posição dentre os Membros da OMC aplicadores dessas medidas. As origens mais afetadas pelas medidas aplicadas pelo Brasil, nesse período, foram China e EUA. Por outro lado, o Brasil também se encontra dentre os Membros mais afetados pela aplicação dessas medidas, 74 medidas, ocupando a décima posição, sendo a Argentina a responsável por quase metade dessas medidas (31). Os setores químico e metalúrgico foram os que apresentaram maior relevância tanto em termos de medidas aplicadas pelo Brasil, como de medidas aplicadas contra exportações brasileiras.³¹

A proliferação da utilização das medidas antidumping, como mecanismo de proteção, pode inibir a estratégia das empresas exportadoras que investem mais agressivamente na redução de custos e melhoria de qualidade a fim de penetrar no mercado externo. Isto pode representar uma redução nas taxas de inovação e crescimento globais.

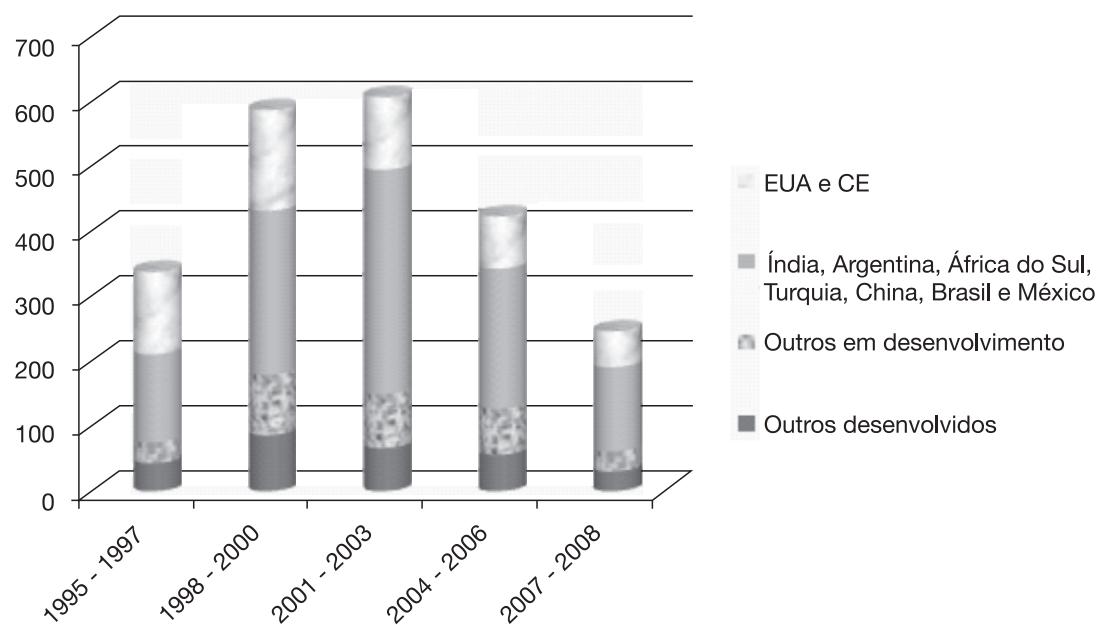


Figura 25.2 Aplicação de medidas antidumping por grupo de países. Fonte: Elaboração Própria, a partir de dados disponíveis no site da OMC (www.wto.org).

25.5.2.4 Medidas de Proteção contra Subsídios Concedidos por Outros Países

O Acordo sobre Subsídios e Medidas Compensatórias (ASMC), da OMC, estabelece regras para que um país possa defender seus interesses no caso de ter sua produção doméstica afetada pela concessão de subsídios.³²

Os países que se sintam prejudicados podem entrar com uma ação na OMC contra o Membro que concede subsídios, cujos prazos para resolução e ônus da prova do país reclamante irão variar em função da categoria do subsídio objeto da reclamação.³³ As categorias estabelecidas para esse fim são: proibidos,³⁴ acionáveis³⁵ e não acionáveis³⁶.

Esta classificação reflete a percepção do grau de distorção ao comércio gerado pelo subsídio. Aqueles que afetam diretamente o comércio, por estimularem as exportações ou restringirem as importações, são considerados proibidos; ao passo que os acionáveis, por serem específicos, podem gerar distorções (mas não obrigatoriamente geram) pelo fato de alterarem as condições de funcionamento do mercado, mudando a alocação de recursos e estimulando a produção de determinados produtos. Já os subsídios de disponibilidade geral, por não afetarem, em princípio, a alocação de recursos, não gerariam efeitos distorcidos sobre o comércio.

Quando a concessão de subsídio específico (proibidos ou não) por um país implicar dano aos produtores domésticos de outro país, em decorrência de importações do produto beneficiado pelo subsídio, o governo do país importador poderá, alternativamente à ação perante a OMC, aplicar medida compensatória. Para esse fim, deverá seguir procedimento similar ao previsto para aplicação de medidas antidumping (isto é, iniciar investigação, dar oportunidade para que outras partes interessadas que não os produtores domésticos apresentem informações e manifestações e determinar o montante de subsídio específico e a existência de dano causado pelas importações do produto beneficiado pelo(s) subsídio(s) específico(s)).

A utilização de medidas compensatórias é bem menos frequente que a de medidas antidumping. De 1995 a 2008, foram aplicadas apenas 128 medidas compensatórias por todos os países Membros da OMC, tendo o Brasil aplicado apenas 7 medidas.

Em relação às regras previstas no ASMC, cabe destacar os seus impactos sobre a formulação da política industrial e seus resultados. Conforme Lemme et al. (2007), o Acordo refletiria a concepção de que políticas horizontais (envolvendo a concessão de subsídios de disponibilidade geral) seriam aceitáveis, por não gerarem efeitos distorcidos sobre o comércio, não sendo, portanto, passíveis de ações perante a OMC e nem sujeitas à aplicação de medidas compensatórias. Já no caso de políticas seletivas, no entanto, envolvendo subsídios específicos (acionáveis), a análise das implicações das regras estabelecidas no âmbito da OMC seria mais complexa. Se a política for implementada para gerar resultados rápidos, outros países podem questioná-la, tendo menor dificuldade para demonstrar a ocorrência dos efeitos adversos dela decorrentes na OMC. Porém, o maior risco, no caso da política implicar aumento de competitividade e, por consequência, aumento

das exportações, é que essas exportações venham a ser obstaculizadas pela imposição de medidas compensatórias. Já no caso de políticas envolvendo subsídios proibidos, deve ser considerado que tanto o seu questionamento no âmbito da OMC implicaria menor ônus da prova para o país reclamante, facilitando, assim, uma “condenação” da política adotada e a retirada do subsídio.³⁷ No caso de subsídios às exportações, os seus efeitos poderiam ser facilmente contrarrestados por meio da aplicação de medidas compensatórias, ou mesmo de medidas antidumping, visto que a concessão desse tipo de subsídio implica preço superior no mercado interno (supostamente protegido) em relação ao preço no mercado internacional.

Cabe, por fim, registrar, no que se refere às regras para subsídios, o tratamento diferenciado conferido para os produtos agrícolas. A criação do Acordo sobre Agricultura (AA), no âmbito da Rodada Uruguai, representou uma solução de compromisso: por um lado, foi possível trazer o comércio agrícola para o âmbito da regulação multilateral; em contrapartida, os elevados subsídios concedidos pelos países desenvolvidos foram referendados, inclusive os subsídios à exportação, ainda que sujeitos a compromisso de redução. Acordou-se que, por um prazo de nove anos, os subsídios agrícolas concedidos em conformidade com os compromissos de redução assumidos no âmbito do AA, não poderiam ser objeto de ações perante a OMC³⁸ e somente poderiam ser objeto de medidas compensatórias, caso as investigações para esse fim, fossem iniciadas com a “devida moderação”, expressão suficientemente vaga para contemplar as mais distintas interpretações.

25.5.2.5 Regras sobre Outras Medidas que Afetam o Comércio de Bens

O Acordo da OMC contempla ainda regras sobre distintos instrumentos, que, ainda que justificáveis, podem ser usados de maneira arbitrária e protecionista, para burlar compromissos de liberalização assumidos. Destacam-se, neste grupo, os Acordos sobre Barreiras Técnicas ao Comércio e o Acordo sobre Medidas Sanitárias e Fitossanitárias, que reconhecem a legitimidade de os governos aplicarem medidas para assegurar proteção da vida ou saúde humana, animal ou vegetal, do meio ambiente, mas que buscam assegurar que essas medidas não sejam aplicadas de maneira arbitrária e que, por essa razão, venham a se constituir em uma restrição ao comércio internacional.³⁹

O Acordo da OMC também estabelece regras relativas aos procedimentos de importação – tais como, licenciamento de importações e valoração aduaneira (determinação da base de cálculo do imposto de importação), entre outros, visando a evitar que os governos criem obstáculos administrativos injustificados às importações.

25.6 Acordos Regionais

Ainda que as regras multilaterais estabeleçam o princípio de não discriminação, abre-se uma exceção para a realização de acordos comerciais regionais. Para os acordos de livre comércio, requer-se que “substancialmente todo” o comércio de bens esteja isento de tarifa entre os países-membros dentro de um período razoável de tempo (segundo entendimento sobre a Interpretação do Artigo XXIV do GATT 1994, parágrafo 3, esse período não deveria exceder dez anos exceto em circunstâncias excepcionais). Para as uniões aduaneiras, requer-se adicionalmente que a proteção média conferida pelos países integrantes da união aduaneira para terceiros países não supere o grau de proteção vigente antes da implementação do acordo.

Os acordos de livre comércio têm como objetivo a liberalização do fluxo de comércio de bens entre os países signatários do acordo, por meio da eliminação das tarifas aplicadas ao comércio intrazona. Nos acordos de livre comércio, os países signatários continuam livres para estabelecer sua política comercial em relação a terceiros países. Para os acordos firmados a partir dos anos de 1990 foram incluídas regras relativas à liberalização do comércio de serviços, a investimentos e propriedade intelectual.

Já nas uniões aduaneiras, ademais de se buscar a liberalização do comércio intrazona, os países membros adotam uma política comercial comum em relação a terceiros países (por exemplo, uma tarifa externa comum, regras para aplicação comum de medidas antidumping, compensatórias e de defesa comercial).

Existem outros tipos de acordos que preveem uma maior integração dos países: o mercado comum, no qual, adicionalmente aos elementos que caracterizam a união aduaneira, se prevê a liberalização dos fluxos de trabalho e capital, e a união econômica e monetária, que inclui o estabelecimento de moeda única.

Assim, quanto maior o grau de integração entre os países, maior a cessão de soberania requerida. Por exemplo, no caso das uniões aduaneiras, quando comparadas às áreas de livre comércio, os países deixam de ter autonomia para determinar a sua política comercial, que passa a ser estabelecida em conjunto pelos países integrantes da união. Adicionalmente, o aprofundamento da integração comercial e econômica entre os países implica a necessidade de harmonização de suas políticas internas, visto que políticas distintas podem ter implicações significativas sobre a alocação dos recursos intrazona.

É justamente por implicar menor restrição aos países na gestão de sua política comercial e políticas internas que a maioria dos acordos negociados ou em negociação se referem a acordos de livre comércio ou acordos de alcance parcial (estes são acordos nos quais se negocia uma dedução da tarifa aplicada para terceiros países).⁴⁰

Observa-se, a partir dos anos de 1990, uma proliferação de acordos regionais/bilaterais. Embora esses acordos sejam um passo na direção da liberalização do comércio, em função da multiplicidade de regras que vão sendo estabelecidas, podem acabar gerando obstáculos ao comércio. Mais especificamente, o fato de os países firmarem diversos acordos com distintos parceiros gera, muitas vezes, a adoção de regras e critérios diferenciados para os diversos países. Com isso, dificulta-se a percepção do custo de importação relativo para cada uma das possíveis origens do produto a ser adquirido, em especial durante o período de implementação.

Adicionalmente, há que se considerar que os acordos firmados entre grupos de países podem gerar efeitos positivos (criação de comércio)⁴¹ e negativos (desvio de comércio)⁴² e que, nem sempre, o resultado líquido auferido por um acordo é positivo.

25.6.1 A EXPERIÊNCIA RECENTE BRASILEIRA

Em 1991, o Brasil assinou o Tratado de Assunção com Argentina, Uruguai e Paraguai, que previa o estabelecimento de um mercado comum, o Mercosul, em 1994, com livre circulação de bens, serviços e fatores produtivos, e uma política comercial comum. Embora tenham ocorridos avanços no processo de integração, com liberalização da maior parte do fluxo de bens entre os países, ainda hoje existem barreiras não tarifárias aplicadas no comércio intrazona. No que se refere à política comercial comum, somente foi possível o estabelecimento de uma Tarifa Externa Comum (TEC), mas que ainda comporta muitas exceções, permitindo-se aos países integrantes do bloco aplicar alíquotas distintas daquelas previstas na TEC para uma ampla gama de produtos. Além disso, ainda se observa a “dupla cobrança” da TEC sobre produtos importados de fora da região. Se um produto é importado pelo Brasil de fora da região, estando sujeito à cobrança da TEC, e depois reexportado para um país do bloco, será novamente cobrado o imposto de importação. Isto implica que os países não se beneficiam de uma das principais vantagens da união aduaneira sobre acordos de livre comércio, que é evitar os custos associados ao exame das regras de origem. Por essas razões, é comum a referência ao Mercosul como uma “união aduaneira imperfeita”.

Além do Mercosul, Brasil, como parte integrante deste bloco, também firmou acordos de livre comércio com Chile, Bolívia, México, Peru, Colômbia, Equador e Venezuela, a partir da segunda metade dos anos 1990. Fora do continente sul-americano, foram também estabelecidos acordos de preferências tarifárias com a Índia, vigente a partir de junho de 2009, com Israel e com a União Aduaneira da África do Sul, que ainda não entraram em vigor.

Notas

1. Nos Estados Unidos, a base para a incidência da tarifa é o valor f.o.b. (*free on board*) da mercadoria, o que significa o valor da mercadoria já embarcada no país de origem. A prática no Brasil é tomar, como base de cálculo para a tarifa de importação, o valor c.i.f. (*cost, insurance and freight*) do bem importado, o que significa incluir os custos de fretes e seguros até o ponto de entrada no país.
2. Por outro lado, pode induzir ao investimento estrangeiro direto no país importador, como foi o caso da indústria automobilística japonesa nos anos 1980 (Chami Batista, 2008).
3. O *Buy American Act* de 1933 (e emendas posteriores) restringe a compra de suprimentos e materiais de construção por agências de governo aos produtos definidos como bens finais produzidos domesticamente (*domestic end-products*). Esses bens têm que ser manufaturados nos Estados Unidos e utilizar componentes domésticos com um peso superior a 50% do custo de todos os componentes (Chami Batista, 2002).
4. Essa prática do exame de similar nacional começou nos anos 1950 no Brasil, mas o Decreto Lei n. 37 de 1966 regulou a prática e ficou conhecido como a “Lei do Similar Nacional”.
5. O texto original da Lei n. 8.248, de 23 de outubro de 1991, conhecida como Lei de Informática, não adotou o conceito de PPB como contrapartida aos seus incentivos fiscais, os quais eram concedidos aos bens de informática e automação fabricados no País com níveis de valor agregado local compatíveis com as características de cada produto. Com a publicação da Lei n. 10.176, de 2001, finalmente o PPB foi incorporado à legislação como contrapartida aos benefícios fiscais da Lei de Informática, somado à obrigatoriedade, já existente, de aplicação de recursos financeiros em Pesquisa e Desenvolvimento (P&D). Disponível em: <<http://www.mdic.gov.br/sitio/interna/interna.php?area=2&menu=1103>> Acesso em 30/07/2012.
6. Esse coeficiente é obtido por meio de uma matriz insumo-produto da economia e a taxa de proteção efetiva é igual a: $e = (n-ab)/(1-a)$, em que ‘e’ é a taxa de proteção efetiva, ‘n’ é a tarifa nominal sobre o produto, ‘b’ é a tarifa nominal sobre os insumos, e ‘a’ é o percentual dos insumos importados.
7. A tarifa efetiva da indústria de automóveis, caminhões e ônibus foi em média de 159% entre 1995 e 1998, chegando a atingir 217,5% em 1996 (Kume, Piani e Souza, 2003).
8. O indicador para o Brasil (2006) foi de 21,9% comparado com: 15,5% Argentina (2006), 9,4% Chile (2006), 10,1% China (2006), 21,7% Índia (2005), 6,6% Indonésia (2006), 23% Malásia (2006), 27,2% México (2006), 19,4% Rússia (2005), 4,1% Turquia (2006).
9. Abreu (2007) faz uma análise cuidadosa da economia política da política comercial brasileira. Ferreira e Facchini (2005) estudam a relação entre a estrutura de proteção e a estrutura industrial no Brasil no período 1988 e 1994. Ver também Oliveira (2009).

10. Os instrumentos de política comercial não são vistos, teoricamente, como os mais adequados para melhorar a distribuição funcional da renda. Metas exógenas são difíceis de racionalizar, exceto quando elas têm como base considerações de natureza política, estratégica, ou moral, que as colocam no primeiro lugar das prioridades de uma ordem de preferência social lexicográfica (Dixit e Norman, 1980).
11. Em competição imperfeita há exceções em que cotas de importação, por exemplo, geram resultados superiores para o país que a tarifa (Helpman e Krugman, 1989).
12. A existência de desemprego ou subemprego persistente de mão de obra e de um mercado de capital e/ou de trabalho incapazes de transferir recursos para setores com grande potencial de rendimentos elevados também seriam sinais de falhas de mercado. O dualismo econômico, presente em países menos desenvolvidos, caracterizado por grandes diferenciais de salários entre trabalhos semelhantes em diferentes setores, seria um dos sinais de que os mercados não estão funcionando bem. Em algumas análises, contudo, a própria política comercial pode agravar a economia dual (Krugman e Obstfeld, 1994).
13. No seu clássico livro sobre política comercial, Corden (1974) distingue bem a defesa do livre-comércio da defesa do *laissez-faire*. É perfeitamente possível defender intervenções de políticas econômicas para, por exemplo, melhorar a distribuição da renda, mas acreditar que, *grosso modo*, o livre-comércio é a melhor alternativa.
14. Por exemplo, para confirmar a suposição de que o país sob análise possui vantagem comparativa potencial na indústria com economias externas, seria necessário estimar o efeito das economias externas estáticas sobre o custo unitário a cada nível de produção no país que serve de referência (País B) e projetar esse mesmo efeito para a indústria a ser desenvolvida no país A. No caso de economias externas, o cálculo seria ainda mais complexo, pois seria necessário projetar a evolução ao longo das curvas de aprendizagem, tanto para os pioneiros como para o país A. Em ambos os casos de economias externas, deveria ser projetada a possível reação das firmas já estabelecidas no país que se beneficia dessas economias. Além de estudar os efeitos da política comercial sobre o bem-estar para cada setor isoladamente (análise de equilíbrio parcial), seria necessário analisar os efeitos das diferentes alternativas para a estrutura de impostos, para justificar maior proteção ou subsídios a determinadas indústrias em detrimento de outras (enfoque de equilíbrio geral).
15. Embora os benefícios sociais da decisão de liberalizar o comércio ultrapassem em muito os seus custos, esses benefícios são divididos por um grande número de pessoas, ao passo que os custos são concentrados em um pequeno grupo. Assim, o grupo “perdedor” se mobiliza para evitar a abertura de mercado, ensejando atividades políticas para capturar o Estado e favorecer seus interesses (*rent seeking*). Devemos a Olson (1965) a ideia de que a atividade política em função de um grupo é um bem público e a identificação do problema que ele denominou ação coletiva: embora haja interesses do grupo como um todo em pressionar por políticas favoráveis, não há nenhum interesse individual em fazer o mesmo.
16. O uso de teoria dos jogos é necessário para o exame adequado das possibilidades de retaliação e suas consequências.
17. A proteção por meio de instrumentos quantitativos, como uma cota de importação, por exemplo, incentivam o conluio das empresas locais.
18. Nos modelos de economias abertas ao comércio com crescimento endógeno (a taxa de progresso tecnológico é determinada pelo modelo), a presença de economias externas dinâmicas no processo de inovação implica que intervenções de políticas no mercado de pesquisa e desenvolvimento podem aumentar o bem-estar. Esse resultado requer, no entanto, que as externalidades sejam país-específicas (os transbordamentos tecnológicos não ultrapassem as fronteiras do país), e a produtividade em pesquisa seja constante. Além disso, se há externalidades positivas e negativas simultaneamente (como no caso dos modelos de melhoria de qualidade dos produtos), a recomendação pode ser tanto um subsídio à atividade de pesquisa e desenvolvimento (caso as externalidades positivas sejam dominantes), ou um imposto sobre essa atividade (no caso oposto). Ver Grossman e Helpman (1991).
19. Os seguintes acordos foram criados na Rodada Uruguai: o Acordo Geral sobre Comércio de Serviços (GATS); o Acordo sobre Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (TRIPS), sigla em inglês; e o Acordo sobre Medidas de Investimento Relacionadas ao Comércio (TRIMs), sigla em inglês.
20. O procedimento de solução de controvérsias estabeleceu prazos para a condução do processo, assim como a possibilidade de recursos às avaliações realizadas pelos painéis, por meio do estabelecimento do Órgão de Apelação. Adotou também a regra de “consenso negativo”, isto é, as recomendações emanadas dos painéis (ou das recomendações dos painéis revistas pelo Órgão de Apelação) somente podem ser rejeitadas se houver consenso dos países neste sentido, garantindo assim a adoção dessas recomendações, distintamente do que se observava no período pré-OMC, quando as recomendações somente eram adotadas se todos os países concordassem (regra do “consenso positivo”).
21. As outras funções da OMC são: facilitar a implementação, administração e operação dos Acordos; administrar o Procedimento de Solução de Controvérsias; e administrar o Mecanismo de Revisão de Políticas Comerciais (TPRM, sigla em inglês).
22. Os países são livres para reduzir o imposto de importação. Porém, para aumentá-lo acima do nível consolidado, terão de entrar em negociações com os principais países fornecedores do produto de que se trate e, se for o caso, dar uma compensação, pelo “rompimento do acordo”.
23. Produtos agrícolas, na definição da OMC, referem-se aos produtos cobertos pelo Acordo sobre Agricultura.
24. Para aplicação de medida de salvaguarda, o governo do país importador deve conduzir uma investigação, dando oportunidade para que importadores e produtores/exportadores estrangeiros se manifestem e apresentem informações. Além disso, deve ser feita notificação à OMC e ser dada a oportunidade para a realização de consultas com os principais países fornecedores do produto em questão. A medida poderá ser aplicada por um prazo de quatro anos, prorrogáveis por igual período, no caso de países desenvolvidos. Para os países em desenvolvimento, o período máximo de aplicação da medida é de dez anos. Na hipótese da medida vigorar por mais de três anos, negociações com os principais países fornecedores deverão ser realizadas com o intuito de se discutir uma compensação, em virtude do rompimento temporário de uma concessão.
25. Por essa razão, o Acordo prevê que, no caso da medida durar mais de 12 meses, esta deverá ser gradualmente liberalizada, com o objetivo de facilitar o ajuste da produção doméstica e, caso dure mais de três anos, o governo que aplicou a medida deverá revê-la na metade de sua vigência para determinar se a medida continua a ser necessária (ou se a sua liberalização pode ser acelerada). No caso da legislação brasileira, estabeleceu-se a obrigação dos produtores, que solicitam a aplicação da medida, apresentarem um compromisso de ajustamento, que indique as medidas a serem implementadas para permitir a recuperação da competitividade da produção doméstica, considerando o grau de proteção acordado. No caso da legislação norte-americana, está prevista a possibilidade de aplicação de medida para que, entre outros objetivos, seja possível o “desmonte” da indústria afetada e sua reconversão.
26. Conforme informação disponível no site da OMC (acesso em 26/07/09), de março de 1995 até novembro de 2008, foram aplicadas 89 medidas, sendo os principais países aplicadores a Turquia (11 medidas), a Índia (nove medidas), o Chile (sete medidas), EUA e Jordânia (ambos com seis medidas). O Brasil até julho de 2009 aplicou duas medidas de salvaguarda sobre importações de: (1) brinquedos, medida que vigorou por dez anos (1996 a 2006); e

- (2) coco ralado, aplicada em 2002, prorrogada em 2006 e com prazo de vigência previsto até 2010. Em setembro de 2008, foi iniciada investigação para aplicação de medidas sobre “mídias ópticas graváveis”, que não foi ainda concluída.
27. Assim como no caso de medidas de salvaguarda, para aplicação de medidas antidumping, o governo do país importador deve realizar uma investigação, normalmente solicitada pelos produtores domésticos que estão se sentindo afetados pelas importações a baixos preços para determinar a existência de dumping e de dano aos produtores domésticos causado pelas importações objeto de dumping.
 28. Em determinadas condições (inexistência de vendas para o mercado interno do país exportador, ou vendas em volume muito limitado), o “valor normal” poderá ser determinado com base no valor construído (custo de produção, despesas e margem razoável de lucro) ou no preço de exportação para terceiro país.
 29. Para fins de determinação do valor normal, vendas abaixo do custo para o mercado interno, se realizadas em volume significativo, entre outras condições, poderão ser desconsideradas para efeito de determinação do valor normal.
 30. A relação existente entre a intensificação da utilização do instrumento e o processo de liberalização comercial fica bastante evidente quando se observa a China. Ao entrar na OMC, em 2001, aquele país assumiu uma série de compromissos de liberalização da sua economia, os quais começaram a ser implementados em 2002. E, a partir daí, a China transformou-se em um dos principais aplicadores de medidas antidumping. Em sete anos (2002 a 2008), aplicou 108 medidas antidumping, sendo o sétimo principal aplicador de medidas antidumping, dentre os países membros da OMC, considerando o período 1995 a 2008.
 31. www.wto.org
 32. Segundo a definição do ASMC, subsídios são definidos como contribuição financeira do governo, real ou potencial, que implique benefício. Os subsídios são classificados em duas grandes categorias: (1) específicos, isto é, quando sua concessão está restrita a determinados produtos, indústrias, empresas e/ou regiões; ou (2) disponibilidade geral.
 33. O Brasil, por exemplo, acionou o Canadá na OMC por conta dos subsídios conferidos a Bombardier (empresa produtora de aeronaves civis), em 1997 e 2001. Também foi acionado pelo Canadá em função dos subsídios à exportação concedidos para a Embraer, em 1996.
 34. São considerados como proibidos os subsídios à exportação (isto é, aqueles cuja concessão esteja condicionada ao desempenho exportador) e subsídios que discriminem contra o produto importado (isto é, aqueles cuja concessão esteja condicionada à utilização do produto doméstico em detrimento do produto estrangeiro). Os subsídios proibidos são, por definição, considerados como subsídios específicos. No caso de subsídios proibidos, o país reclamante deve demonstrar apenas a existência do subsídio que se enquadre nesta categoria e os prazos para a conclusão do caso são mais céleres. Se for considerado que de fato é concedido um subsídio proibido, o país reclamado deverá eliminar o subsídio em questão em prazo a ser acordado, ou não sendo possível, em prazo a ser arbitrado.
 35. Os subsídios acionáveis se referem aos demais subsídios específicos. No caso de reclamação relativa a subsídio acionável, além de demonstrar a concessão de subsídio específico pelo país reclamado, o reclamante deverá demonstrar os efeitos adversos que o subsídio está causando para os seus interesses (quer seja em seu mercado interno, quer seja no mercado do país que concede o subsídio ou em terceiro mercado, ou ainda, no mercado internacional, por meio de efeitos de depressão de preços). Na hipótese de ser determinado que o subsídio é específico e está causando prejuízo aos interesses do país reclamante, não necessariamente o subsídio deverá ser eliminado, podendo ser adotadas medidas para anular os efeitos adversos causados pelo subsídio.
 36. Os subsídios não acionáveis correspondem aos subsídios de disponibilidade geral. Os países que pretendam adotá-los têm a obrigação de notificar à OMC previamente, dando, assim, a oportunidade para que outros países solicitem ao Secretariado uma avaliação sobre se o subsídio notificado de fato se enquadra na categoria “não acionável”. Se confirmado que o programa notificado é “não acionável”, nenhuma ação contra o país que o adota pode ser tomada.
 37. Cabe registrar a existência de interpretações distintas sobre o significado exato da expressão “retirada do subsídio – withdrawal of the subsidy”.
 38. Ainda durante a vigência da Cláusula de Paz, o Brasil iniciou duas ações na OMC sobre subsídios agrícolas: a controvérsia do açúcar contra a União Europeia e a do algodão contra os Estados Unidos. Nos dois casos, os países reclamados foram condenados.
 39. Esses acordos recomendam a harmonização das medidas adotadas pelos países Membros e, para esse fim, a adoção de medidas que sejam embasadas em normas aceitas internacionalmente. Estabelecem, adicionalmente, a obrigatoriedade de notificações à OMC relativas à criação de novas medidas ou alteração daquelas já existentes, para garantir a transparência das exigências estabelecidas pelos países que devem ser cumpridas não apenas pelos produtos importados, mas também pelos produtos domésticos.
 40. “If we take into account RTAs which are in force but have not been notified, those signed but not yet in force, those currently being negotiated, and those in the proposal stage, we arrive at a figure of close to 400 RTAs which are scheduled to be implemented by 2010. Of these RTAs, free trade agreements (FTAs) and partial scope agreements account for over 90%, while customs unions account for less than 10%.” Disponível em: <www.wto.org>. Acesso em 30/07/2012.
 41. Entende-se que ocorre criação de comércio quando um fornecedor local é substituído por um fornecedor do outro país com o qual o acordo foi firmado.
 42. Ocorre desvio de comércio quando um fornecedor estrangeiro, de um terceiro país, é substituído por um fornecedor do país com o qual se firmou o acordo, sendo a substituição decorrente da redução de tarifa promovida pelo acordo firmado.

Bibliografia

- Abreu, M. P. (2007) *Comércio Exterior: Interesses do Brasil*, Rio de Janeiro: Elsevier Editora Ltda.
- Chami Batista, J. (2002) Potencial para o Brasil das compras governamentais nos países do hemisfério: elementos para demandas brasileiras sobre o tema no âmbito da ALCA, Relatório Final Para o Ministério das Relações Exteriores, *mimeo*, julho.
- Chami Batista, J. (2008) Competing for the U.S. Import Market: NAFTA and Non-NAFTA Countries, em McKinney and Gardner (ed) *Economic Integration in the Americas*, Routledge Publishing Co.
- CNI e INMETRO (2002) Barreiras Técnicas às Exportações: o que são e como superá-las, CNI-Confederação Nacional da Indústria e INMETRO-Instituto Nacional de Metrologia, Normalização e Qualidade Industrial.

- Corden, W. M. (1974) *Trade Policy and Economic Welfare*, Oxford: Clarendon Press.
- Dixit, A.; Norman, V. (1980) *Theory of International Trade*, Cambridge University Press.
- Ferreira, P. C.; Facchini, G. (2005) Trade Liberalization and industrial concentration: Evidence from Brazil, *The Quarterly Review of Economics and Finance*, v. 45, p. 432-446.
- Grossman, G.; Helpman, E. (1991) *Innovation and Growth in the Global Economy*, The MIT Press.
- Helpman, E.; Krugman, P. (1989) *Trade Policy and Market Structure*, Cambridge e London: The MIT Press.
- Krugman, P. (1987) Is Free Trade Passé? *The Journal of Economic Perspectives*, v. 1, n. 2, p. 131-144, 1987.
- Krugman, P. (1994) The Myth of Asia's Miracle, *Foreign Affairs*, November.
- Krugman, P.; Obstfeld, M. (1994) *International Economics Theory and Policy*, 3rd edition, HarperCollins College Pub.
- Kume, H.; Piani, G.; de Souza, C. F. B. (2003) A Política Brasileira de Importação no Período 1987-1998: Descrição e Avaliação, in: Corseuil, C. H.; Kume, H. (Org.). *A abertura comercial brasileira nos anos 1990: impactos sobre emprego e salário*. 1. ed. Rio de Janeiro: Ipea, v. 1.
- Lemme, M. C.; Naidin, L.; Gadelha, M. F. (2007) Subsídios em Doha: Propostas de Negociação e Implicações para os Interesses Brasileiros, in: Veiga, P. M. (org.) (2007) *Comércio e Política Comercial no Brasil*. 1. ed. São Paulo: Singular.
- Moreira M. M.; Correa, P. (1986) Abertura Comercial e Indústria: O que se pode esperar e o que se vem obtendo, *Texto para Discussão* n. 49, BNDES.
- Olson, M. (1965) *The Logic of Collective Action*, Cambridge: Harvard University Press.
- OMC (2003) *Comprendre l'OMC*, Genéve, OMC.
- Sampaio Oliveira, G. A. (2009) The Political Economy of Brazilian Trade Policy: Domestic and International Determinants – Empirical Testing and Data Investigation, XXXVII Encontro Nacional de Economia, ANPEC, Foz do Iguaçú, dezembro.
- Stiglitz, J. E. (1996) Some Lessons from the East Asian Miracle, *The World Bank Research Observer*, v. 11, n. 2, August.
- WTO & ITC (UNCTAD/WTO) (2008) *World Tariffs Profile 2008*, Switzerland, WTO Secretariat.

Política Ambiental

Maria Cecília J. Lustosa e Carlos Eduardo F. Young

26.1 Introdução

A Política Ambiental é o conjunto de metas e instrumentos que visam reduzir os impactos negativos da ação antrópica – do homem – sobre o meio ambiente. Como toda política, possui justificativa para sua existência, fundamentação teórica, metas, instrumentos e prevê penalidades para aqueles que não cumprem as normas estabelecidas. Interfere nas atividades dos agentes econômicos e, portanto, a maneira pela qual é estabelecida influencia as demais políticas públicas, inclusive as políticas industrial e de comércio exterior. Por outro lado, as políticas econômicas favorecem um tipo de composição da produção e do consumo que tem impactos importantes sobre o meio ambiente.

A importância da Política Ambiental tem sido crescente, principalmente nos países industrializados, e seus efeitos sobre o comércio internacional podem ser percebidos com o surgimento de barreiras não tarifárias. Como cada país possui problemas ambientais específicos, há diferenças nos princípios e tipos de instrumentos de Política Ambiental adotados, mas há traços gerais que são comuns a todos os países.

O objetivo deste capítulo é apresentar, na próxima seção, as principais razões para a adoção de uma Política Ambiental. Em seguida, discutimos as possíveis soluções econômicas para os problemas ambientais e, posteriormente, apresentamos os principais instrumentos de Política Ambiental. Na Seção 26.5 são analisadas as relações entre Política Ambiental e comércio internacional. Finalmente, descrevemos as práticas de Política Ambiental no Brasil. Ao final do capítulo há um guia de leitura contendo algumas referências básicas sobre o tema.

26.2 Razões para Adoção da Política Ambiental

A Política Ambiental é necessária para induzir ou forçar os agentes econômicos a adotarem posturas e procedimentos menos agressivos ao meio ambiente, ou seja, reduzir a quantidade de poluentes lançados no ambiente e minimizar a depleção dos recursos naturais.¹ No caso das indústrias, os recursos naturais são transformados em matérias-primas e energia, gerando impactos ambientais iniciais (desmatamento, emissões de gases poluentes, erosão de solos, entre outros). As matérias-primas e energia são os insumos da produção, tendo como resultados o produto final e os rejeitos industriais – fumaça, resíduos sólidos e efluentes líquidos. Após o consumo do produto, temos o lixo como resultado final de todo processo de produção e consumo (Figura 26.1). Como os recursos naturais utilizados nos processos industriais são finitos, e muitas vezes não renováveis, a

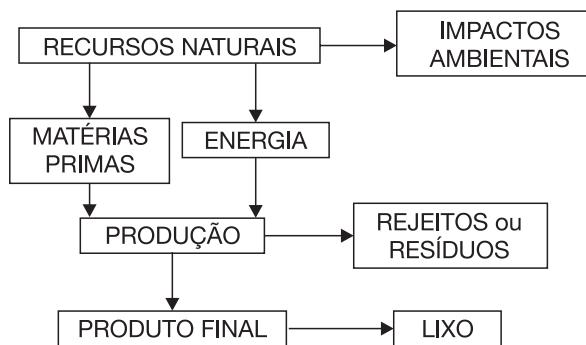


Figura 26.1 O processo de produção e a geração de impactos ambientais.

utilização deve ser racional a fim de que o mesmo recurso possa servir para a produção atual e também para as gerações futuras – esse é o princípio do *Desenvolvimento Sustentável*.²

O atual padrão tecnológico da produção industrial é intensivo em energia e matérias-primas. Além do mais, não é possível que uma tecnologia aproveite 100% dos insumos, sem gerar resíduos. Juntamente com a produção surgem os rejeitos industriais, e se suas quantidades forem maiores que a capacidade de absorção do meio ambiente, eles geram a poluição. Essa tem efeitos negativos sobre o bem-estar da população e sobre a qualidade dos recursos naturais, afetando a harmonia dos ecossistemas e aumentando os gastos públicos – o Quadro 26.1 exemplifica os gastos públicos com doenças relacionadas à poluição no Brasil. Se todos os países do mundo adotassem o mesmo padrão produtivo, os recursos naturais iriam se esgotar rapidamente e o planeta ficaria sem condições de ser habitado, dado o surgimento de problemas ambientais globais e locais.³

A poluição industrial está associada ao padrão de especialização da economia e à escala de produção. Ou seja, como algumas indústrias são mais poluidoras do que outras (Tabela 26.1), a estrutura industrial será mais poluidora quando for especializada em indústrias de grande potencial de emissões de poluentes. Quanto maior a escala de produção, maiores são as emissões industriais⁵ em termos absolutos.

A sistematização da Política Ambiental é recente, pois somente a partir do início do século XX a legislação ambiental começou a dar os primeiros passos. Isso não significa que não havia problemas ambientais no início da Revolução Industrial, pois a base energética da atividade econômica era a queima de carvão mineral, uma importante fonte de poluição do ar. Além do mais, o rápido processo de urbanização sem a infraestrutura adequada trouxe problemas ambientais, como poluição da água e geração de volumes crescentes de lixo sem disposição adequada.

A acumulação de poluentes e resíduos, juntamente com o aumento da atividade industrial e da concentração da população nos centros urbanos, começou a tomar proporções alarmantes. Surge, portanto, a necessidade de sistematizar as normas de conduta em relação ao meio ambiente.

QUADRO 26.1 GASTOS COM SAÚDE DEVIDOS À POLUIÇÃO ATMOSFÉRICA E HÍDRICA NO BRASIL

A poluição do ar é um dos grandes problemas atuais dos centros urbanos e industriais, devido ao rápido processo de urbanização e industrialização no Brasil. A excessiva concentração de poluentes no ar causa graves problemas para a saúde humana, sendo o principal responsável por doenças respiratórias como bronquite e bronquiolite agudas, gripe, alergias, bronquites crônicas, enfisema, asma, bronquiectasia, entre outras. As crianças e os idosos são as principais vítimas.

Tal como o ar, a água é um dos recursos naturais mais usados pelo homem, seja para desidratação humana ou animal, irrigação, geração de energia, insumo industrial, higiene pessoal, transporte, lazer e outros usos. A poluição hídrica é prejudicial à saúde humana, gerando várias doenças como cólera, infecções gastrintestinais, febre tifoide, poliomielite, amebíase, esquistossomose e shigelose.

Segundo estudo realizado pelo IPEA/RJ,⁴ os gastos médicos (realizados pelo sistema Inamps) associados à poluição hídrica doméstica no Brasil no ano de 1989 foram de US\$ 40,2 milhões e os gastos hospitalares na cidade de São Paulo com doenças causadas por poluição atmosférica para o mesmo ano foram de US\$ 785 mil. Os custos médios de saúde *per capita* associados à poluição hídrica foram de US\$ 2,97 e US\$ 0,84 associados à poluição atmosférica (para as populações de São Paulo, Rio de Janeiro e Cubatão).

É importante lembrar que esses são gastos com tratamento de doenças, sem contar as perdas econômicas que ocorrem devido à morbidade e à mortalidade causadas pela poluição. Ou seja, pessoas doentes perdem dias de trabalho e a produtividade cai, além da produção que foi sacrificada devido à morte de trabalhadores e da mão de obra futura.

Você já imaginou o quanto poderá ser economizado com a redução da poluição?

TABELA 26.1 Indústrias com Maior Potencial de Emissão de Poluentes

	Tipo de poluente	Tipo de indústria
Poluentes da água	Carga orgânica (DBO)	Metalurgia de não ferrosos; papel e gráfica; químicos, não petroquímicos; açúcar
	Sólidos suspensos	Siderurgia
Poluentes do ar	Dióxido de enxofre (SO_2)	Metalurgia de não ferrosos; siderurgia; refino de petróleo e petroquímica
	Dióxido de nitrogênio (NO_2)	Refino de petróleo e petroquímica; siderurgia
	Monóxido de carbono (CO)	Siderurgia; metalurgia de não ferrosos; químicos diversos; refino de petróleo e petroquímica
Poluentes do solo	Compostos orgânicos voláteis	Refino de petróleo e petroquímica; siderurgia; químicos diversos
	Material particulado inalável	Siderurgia; óleos vegetais e gorduras para alimentação; minerais não metálicos

Fonte: Young, C.E.F. e Pereira, A.A. (2000). Controle ambiental, competitividade e inserção internacional: uma análise da indústria brasileira. XXVIII Encontro Nacional de Economia, Campinas: ANPEC.

26.3 Soluções Econômicas para os Problemas Ambientais

Nesta seção vamos discutir a Política Ambiental sob uma perspectiva econômica, mostrando soluções teóricas para os problemas ambientais. Inicialmente, veremos que é possível criar um mercado para que poluidores e vítimas da poluição possam chegar a um acordo sobre o nível de poluição “ideal”, mas somente será possível em casos particulares, quando os direitos de propriedade estiverem bem definidos. Em outra abordagem, a poluição é vista como uma externalidade⁶ negativa e a solução para o problema é que o poluidor internalize as externalidades – princípio do poluidor-pagador.

26.3.1 A LIVRE NEGOCIAÇÃO (OU TEOREMA DE COASE)

Se há poluidores e vítimas da poluição, por que uns não negociam com os outros e tentam resolver o problema sem a interferência pública por meio da Política Ambiental? Em outras palavras, por que não se cria um mercado para negociar a quantidade de poluição aceitável?

Essa questão foi levantada por Ronald Coase (prêmio Nobel de 1991) em seu artigo *The Problem of Social Cost*, publicado em 1960, no qual argumenta que a solução dos problemas ambientais pode surgir por meio da livre negociação entre as partes envolvidas. O argumento central é que as partes interessadas, por meio de um processo de barganha, irão chegar a um consenso, resultando numa solução ótima (ou seja, melhor do que qualquer outra imposta “de fora”).

A hipótese básica do argumento é que quanto maior a redução na poluição, maior o custo marginal de abatê-la⁷ e menor é o benefício marginal para a parte afetada. A partir do nível de poluição Q_3 (Figura 26.2) inicia-se um processo de barganha e as partes

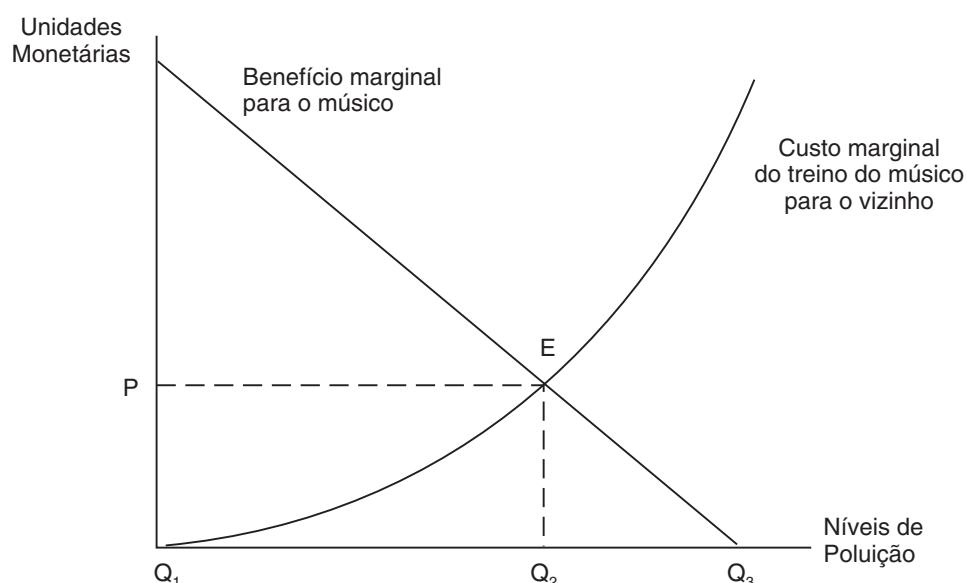


Figura 26.2 Produção física, produto industrial com alto potencial poluidor e total, Brasil, 1981/99 (1981 = 100).

QUADRO 26.2 EXEMPLO DE LIVRE NEGOCIAÇÃO ENTRE POLUIDOR E VÍTIMA DA POLUIÇÃO

Um baterista profissional resolveu praticar seu instrumento à noite, gerando muito barulho (ponto Q_3 na Figura 26.2). Mas esse é o mesmo período em que Cadu, um estudante de economia que estuda pela manhã e trabalha à tarde, tem para estudar. Como não havia nenhuma restrição legal à poluição sonora, Cadu resolveu ir até o apartamento do vizinho e negociar uma solução para o problema. O baterista, por sua vez, disse que quanto mais ensaiasse maior seriam seus rendimentos. O que fazer? Como na aula do dia Cadu aprendeu o teorema de Coase, ele propôs ao músico uma certa quantia para que a sessão de bateria não mais ocorresse (nível de poluição zero ou Q_1). Mas o baterista não aceitou, pois a quantia era inferior ao que receberia tocando. Contudo, sabendo que a produtividade do ensaio do baterista é decrescente com o tempo da sessão, e que a irritação de Cadu é crescente a cada minuto adicional que o vizinho pratica seu instrumento, eles foram negociando até que chegaram a um acordo: o músico tocaria menos à noite (nível de poluição Q_2) recebendo do estudante de economia uma compensação equivalente a P pelo tempo de redução do ensaio, obtido na interseção entre a curva de custo marginal de abatimento da poluição (custo do tempo de ensaio) com a curva benefício (menos irritação) marginal do estudante em escutar o barulho. Note que, pelo teorema de Coase, o mesmo ponto Q_2 de poluição seria alcançado se existisse uma lei de poluição sonora que atribuisse à vítima (no caso, o estudante) o direito de interromper a atividade causadora do barulho. Nesse caso, seria o baterista quem compensaria Cadu pelo tempo do ensaio, ao mesmo valor P que iguala o custo marginal de abatimento da poluição com benefício marginal de abatimento da poluição de suportar o barulho.

envolvidas irão negociar até o ponto em que o custo marginal de reduzir a poluição seja igual ao benefício marginal de reduzi-la – ponto E . Nesse ponto, a vítima da poluição não estará mais disposta a pagar um valor adicional para o agente poluidor para que este reduza a poluição. Ou seja, ela prefere “suportar” um pouco de poluição (Q_2) a gastar mais (acima de P) e o poluidor só aceitará reduzir ainda mais seus níveis de poluição por uma quantia maior do que a vítima está disposta a pagar (ver Quadro 26.2).

Entretanto, apesar da existência hipotética de uma solução de mercado, a livre negociação entre as partes envolvidas muitas vezes não é viável na prática, seja por causa do grande número de vítimas ou poluidores envolvidos, seja porque os direitos de propriedade sobre o ambiente não estão definidos. Por exemplo, como se daria a negociação entre as vítimas da poluição do ar por emissão veicular nos grandes centros urbanos com os usuários dos automóveis? Como negociar quando o poluidor também é uma das vítimas, ou não dispõe de informações adequadas sobre as perdas que lhe são infligidas? Ou seja, os direitos de propriedade devem estar bem delimitados para que possibilite a criação de um mercado.

Assim, a livre negociação só será possível em casos particulares, envolvendo poucos negociadores, devido a uma série de problemas na sua implementação:

1. Altos custos de transação, ou seja, os custos de reunir as diversas vítimas, de contratar advogados, os altos custos processuais, entre outros, podem ser tão elevados que não compensam os ganhos que serão obtidos.
2. Grande parte dos problemas ambientais afetam bens que não têm direito de propriedade bem definido, como parte dos recursos marinhos, o ar, a água e outros bens livres – ou seja, a livre negociação somente é possível quando os direitos de propriedade estão bem definidos.
3. Negociação entre gerações, isto é, há problemas ambientais cujas vítimas são as gerações futuras e não há um consenso em torno de quem irá negociar em nome delas.

Na prática, dadas as limitações da solução de mercado – a livre negociação entre as partes envolvidas –, formas diretas de intervenção são necessárias para fazer com que os poluidores, que causam problemas ambientais para diversas pessoas, assumam os custos da poluição ou reduzam seus níveis de emissão de poluentes.

QUADRO 26.3 EXEMPLO DA FÁBRICA DE CIMENTO QUE POLUI O AR

“Observe-se que esta situação só ocorre porque, como se disse anteriormente, o ar é um bem público, ou melhor, não é de propriedade de ninguém. Se o dono da fábrica fosse também o dono do ar (...), certamente tomaria todo cuidado para conservar esse ar em condições adequadas à produção do cimento. Igualmente, se as pessoas da cidade tivessem comprado o ar da cidade (e, portanto, possuíssem sobre ele um direito exclusivo), o dono da fábrica teria que incorporar uma indenização por estar poluindo o ar da cidade, ou recorreria a filtros de controle para não o poluir. Num e outro caso (...) percebe-se que o resultado (a alocação de recursos) é distinto do que se verifica na prática com os bens públicos. Pela natureza deste tipo de bens, há uma falha de mercado envolvida. Os custos privados diferem dos custos sociais, e uma empresa que maximiza lucro toma decisões que não são socialmente eficientes.”

Fonte: MARGULIS, S. (org.) (1996). Meio Ambiente: aspectos técnicos e econômicos. Brasília:IPEA, Cap. 6, p.136.

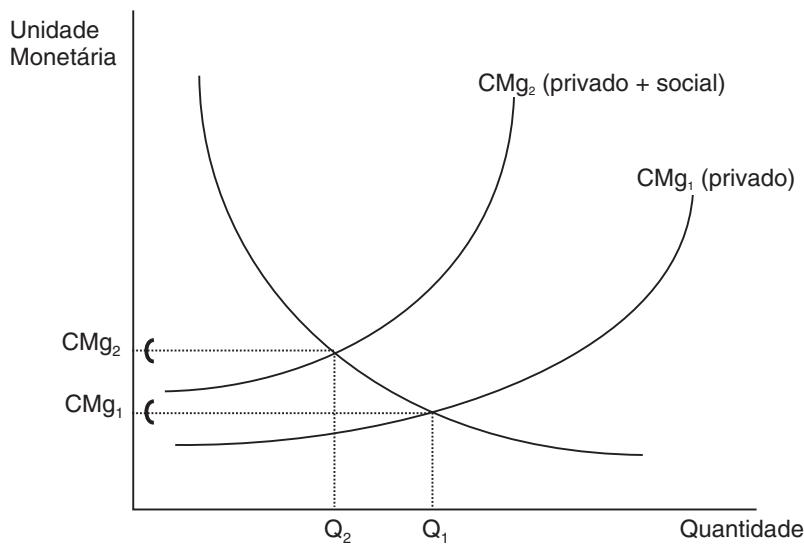


Figura 26.3 Livre negociação entre o poluidor e a vítima da poluição.

26.3.2 A INTERNALIZAÇÃO DAS EXTERNALIDADES

Os mercados podem apresentar falhas importantes no seu funcionamento quando parte dos custos envolvidos na produção ou consumo de um recurso não é plenamente capturada por aqueles que se beneficiam de tais atos. Assim, esses custos não são contabilizados *privadamente*, não havendo motivação para que seus causadores corrijam suas ações. Mas isso não significa que eles não acabem incidindo sobre outras parcelas da sociedade. Diz-se, então, que tais custos são *socializados* por meio da geração de externalidades.

Microeconomicamente, isso quer dizer que as curvas de custo dos que são geradores da emissão (Figura 26.3) não incorporam plenamente todos os custos envolvidos (CMg_1) e por isso as quantidades produzidas (Q_1) excedem o nível de ótimo social (Q_2).

Qual a solução para esse problema? Fazer com que o poluidor incorpore as externalidades, ou seja, os custos sociais também devem ser adicionados ao custo marginal privado, fazendo com que a curva de custo marginal se desloque para a esquerda (CMg_2). Assim, a fim de que o poluidor incorpore as externalidades, o custo marginal a ser considerado deve ser CMg_2 e a quantidade produzida deve ser menor, atingindo o ótimo social em Q_2 . Este é o princípio do *poluidor-pagador*.

26.4 Instrumentos de Política Ambiental

Os instrumentos de Política Ambiental têm a função de internalizar o custo externo ambiental e podem ser divididos em três grupos: instrumentos de comando e controle (ou regulação direta), instrumentos econômicos (ou de mercado) e instrumentos de comunicação (Quadro 26.4). Todos eles possuem vantagens e desvantagens, razão pela qual a análise da experiência de diversos países evidencia uma combinação desses diversos tipos de instrumentos em suas Políticas Ambientais.

QUADRO 26.4 TIPOLOGIA E INSTRUMENTOS DE POLÍTICA AMBIENTAL

Comando e controle	Instrumentos Econômicos	Instrumentos de Comunicação
- Controle ou proibição de produto	- Taxas e tarifas	- Fornecimento de informação
- Controle de processo	- Subsídios	- Acordos
- Proibição ou restrição de atividades	- Certificados de emissão transacionáveis	- Criação de redes
- Especificações tecnológicas	- Sistemas de devolução de depósitos	- Sistema de gestão ambiental
- Controle do uso de recursos naturais		- Selos ambientais
- Padrões de poluição para fontes específicas		- Marketing ambiental

Fonte: Elaboração própria com base em Kemp, R., Smith, K. e Becher, G. (2000). *How should we study the relationship between environmental regulation and innovation?* Relatório final do programa de pesquisa DGIII-IPTS e Almeida, L. T. (1998). *Política Ambiental: uma análise econômica*. Campinas:Papirus.

26.4.1 INSTRUMENTOS DE COMANDO E CONTROLE

Os instrumentos de comando e controle são também chamados de instrumentos de regulação direta, pois implicam o controle direto sobre os locais que estão emitindo poluentes. O órgão regulador estabelece uma série de normas, controles, procedimentos, regras e padrões a serem seguidos pelos agentes poluidores e também diversas penalidades (multas, cancelamento de licença, entre outras) caso eles não cumpram o estabelecido.

Esse procedimento requer uma fiscalização contínua e efetiva por parte dos órgãos reguladores, implicando altos custos de implementação. Os instrumentos de comando e controle são eficazes no controle dos danos ambientais, mas podem ser injustos por tratar todos os poluidores da mesma maneira, sem levar em conta diferenças de tamanho da empresa e a quantidade de poluentes que lança no meio ambiente. São exemplos de instrumentos de comando e controle: exigência de utilização de filtros em chaminés das unidades produtivas, fixação de cotas para extração de recursos naturais (madeira, pesca e minérios), concessão de licenças para funcionamento de fábricas e obrigatoriedade de substituição da fonte energética da unidade industrial – substituição de lenha por energia hidrelétrica em siderúrgicas, por exemplo.

26.4.2 INSTRUMENTOS ECONÔMICOS

Os instrumentos econômicos são também denominados de instrumentos de mercado e visam à internalização das externalidades ou de custos que não seriam normalmente incorridos pelo poluidor ou usuário.

As principais vantagens da utilização dos instrumentos econômicos em relação aos de comando e controle são:

1. Permitir a geração de receitas fiscais e tarifárias – por meio da cobrança de taxas, tarifas ou emissão de certificados – para garantir os recursos para pagamento dos incentivos e prêmios ou capacitar os órgãos ambientais. É considerado um *duplo-dividendo*, pois além da melhoria ambiental, gera receitas para os órgãos reguladores;
2. Considerar as diferenças de custo de controle entre os agentes e, portanto, alocar de forma mais eficiente os recursos econômicos à disposição da sociedade, permitindo que aqueles com custos menores tenham incentivos para expandir as ações de controle. Portanto, com os instrumentos econômicos a sociedade incorre em custos de controle inferiores àqueles que seriam incorridos se todos os poluidores ou usuários fossem obrigados a atingir os mesmos padrões individuais;
3. Possibilitar que tecnologias menos intensivas em bens e serviços ambientais sejam estimuladas pela redução da despesa fiscal que será obtida em função da redução da carga poluente ou da taxa de extração;
4. Atuar no início do processo de uso dos bens e serviços ambientais;
5. Evitar os dispêndios em pendências judiciais para aplicação de penalidades;
6. Implementar um sistema de taxação progressiva ou de alocação inicial de certificados segundo critérios distributivos em que a capacidade de pagamento de cada agente econômico seja considerada.

São exemplos de instrumentos econômicos: empréstimos subsidiados para agentes poluidores que melhorarem seu desempenho ambiental, taxas sobre produtos poluentes, depósitos reembolsáveis na devolução de produtos poluidores – o antigo depósito sobre vasilhames de vidro – e licenças de poluição negociáveis – a fábrica tem um patamar máximo de emissões e caso não o utilize, pode negociar sua licença “para poluir” com terceiros.

26.4.3 INSTRUMENTOS DE COMUNICAÇÃO

Os instrumentos de comunicação são utilizados para conscientizar e informar os agentes poluidores e as populações atingidas sobre diversos temas ambientais, como: danos ambientais causados, atitudes preventivas, mercados de produtos ambientais, tecnologias menos agressivas ao meio ambiente e aumento da cooperação entre os agentes poluidores para buscar soluções ambientais. São exemplos de instrumentos de comunicação: a educação ambiental, a divulgação de benefícios para as empresas que respeitam o meio ambiente e os selos ambientais.

26.5 Política Ambiental e Comércio Internacional

A Política Ambiental de diferentes países pode influenciar nos fluxos de comércio internacional. Quando o produto e/ou seu método de produção causam problemas ambientais, o país importador pode colocar barreiras ao comércio internacional. Essas são barreiras não tarifárias, também chamadas de *barreiras verdes*, pois restringem o comércio internacional com a finalidade de proteger o meio ambiente.

Quais são os principais problemas ambientais causados pelo comércio internacional?

1. *Danos ambientais causados pelo uso de um produto:* o país importador teria problemas ambientais por causa do consumo do produto importado. Nesse caso, o país produtor estaria exportando o produto juntamente com problemas ambientais. Por exemplo, o uso de produtos importados que contenham gases tipo CFC causa danos ao meio ambiente, visto que contribui para a destruição da camada de ozônio.
2. *Danos ambientais causados por processos e métodos de produção (PPMs)⁸:* a maneira pela qual o produto foi produzido causa danos ambientais no próprio país produtor ou em outros países. Esses danos podem ser:
 - *Po*luição transfronteiriça ocorre quando determinados PPMs afetam negativamente o meio ambiente de outros países vizinhos ou da mesma região. Ocorre com frequência quando há poluição do ar ou dos recursos hídricos comuns. Por exemplo, a emissão de dióxido de enxofre (SO_2) em um país pode provocar chuva ácida em outro;
 - *Espécies migratórias e recursos vivos comuns* são ameaçados de extinção quando são praticados PPMs agressivos ao meio ambiente. São exemplos: a extinção de espécies aquáticas devido a métodos de pesca nocivos, geralmente com redes de malha de tamanho impróprio, capturando indiscriminadamente diversas espécies. As tartarugas marinhas foram ameaçadas de extinção por causa do método de pesca inadequado de camarões;
 - *Preocupações com o meio ambiente global* surgem quando os PPMs afetam recursos comuns a todos os países. Por exemplo, os impactos negativos causados na camada de ozônio pela utilização de gases tipo clorofluorcarbono (CFC); as mudanças climáticas causadas pela derrubada de florestas tropicais nativas, resultado da extração irracional de madeiras e sem reposição da biomassa; ou o aquecimento global resultante, entre outras causas, da alta concentração de dióxido de carbono⁹ – CO_2 – na atmosfera, provocado pela queima de combustíveis fósseis em processos produtivos, além de outras fontes emissoras;
 - *Preocupações com o meio ambiente local* ocorrem quando os PPMs levam a danos ambientais nos limites geográficos do país. A poluição da água, terra ou ar que não cause efeitos transfronteiriços são exemplos desse tipo de problema.

Assim, por causa desses problemas ambientais que podem surgir, os países que importam mercadorias do Brasil podem impor barreiras verdes, reduzindo as exportações e consequentemente a produção industrial, diminuindo o produto, a renda e o emprego gerados pela indústria (Quadro 26.5).

26.6 Política Ambiental no Brasil

26.6.1 HISTÓRICO DA POLÍTICA AMBIENTAL

O atraso no estabelecimento de normas ambientais e agências especializadas no controle da poluição industrial demonstra que, de fato, a questão ambiental não configurava entre as prioridades de política pública. Até a década de 1970, não existia um órgão especificamente voltado ao controle ambiental. As legislações existentes tratavam da exploração de alguns recursos naturais, por meio de medidas isoladas:

QUADRO 26.5 A PROIBIÇÃO DA IMPORTAÇÃO DE GASOLINA BRASILEIRA PELOS EUA

Um exemplo de restrição comercial baseada em regulamentação ambiental foi o caso das gasolinas brasileira e venezuelana exportadas para os EUA. Em 1995, para complementar o *Clean Air Act*, a agência ambiental norte-americana (EPA) estabeleceu a Regra da Gasolina (*Gasoline Rule*), ou seja, impôs padrões de qualidades para a gasolina utilizada pelos consumidores norte-americanos. A exigência era de que somente gasolina de determinada especificação (*reformulated gasoline*) poderia ser vendida nas áreas mais poluídas e, no restante do país, somente a gasolina convencional (*conventional gasoline*), cujo padrão não poderia ser mais sujo do que do ano-base de 1990. Entretanto, para as refinarias norte-americanas que já operavam em 1990, foi estabelecido um patamar mínimo individual com base no padrão de qualidade da gasolina que produzia em 1990. Ou seja, os padrões individuais para as refinarias domésticas poderiam ser inferiores aos estabelecidos para a gasolina importada, o que impôs ao Brasil e à Venezuela processos mais onerosos que os domésticos. Esses países reivindicaram que a Regra da Gasolina era inconsistente, com o artigo III do GATT,¹⁰ e não estava coberta pelo Artigo XX.¹¹ Os Estados Unidos argumentaram que a Regra da Gasolina era consistente com o artigo III, e estava justificada nas exceções contidas no artigo XX do GATT, parágrafos (b), (g) e (d). Os EUA perderam em todas as instâncias julgadas no GATT/OMC, pois os painéis não aceitaram os argumentos apresentados e os EUA foram forçados a voltar atrás e retirar a restrição às gasolinas estrangeiras.

1. Código Florestal de 1934 (Decreto n. 23.793), tratava da questão das matas nativas, sendo posteriormente reformado pela Lei n. 4.771 de 1965.
2. Código de Águas de 1934 (Decreto n. 24.643), estabelecendo normas de uso dos recursos hídricos, com especial atenção ao seu aproveitamento hidrelétrico.
3. Comissão Executiva da Defesa da Borracha de 1947 (Lei n. 86), estabelecendo medidas que visam à assistência econômica da borracha natural, reestruturada posteriormente em 1967.
4. Superintendência do Desenvolvimento da Pesca (SUDEPE) de 1962 (Lei Delegada n. 10), vinculada ao Ministério da Agricultura.

Somente em 1973 a questão ambiental passou a ser tratada com uma estrutura independente, seguindo a recomendação da Conferência das Nações Unidas sobre o Meio Ambiente. Criou-se a Secretaria Especial do Meio Ambiente – SEMA (Decreto n. 73.030), vinculada ao Ministério do Interior. A estrutura do sistema de gestão ambiental tomou por modelo a experiência norte-americana, caracterizada por dois elementos básicos: um grande nível de descentralização e um acentuado viés regulatório, baseado nos instrumentos de comando e controle, favorecendo a regulação direta das empresas e, por isso, demandando recursos humanos e técnicos para o controle que, no caso brasileiro, estão muito acima das disponibilidades dos órgãos fiscalizadores.

Embora leis e normas tenham sido criadas ainda na década de 1970, apenas em 1981 é que a Lei n. 6.938 estabeleceu os objetivos, as ações e os instrumentos da *Política Nacional do Meio Ambiente*, objetivando não só a preservação, melhoria e recuperação da qualidade ambiental, mas também comprometendo-se a assegurar as condições para o desenvolvimento socioeconômico, aos interesses da segurança nacional e à proteção da dignidade da vida humana, atendendo os princípios já previstos na Constituição e constituindo outros que asseguram a tutela jurídica do meio ambiente. Os instrumentos explicitados são (art. 9):

1. O estabelecimento de padrões de qualidade ambiental
2. O zoneamento ambiental.
3. A avaliação de impactos ambientais.
4. O licenciamento e a revisão de atividades efetiva ou potencialmente poluidoras.

Foi constituído o Sistema Nacional do Meio Ambiente (SISNAMA) e também criou-se o Conselho Nacional do Meio Ambiente (CONAMA) (Quadro 26.6).

QUADRO 26.6 O AMBIENTE INSTITUCIONAL DA POLÍTICA AMBIENTAL NO BRASIL

O ambiente institucional é regulamentado em três esferas de poder – federal, estadual e municipal (incluindo o Distrito Federal). São três órgãos reguladores no âmbito federal:

1. Ministério do Meio Ambiente (MMA) – é o responsável pela formulação, planejamento, coordenação, supervisão e controle da Política Nacional do Meio Ambiente e dá as diretrizes nacionais para o meio ambiente. É o órgão central do SISNAMA.
2. Conselho Nacional do Meio Ambiente (CONAMA) – é o órgão consultivo e deliberativo do SISNAMA e está vinculado ao MMA. O CONAMA é um órgão colegiado cujos membros são representantes do Governo e da sociedade civil que têm envolvimento com as questões ambientais e que tem a finalidade de assessorar, estudar e propor as diretrizes de políticas governamentais para o meio ambiente e os recursos naturais. É composto de dez câmaras técnicas permanentes e oito temporárias, que discutem questões relativas a gerenciamento costeiro, energia, controle ambiental, ecossistemas, recursos hídricos, recursos naturais e outros temas relevantes. A determinação dos padrões de qualidade ambiental é de competência do CONAMA. Estes parâmetros são normalmente baseados na experiência internacional, como no caso dos padrões de qualidade do ar (determinados a partir de padrões internacionais, como os da *Environment Protection Agency* – EPA – a agência de proteção ambiental norte-americana).
3. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) – foi criado em 1989 e assumiu os direitos, créditos, obrigações e receitas dos órgãos reguladores extintos.¹² É o órgão executor do SISNAMA e vinculado ao MMA. Cabe ao IBAMA (no nível federal) a responsabilidade pelo controle e fiscalização de atividades capazes de provocar a degradação ambiental. A exigência de estudos de impacto ambiental (EIA) e relatórios de impacto ambiental (RIMA) constitui a base de avaliação para o licenciamento das atividades efetiva ou potencialmente poluidoras, sendo este o principal instrumento disponível para a gestão ambiental.
4. Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) – foi criado a partir da reestruturação do IBAMA em 2007 e é o responsável pela administração das Unidades de Conservação (UCs) e da pesquisa sobre biodiversidade, além de sua proteção e conservação. Tais atribuições eram de responsabilidade do IBAMA. Está vinculado ao MMA e integra o SISNAMA.

Nos âmbitos estadual e municipal, o controle e a fiscalização de atividades que têm impactos negativos sobre o meio ambiente é de responsabilidade dos órgãos ou entidades estaduais e municipais – como o Instituto Estadual do Ambiente – INEA no Rio de Janeiro e a Companhia de Tecnologia de Saneamento Ambiental (CETESB) em São Paulo. As multas e outras penalidades aos agentes que violam os padrões estabelecidos são determinadas de forma diferenciada pelas agências estaduais de controle. Não há hierarquia entre as agências federais, estaduais e municipais, sendo umas independentes das outras.

Em 1996, foi instituída a Política Nacional de Recursos Hídricos (Lei n. 9.433) e criou-se o Sistema Nacional de Gerenciamento de Recursos Hídricos e o Conselho Nacional de Recursos Hídricos. Em 2000, foi criada a Agência Nacional de Águas – ANA (Lei n. 9.984), responsável pela implementação da Política Nacional de Recursos Hídricos e pela coordenação do sistema nacional de gerenciamento de recursos hídricos. Ainda relacionado aos recursos hídricos, a Lei n.10.881 de 2004 dispõe sobre os contratos de gestão entre a ANA e entidades delegatárias das funções de agências de águas relativas à gestão de recursos hídricos de domínio da união. Em 2010, foi estabelecida a política nacional de segurança de barragens destinadas à acumulação de água para quaisquer usos, à disposição final ou temporária de rejeitos e à acumulação de resíduos industriais, e também foi criado o sistema nacional de informações sobre segurança de barragens, ambos na forma da Lei n.12.334.

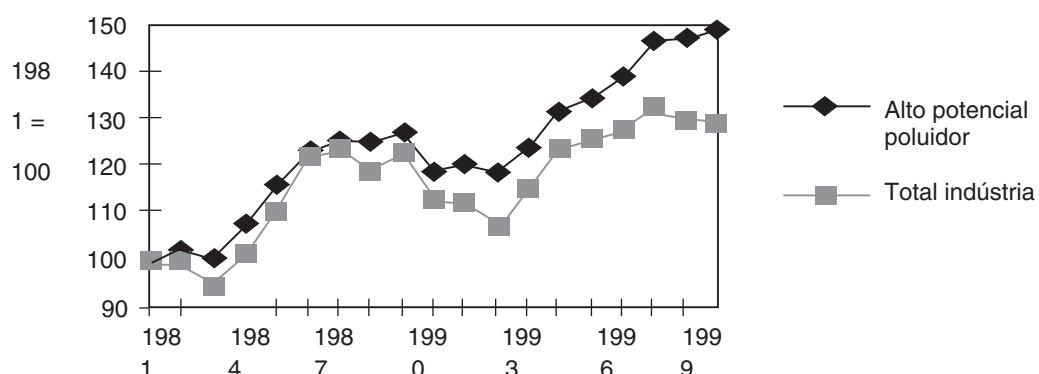
No que tange a outros temas importantes para o meio ambiente, em 1998, as condutas e atividades lesivas ao meio ambiente passam a sofrer sanções penais e administrativas estabelecidas pela Lei n. 9.605 – a Lei de Crimes Ambientais. O Sistema Nacional de Unidades de Conservação da Natureza – SNUC foi instituído em 2000, pela Lei n. 9.985. Em 2006, a Lei n. 11.284 regulamentou a gestão de florestas públicas para a produção sustentável, instituindo o Serviço Florestal Brasileiro – SFB na estrutura do MMA e criando o Fundo Nacional de Desenvolvimento Florestal – FNDF e a Lei n. 11.428 ordenou a utilização e proteção da vegetação nativa do bioma da Mata Atlântica. A Política Nacional de Resíduos Sólidos foi instituída pela Lei n. 12.305

26.6.2 PRINCIPAIS CARACTERÍSTICAS DA POLÍTICA AMBIENTAL BRASILEIRA

A questão ambiental não foi prioridade no processo de industrialização brasileiro. Desde o estabelecimento de indústrias intensivas em emissões, vindas dos países desenvolvidos nos anos 1970 para produzir bens intermediários, aos vazamentos de óleo do setor petrolífero ocorridos no ano 2000, são vários os exemplos de descaso do setor industrial brasileiro com a questão ambiental. Uma das consequências desse relativo descaso com a questão ambiental é a presença cada vez mais importante de indústrias intensivas em recursos naturais e energia, ou que apresentam um alto potencial poluidor. Isso é demonstrado pela Figura 26.4, na qual observamos que o crescimento das indústrias de alto potencial poluidor no período 1981-1999 foi nitidamente superior ao da média geral da indústria, sugerindo uma especialização relativa em atividades potencialmente “sujas”.

Uma série de razões pode ser apontada para explicar essa intensificação das atividades poluentes na composição setorial do produto industrial. Em primeiro lugar, o atraso no estabelecimento de normas ambientais e agências especializadas no controle da poluição industrial demonstra que, de fato, a questão ambiental não configurava entre as prioridades de política pública – apenas na segunda metade dos anos 1970 foi criado o primeiro órgão especificamente para esse fim, a Fundação Estadual de Engenharia do Meio Ambiente no Rio de Janeiro (FEEMA).

Em segundo lugar, a estratégia de crescimento associada à industrialização por substituição de importações no Brasil privilegiou setores intensivos em emissão. A motivação inicial do processo de industrialização por substituição de importações era baseada na percepção de que o crescimento de uma economia periférica não poderia ser apenas sustentado em produtos



Fonte: IBGE - Departamento de Indústria (DIND)

Figura 26.4 Custos marginais privados com a incorporação das externalidades.

diretamente baseados em recursos naturais (extração mineral, agricultura ou outras formas de aproveitamento de vantagens comparativas absolutas definidas a partir da dotação de recursos naturais). Contudo, embora o Brasil tenha avançado na consolidação de uma base industrial diversificada, esse avanço esteve calcado no uso indireto de recursos naturais (energia e matérias-primas baratas), em vez de expandir-se por intermédio do incremento na capacidade de gerar ou absorver progresso técnico – chave para o crescimento sustentado, mas que ficou limitado a algumas áreas de excelência.

Essa concentração em atividades intensivas em emissão aumentou ainda mais a partir da consolidação dos investimentos do II Plano Nacional de Desenvolvimento (II PND), que resultou em forte expansão de indústrias de grande potencial poluidor – especialmente dos complexos metalúrgico e químico/petroquímico – sem o devido acompanhamento de tratamento dessas emissões. Outro fator que contribuiu para o incremento de atividades industriais poluidoras foi a tendência de especialização do setor exportador em atividades potencialmente poluentes.

Dadas essas especificidades da industrialização brasileira em relação ao meio ambiente, a Política Ambiental vai adquirir contornos específicos, com ampla utilização dos instrumentos de comando e controle. O modelo de licenciamento concentrou a maior parte dos avanços em poucos estados onde as agências de meio ambiente desenvolveram-se de forma mais efetiva, como o INEA e a CETESB.

Contudo, existe grande heterogeneidade entre essas agências e, na maioria dos estados, existe enorme carência técnica, financeira e de pessoal que impede uma ação efetiva. Vários desses problemas foram agravados pela crise fiscal vivenciada por diversas administrações estaduais e falta de apoio político, resultando no declínio da qualidade da ação das agências ambientais. As lacunas deixadas por essa crescente paralisia estadual têm sido parcialmente preenchidas pela maior atuação de agências municipais, principalmente nas regiões metropolitanas. Essa multiplicidade de níveis administrativos pode levar a conflitos de competência, com prejuízo para as populações envolvidas.

Mesmo onde houve fortalecimento das agências ambientais, a qualidade ambiental não necessariamente melhorou, como em São Paulo. Isso se deve, além das dificuldades internas do setor ambiental, à falta de investimento em infraestrutura e serviços urbanos que são de outras competências administrativas (saneamento, transporte público, coleta de lixo, habitação popular); à persistência de grandes bolsões de pobreza (proliferação de favelas e outros ambientes degradados); e padrões de consumo que resultam em agravamento das condições ambientais (o rápido crescimento da frota de automóveis particulares é o exemplo mais gritante).

Os aspectos ambientais ainda estão pouco integrados na formulação de políticas públicas. O problema é agravado pela falta de informações sobre a extensão e relevância dos problemas resultantes da degradação ambiental. A criação de um sistema de indicadores ambientais que compile dados obtidos pelas agências de controle poderia facilitar essa integração, definindo áreas de prioridade de ação.

Essa série de problemas levou a ser questionado o atual sistema de gestão que se baseia nos instrumentos de comando e controle. Em primeiro lugar, a ação desses órgãos é reativa, sendo que a expansão de suas atividades é normalmente resultado do agravamento de problemas não resolvidos. Os procedimentos atuais de estudo de impacto ambiental (EIA) e relatório de impacto ambiental (RIMA) são passíveis de várias críticas, pois existe pouca clareza quanto aos critérios adotados no seu enquadramento como instrumentos de avaliação de impactos ambientais. Na prática, não se observa a apresentação de alternativas tecnológicas e locacionais, e as áreas de influência consideradas são bastante restritas. Além disso, o princípio de independência da equipe responsável pela elaboração dos estudos entra em contradição com a prática de se permitir ao proponente do projeto a indicação e contratação dessa equipe.

A fiscalização também apresenta sérios problemas. Os principais referem-se à já mencionada escassez de recursos humanos e financeiros, em virtude da crise do estado brasileiro em seus diversos níveis e à fraca integração entre esses distintos níveis de governo. Além disso, a rigidez do sistema de normas atualmente vigente retira flexibilidade dos gestores ambientais, tornando-os muito pouco efetivos no sentido de criar estímulos aos agentes sociais para que adotem práticas mais adequadas ao ambiente.

Uma vez atendidos os padrões de emissão, há pouco interesse do agente para que melhore ainda mais sua performance. Por outro lado, no caso de indústrias estabelecidas sob um padrão tecnológico prévio à definição dos padrões ambientais, os custos de readaptação podem ser significativos, o que levaria em alguns casos ao fechamento da indústria, com enorme custo social.

A saída apontada por grande parte da literatura para esse problema é a incorporação de instrumentos econômicos baseados no princípio do poluidor-pagador, no qual emissões passam a ser cobradas mesmo estando em conformidade com os padrões máximos, mas ao mesmo tempo permitindo que os agentes emissores negociem entre si seus próprios limites de emissão para minimizar os custos sociais de ajuste.¹³ Indústrias mais antigas, cujo custo de readaptação seja mais alto, poderiam se beneficiar por meio da negociação com outros agentes mais eficientes no controle ambiental. Além disso, ao contrário de um único padrão,

haveria a possibilidade de impor um “preço” mais elevado nas emissões que resultem em maior poluição e valores mais baixos onde o problema não é relevante, pois a concentração de poluentes resultante de emissões pode variar consideravelmente devido às circunstâncias do local afetado.

Em suma, embora a experiência brasileira, especialmente em alguns estados do Sudeste e do Sul do país, possa ser considerada como avançada se comparada com outros países latino-americanos, o modelo de gestão resultou em avanços limitados no controle da poluição e outras formas de degradação. Problemas importantes permanecem sem solução e, se comparados com padrões de países desenvolvidos, os indicadores de qualidade ambiental no Brasil ainda estão bastante abaixo do satisfatório. Se por um lado a dinâmica errática de crescimento econômico, a acelerada urbanização e a crise do Estado ocorridas a partir dos anos 1980 podem ser apontadas como parte da questão, por outro lado o modelo de gestão adotado mostrou-se inadequado para tratar de diversos problemas. Sendo assim, os próprios gestores ambientais reconhecem a necessidade de se buscar formas mais eficientes de controle.

Notas

1. Os recursos naturais podem ser classificados em três tipos: os renováveis – fauna e flora –, os não renováveis – minerais e fósseis – e os livres – água, ar, luz solar e outros que existem em abundância.
2. O Desenvolvimento Sustentável é definido no estudo da Comissão Mundial sobre Meio Ambiente e Desenvolvimento, intitulado *Nosso Futuro Comum*, editado em 1991 pela Fundação Getúlio Vargas, como “aquele que atende às necessidades do presente sem comprometer a possibilidade de as gerações futuras atenderem às suas próprias necessidades” (p. 46). Centra-se em três eixos principais: crescimento econômico, equidade social e equilíbrio ecológico.
3. São exemplos de problemas ambientais globais o aquecimento global (efeito estufa) e a destruição da camada de ozônio. Os problemas ambientais locais são, entre outros, a poluição de recursos hídricos e a degradação dos solos.
4. Seroa da Motta, R. (coord.) (1995). *Contabilidade Ambiental: teoria, metodologia e estudos de caso no Brasil*. Rio de Janeiro: IPEA, caps. 8 e 9.
5. É importante distinguir emissões industriais de poluição industrial. As emissões são os resíduos da atividade industrial, que são em parte absorvidas pelo meio ambiente. Quando a capacidade assimilativa do meio ambiente é inferior à quantidade de emissões surge, então, a poluição.
6. Sobre o conceito de externalidade ver Hal Varian, *Microeconomia: princípios básicos*. Rio de Janeiro: Editora Campus, 1999.
7. *Abatimento*, segundo o Dicionário Brasileiro de Ciências Ambientais, é a “diminuição de uma parcela de poluição por meio de técnicas de controle ambiental”.
8. Refere-se ao termo em inglês, *processes and production methods* (PPMs).
9. As emissões de CO₂ na atmosfera são preocupantes, pois seu tempo médio de decaimento é de 100 anos. Logo, as emissões desse gás hoje irão produzir efeitos na atmosfera por cerca de 100 anos.
10. O artigo III do GATT estabelece que não se pode tratar o produto estrangeiro diferentemente do nacional.
11. Este artigo do GATT estabelece exceções às regras, as quais incluem as questões ambientais.
12. Foram extintos a SUDEHVEA (Superintendência da Borracha), o IBDF (Instituto Brasileiro de Desenvolvimento Florestal), a SEMA (Secretaria Especial do Meio Ambiente) e a SUDEPE (Superintendência do Desenvolvimento da Pesca) pelas Leis n. 7.732 e n. 7.735.
13. A adoção de instrumentos econômicos permite que um agente emita acima de um padrão médio estabelecido, desde que outros agentes decidam reduzir seu nível de emissão por meio de compensações financeiras, diretas (venda de certificados de emissão) ou indiretas (redução do imposto a pagar). Dentro do sistema vigente, o fato de um agente emitir em um nível muito abaixo do padrão legal não lhe garante vantagem adicional em relação à situação em que está apenas ligeiramente abaixo do padrão.

GUIA DE LEITURA

Para que você possa saber mais sobre Política Ambiental, sugerimos uma breve lista de livros e sites interessantes:

- *Economia do Meio Ambiente*, de Peter May, Maria Cecilia Lustosa e Valéria da Vinha (organizadores), publicado pela Campus em 2003.
Em vários capítulos de diferentes autores, o livro trata de vários aspectos da Economia do Meio Ambiente, desde o enfoque da economia neoclássica como o da economia ecológica. Outros temas como agricultura, comércio internacional, recursos hídricos e mudanças climáticas são tratados em suas interfaces com o meio ambiente.
- *Política Ambiental: uma análise econômica*, de Luciana Togeiro de Almeida, publicado pela Papirus em 1998.
Mostra diferentes enfoques teóricos (neoclássico, institucionalista e evolucionista) dos instrumentos de Política Ambiental, descreve a experiência de Política Ambiental de vários países e apresenta os instrumentos privilegiados de Política Ambiental no Brasil.
- *Economic Adjustment Policies and the Environment: a case study of Brazil*, de Carlos Eduardo Frickmann Young, tese de doutorado da University of London, 1996.

Analisa as consequências ambientais dos programas de ajustamento para a crise da dívida externa dos países em desenvolvimento no início dos anos 1980, examinando as relações entre as políticas de ajustamento e o uso dos recursos naturais, concentrando-se no caso brasileiro.

- *A Regulamentação Ambiental: Instrumentos e Implementação*, de Sérgio Margulis, publicado pelo IPEA/RJ como Texto para Discussão n. 437, em 1996.

Apresenta os principais instrumentos de Política Ambiental e relata a experiência de diversos países no assunto.

- *Meio Ambiente: aspectos técnicos e econômicos*, de vários autores, editado por Sérgio Margulis e publicado pelo IPEA/Brasília, em 1996.

Enfoca diversos assuntos relacionados com o meio ambiente, como poluição do ar e da água, avaliação de impactos ambientais, análise de custo-benefício e serve como introdução à economia dos recursos naturais e à economia do meio ambiente, abordando questões relativas à Política Ambiental.

- *Natural Resources and Environmental Economics*, de Roger Perman, Yue Ma e James McGilvray, publicado pela Longman, em 1996.

Apresenta uma cobertura bastante ampla sobre economia ambiental, incluindo a economia dos recursos naturais e da poluição, a política de controle da poluição, as contas ambientais e os problemas da poluição global. Relaciona temas como crescimento demográfico, crescimento econômico e economia do bem-estar social ao meio ambiente.

<http://www.ie.ufrj.br/gema/index.php> Grupo de Economia do Meio Ambiente e Desenvolvimento

Sustentável – IE/UFRJ

<http://www.ecoeco.org.br>

Sociedade Brasileira de Economia Ecológica (ECOECO)

<http://www.mma.gov.br>

Ministério do Meio Ambiente

<http://www.ibama.gov.br>

Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais

Renováveis (Ibama) <http://www.ecoeco.org>

International Society for Ecological Economics

<http://www.epa.gov>

Agência de Proteção Ambiental dos EUA

(*Environmental Protection Agency – EPA*)



PARTE VII

GUIA PARA ANÁLISES EMPÍRICAS

Modelos Estatísticos

Lucia S. Kubrusly

27.1 Introdução

As técnicas empíricas em Economia Industrial para quantificar e predizer as formas de comportamento das empresas têm sido muito utilizadas. A técnica mais usada por diversos autores foi a análise de regressão. Estes autores estavam interessados em examinar a existência de relação entre a estrutura de mercado e a lucratividade, e saber se a existência de barreiras à entrada elevadas permitia a obtenção de lucros mais altos e como a estrutura do mercado afetava a taxa de avanço tecnológico, e outras questões desse tipo.

A análise de regressão é adequada quando as variáveis envolvidas são quantitativas, de preferência tenham distribuição normal e, principalmente, quando existe uma hipótese teórica, como por exemplo um modelo econômico, que estabeleça a relação entre as variáveis envolvidas.

Mais recentemente têm crescido o uso de outras técnicas estatísticas para a análise de indústria devido principalmente à grande quantidade e diversidade dos dados coletados e à falta de modelos teóricos que estabeleçam relações funcionais entre tantas variáveis. As técnicas de análise exploratória de dados têm sido cada vez mais utilizadas, devido ao fato de não necessitarem de hipóteses iniciais nem quanto às relações existentes entre as variáveis, nem com relação à forma das distribuições de probabilidades das variáveis envolvidas. A desvantagem é que os resultados obtidos permitem afirmações mais fracas, do ponto de vista teórico.

O objetivo deste capítulo é discutir algumas destas técnicas estatísticas alternativas exemplificando o uso destes modelos na análise da indústria. Para isso é necessário introduzir alguns conceitos básicos de análise de dados. O primeiro seria o conceito de *banco de dados* que em geral é formado por diversas informações (variáveis) observadas para um conjunto de elementos ou objetos. Esses objetos formam a amostra ou o universo de estudo. As informações podem ser dispostas numa matriz que chamamos matriz de dados. Considerando um conjunto de p variáveis, observadas para n elementos, podemos escrever:

$$X(n \times p) = \begin{bmatrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{bmatrix}$$

onde X_{ij} é o valor da j -ésima variável, observada para o i -ésimo elemento.

O problema de análise de dados assume diferentes aspectos, dependendo do tamanho do banco de dados (isto é, do número de variáveis), e também do tipo de variável observada. Quanto aos tipos de variáveis, estas podem ser quantitativas (discretas ou contínuas), ordinais e categóricas. Variáveis quantitativas são as que expressam características de natureza numéricas como por exemplo: preço, custo, dias trabalhados durante o ano, número de peças defeituosas num processo de produção etc. Variáveis categóricas, por outro lado, descrevem em geral características nominais de indivíduos ou objetos, como por exemplo sexo, região de procedência, profissão etc. Variáveis desse tipo podem ser analisadas pelas frequências com que ocorrem numa dada amostra. As variáveis ordinais permitem uma ordenação, mas não podem ser tratadas como quantidades. Um exemplo seria o grau de instrução de um indivíduo, medido por “primeiro grau”, “segundo grau” e “terceiro grau”. Observe que existe uma ordem entre esses valores, mas não se pode dizer que aquele que tem o “segundo grau” tem duas vezes mais instrução do que aquele que tem o “primeiro grau”.

Quanto às técnicas estatísticas apresentadas, daremos ênfase aos modelos voltados para analisar a associação entre variáveis. Em geral, a análise da associação entre variáveis permite verificar se o valor de uma certa variável X está relacionado com o valor que outra variável Y assume, para um mesmo elemento.

O capítulo será dividido em quatro seções. Na Seção 27.2 será discutida a análise de dados categóricos, utilizando-se as tabelas de contingência e algumas medidas de associação de variáveis categóricas. A Seção 27.3 será dedicada às primeiras medidas de associação para variáveis quantitativas e à análise de correlações. As duas últimas seções são dedicadas aos modelos de análise estatística multivariada. Serão discutidos os modelos de Análise de Componentes Principais e Análise de Grupamento (em inglês, *Cluster Analysis*).

27.2 Associação de Variáveis Categóricas

Variáveis categóricas são variáveis nominais que, em geral, dividem uma amostra em categorias, como, por exemplo, sexo ou estado civil ou região geográfica etc. O instrumento que permite analisar simultaneamente duas variáveis categóricas é a *Tabela de Contingência*. Esta é uma tabela de frequência de dupla entrada construída conforme a definição a seguir.

Definição 1: Sejam X e Y duas variáveis categóricas assumindo as categorias x_1, x_2, \dots, x_p e y_1, y_2, \dots, y_q , respectivamente. Considere que X e Y foram observadas sobre n objetos, e seja n_{ij} o número de indivíduos que assumem a categoria simultaneamente. É possível dispor essas informações numa *tabela de contingência*:

	$y_1 \dots y_q$	
x_1	$n_{11} \dots n_{1q}$	n_1
\vdots	$\vdots \ddots \vdots$	\vdots
x_p	$n_{p1} \dots n_{pq}$	n_p
$n_1 \dots n_q$		n

$$\text{onde: } n_{\cdot i} = \sum_j n_{ij} \quad \text{e} \quad n_{\cdot j} = \sum_i n_{ij}$$

A tabela de contingência pode ser usada para investigar a associação entre as variáveis X e Y . Essa análise é baseada no fato de que se X e Y são independentes, então

$$n_{ij}/n \cong (n_{\cdot i}/n)(n_{\cdot j}/n).$$

A partir dessa ideia, foi definida a estatística χ^2 que mede o grau de associação entre duas variáveis categóricas:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - (n_{\cdot i} \cdot n_{\cdot j} / n))^2}{n_{\cdot i} \cdot n_{\cdot j} / n}$$

Note que o numerador da estatística χ^2 será próximo de zero se X e Y forem independentes. Por outro lado, quanto maior a associação entre X e Y maior o valor da estatística χ^2 . Assim, essa estatística serve para medir o grau de associação entre duas variáveis categóricas. A partir dessa estatística é possível estabelecer um teste para independência entre duas variáveis (o teste χ^2 de independência). No entanto, como medida de associação, essa estatística nem sempre é adequada por não ser limitada superiormente, isto é, não existe um valor que corresponda à máxima associação. Para resolver esse problema, K. Pearson propôs o seguinte coeficiente:

Definição 2: Sejam X e Y duas variáveis aleatórias categóricas assumindo as categorias x_1, x_2, \dots, x_p e y_1, y_2, \dots, y_q respectivamente. Suponha que X e Y foram observadas sobre n elementos, e suponha que foi obtida a estatística do χ^2 . O coeficiente de contingência C é definido por:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Esse coeficiente é nulo quando X e Y são perfeitamente independentes e aumenta na medida em que a associação entre X e Y aumenta. Além disso, é possível provar que C nunca assume valores maiores que a unidade. Mas mesmo no caso da associação entre X e Y ser perfeita, C pode não atingir o valor 1. Na verdade, o valor máximo de C depende da dimensão da tabela de contingência considerada, e é dado por:

$$C_{\max} = \sqrt{\frac{k-1}{k}}$$

onde k é o mínimo entre o número de categorias de X e o número de categorias de Y . Note que se o número de categorias de X e Y for suficientemente grande, C aproxima-se do valor 1. Assim, o coeficiente de contingência assume valores em uma faixa limitada,

$$0 \leq C \leq \sqrt{\frac{k-1}{k}}$$

sendo que:

$$C \cong 0 \Rightarrow \text{independência} \quad C \cong \sqrt{\frac{k-1}{k}} \Rightarrow \text{associação perfeita}$$

Exemplo 1

Num estudo sobre a informatização da indústria cafeeira no estado de Minas Gerais realizado por Zambalde, A. L. (2000), foram obtidas diversas informações sobre as empresas do setor, tais como: tipo de organização, tipo de administração, número de sacas comercializadas, número de funcionários, grau de escolaridade dos funcionários, tempo em que a empresa usa equipamentos de informática, tipos de equipamentos etc. Deste banco de dados, tomemos, por agora, as seguintes variáveis categóricas:

ORGANIZAÇÃO – assumindo as categorias: Produtoras, Cooperativas, Torrefadoras e Solubilizadoras.

ADMINISTRAÇÃO – assumindo as categorias: Contratada, Cooperada, Familiar e Societária.

A seguir, apresentamos a tabela de contingência, a estatística χ^2 e o coeficiente de contingência C .

TABELA 27.1 Tabela de Contingência

ORGANIZAÇÃO	ADMINISTRAÇÃO				Total
	contratada	cooperada	familiar	societária	
Cooperativa	5	15	0	0	20
Produtora	4	0	6	0	10
Solubilizadora	2	0	0	0	2
Torrefadora	1	0	21	9	31
Total	12	15	27	9	63

$$\begin{aligned}\chi^2 &= 69,76; \\ C &= 0,725 \quad (C_{\max} = 0,866).\end{aligned}$$

Apenas observando a Tabela 27.1, nota-se que administração “cooperada” e “societária” só ocorre nas “cooperativas” e nas “torrefadoras”, respectivamente. As “produtoras” têm administração ou “contratada” ou “familiar”, mas as “torrefadoras” em geral adotam a administração “familiar”. O coeficiente de contingência é alto, indicando forte dependência entre o tipo de organização e o tipo de administração adotada.

Ainda examinando o mesmo banco de dados, vamos agora verificar se existe associação entre o tipo de organização e o número de sacas comercializadas. Neste caso, temos uma variável categórica (organização) e uma variável contínua (número de sacas). Para efetuar essa análise devemos transformar a variável contínua em variável categórica, conforme mostramos a seguir.

ORGANIZAÇÃO – Produtoras, Cooperativas, Torrefadoras, Solubilizadoras.

NÚMERO DE SACAS – Pequeno (até 5.000)

Médio (de 5.001 a 90.000)

Grande (maior que 90.000)

A seguir apresentamos a tabela de contingência, a estatística χ^2 , e o coeficiente de contingência C.

TABELA 27.2 Tabela de Contingência

ORGANIZAÇÃO	NÚMERO DE SACAS			Total
	grande	médio	pequeno	
Cooperativa	17	3	0	20
Produtora	2	6	2	10
Solubilizadora	0	2	0	2
Torrefadadora	2	13	16	31
Total	21	24	18	63

$$\chi^2 = 42,33; \\ C = 0,634 (C_{\max} = 0,816).$$

Olhando a Tabela 27.2, vemos que as “cooperativas” estão concentradas na categoria “grande”, as “produtoras” são em geral classificadas em “médio”, e as “torrefadoras” estão concentradas nas categorias “médio” e “pequeno”. Como as categorias “pequeno”, “médio” e “grande” estão distribuídas de forma quase uniforme na amostra total, essa leitura já permite concluir que as variáveis não são independentes. O coeficiente de contingência fornece uma medida bastante alta para o grau de associação entre as variáveis. Esse resultado nos permite concluir que a quantidade de café comercializada depende do tipo de organização considerada.

É interessante ressaltar que nas duas análises mostradas anteriormente, a leitura da tabela permitiu verificar a existência de dependência entre as variáveis. O coeficiente C forneceu o grau de associação existente. Portanto, esses instrumentos se completam na análise da associação entre variáveis categóricas. Na seção seguinte, passaremos à análise da associação de variáveis quantitativas.

27.3 Associação de Variáveis Quantitativas

Variáveis quantitativas são as mais adequadas para a formulação de modelos estatísticos, e para se convencer disso basta considerar a enorme quantidade de estatísticas definidas para esse tipo de variáveis. A seguir definiremos algumas das estatísticas mais utilizadas para análise univariada, e depois passaremos às estatísticas usadas para avaliar a associação de variáveis quantitativas.

Definição 3: Seja X uma variável quantitativa observada para n elementos. A sua média amostral é definida por:

$$\bar{X} = \frac{1}{n} \left(\sum_i X_i \right)$$

Esta é uma medida de tendência central de uma série de dados. Informa em torno de que valor os dados estão dispostos. Outras medidas muito utilizadas são a variância e o desvio padrão, definidas a seguir.

Definição 4: Seja X uma variável quantitativa observada para n elementos. A sua variância amostral é definida por:

$$Var(X) = \frac{1}{n} \left(\sum_i (X_i - \bar{X})^2 \right)$$

Definição 5: Seja X uma variável quantitativa observada para n elementos. O seu desvio padrão amostral é definido por:

$$DP(X) = \sqrt{Var(X)}$$

Essas duas últimas estatísticas informam sobre a dispersão de uma variável, isto é, sobre o espalhamento de uma série de dados.

Voltando agora ao problema de analisar a associação entre duas variáveis quantitativas, o coeficiente de correlação amostral é a medida mais utilizada. Esta medida é baseada no grau de relação linear entre duas variáveis X e Y . Antes de definir o coeficiente de correlação, é necessário definir a covariância de duas variáveis quantitativas.

Definição 6: Sejam X e Y duas variáveis quantitativas observadas para n elementos. A covariância amostral de X e Y é definida por:

$$S_{xy} = \frac{1}{n} \left(\sum_i (X_i - \bar{X})(Y_i - \bar{Y}) \right)$$

sendo \bar{X} e \bar{Y} as médias amostrais de X e Y , respectivamente.

A covariância de X e Y é uma medida da *relação linear* entre essas variáveis. Valores negativos indicam relação linear inversa, isto é, valores altos para X estão associados a valores baixos para Y . Valores positivos de S_{xy} indicam relação linear direta, ou seja, valores altos de X correspondem a valores altos de Y . Valores de S_{xy} próximos de zero indicam ausência de relação linear entre as variáveis. É possível mostrar que se X e Y são independentes, então $S_{xy}=0$. A recíproca não é verdadeira. Isto é, pode ocorrer $S_{xy}=0$ e X , Y não serem independentes. Na verdade basta que a relação entre as variáveis seja não linear, para que isso ocorra.

Pela definição da covariância amostral pode-se notar que seu valor depende da unidade de medida utilizada. Para se obter uma medida de associação que independa da unidade de medida das variáveis, foi definido o coeficiente de correlação amostral.

Definição 7: Sejam X e Y duas variáveis quantitativas observadas para n elementos. O coeficiente de correlação amostral de X e Y é definido por:

$$r_{xy} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{n DP(X) DP(Y)},$$

sendo $DP(X)$ e $DP(Y)$ os desvios-padrão amostrais de X e Y , respectivamente. O coeficiente de correlação amostral é definido de tal forma que $-1 \leq r_{xy} \leq 1$.

A interpretação do coeficiente de correlação de X e Y é baseada no fato de que, se $r_{xy}=\pm 1$, existe relação linear perfeita entre as variáveis (associação direta se $r=+1$, associação inversa se $r=-1$); se $r_{xy}=0$, não existe relação linear entre as variáveis. Dessa forma, este coeficiente fornece um *grau de associação linear* entre as variáveis.

Na análise de correlação amostral usualmente é feito um teste de hipótese (para ausência de correlação) sobre cada coeficiente estimado. Uma apresentação detalhada sobre testes estatísticos está fora do escopo deste capítulo. O leitor interessado encontrará uma boa introdução sobre o assunto em Bussab, W. e Morettin, P. (2005). Diremos apenas que em um teste de hipótese, fazemos uma afirmação sobre alguma característica da variável de estudo e verificamos se esta afirmação é válida, tendo como base as informações amostrais, isto é, os valores observados para a variável em questão. No caso do teste para correlação a hipótese estatística mais frequente é: “não existe correlação entre as variáveis”, ou melhor, $r_{xy}=0$. Portanto, uma vez realizado o teste de hipótese, é possível verificar se existe ou não correlação entre cada par de variáveis e, no caso de existir, qual o grau de associação de cada par, avaliado pelo valor do coeficiente estimado.

Frequentemente, em análise de dados, trabalhamos com um número $p > 2$ de variáveis. Nesse caso, as covariâncias e correlações podem ser calculadas para todos os pares possíveis de variáveis, e é usual apresentá-las sob a forma de matrizes:

$$S(p \times p) = \begin{bmatrix} S_{11} & \dots & S_{1p} \\ \vdots & \ddots & \vdots \\ S_{p1} & \dots & S_{pp} \end{bmatrix} \quad R(p \times p) = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}.$$

As matrizes de covariância e de correlação são matrizes simétricas cujos elementos S_{ij} no caso da matriz de covariância, ou r_{ij} no caso da matriz de correlação, fornecem a covariância e a correlação entre as variáveis X_i e X_j , respectivamente. A diagonal da matriz de covariância fornece as variâncias das variáveis, ao passo que a matriz de correlação tem diagonal unitária. Essas duas matrizes podem ser usadas como instrumento de análise, e podem também servir de base para outros métodos estatísticos, como será visto no modelo de Análise de Componentes Principais, que será apresentado na Seção 27.4.

Exemplo 2

Voltando ao estudo sobre a informatização da indústria cafeeira no estado de Minas Gerais, nessa seção serão analisadas algumas das variáveis contínuas observadas na pesquisa. O universo de análise é composto de 63 empresas entre cooperativas, produtoras, torrefadoras e solubilizadoras. Utilizaremos a matriz de correlação como instrumento de análise. As variáveis escolhidas para essa aplicação foram:

nsacas – número de sacas de 60 kg de café comercializadas;

nfuncio – número de funcionários que trabalham na empresa;

estudo – é a média dos anos de estudo dos funcionários da empresa;

tempo – indica o tempo (em anos) que a empresa já vem usando equipamentos de informática;

equip1 – quantidade de microcomputadores em rede + quantidade de microcomputadores independentes;

equip2 – quantidade de impressoras (matriciais, jato de tinta, laser).

TABELA 27.3 Matriz de Correlação (N = 63)

	nsacas	nfuncio	estudo	tempo	equip1	equip2
nsacas	1,00	0,615*	0,237	0,406*	0,787*	0,741*
nfuncio		1,00	0,122	0,558*	0,717*	0,746*
estudo			1,00	0,271*	0,247	0,273*
tempo				1,00	0,457*	0,532*
equip1					1,00	0,843*
equip2						1,00

*(rejeitada a hipótese de não correlação no nível de 5 %, isto é, provavelmente existe correlação).
negrito (coeficientes > 0,70, indicando forte correlação).

A estrutura das correlações é bastante forte, e todas as variáveis apresentam a maioria dos coeficientes de correlação significativamente diferentes de zero. As maiores correlações foram observadas entre “nsacas” e “nfuncio” com “equip1”, “equip2”, e também entre os diferentes tipos de equipamentos. Isto indica que as maiores empresas tendem a ter mais equipamentos, e quem tem mais equipamento de certo tipo tende a ter mais equipamento do outro tipo. Observa-se também uma correlação moderada entre “tempo” e “nsacas”, “nfuncio”, “equip1”, “equip2”, mostrando uma tendência moderada das empresas maiores possuírem os equipamentos dos dois tipos, há mais tempo.

Nesse exemplo, a análise de associação entre as variáveis foi feita diretamente da matriz de correlação. Se, no entanto, o número de variáveis cresce muito, a análise dessa matriz torna-se muito difícil. Na próxima seção, será apresentado um modelo estatístico adequado para analisar as associações das variáveis, dispensando a inspeção direta da matriz de correlação.

27.4 Análise de Componentes Principais

Considere um conjunto de p variáveis observadas sobre n elementos. Já vimos que podemos dispor essas informações em uma matriz de dados $X(n \times p)$. A análise de componentes principais tem como objetivo descrever a configuração dos elementos no espaço das variáveis. Na Figura 27.1 é apresentado um exemplo em que duas variáveis X_1 e X_2 são observadas sobre sete elementos. Note que esses elementos formam uma “nuvem de pontos” com coordenadas nos eixos X_1 e X_2 .

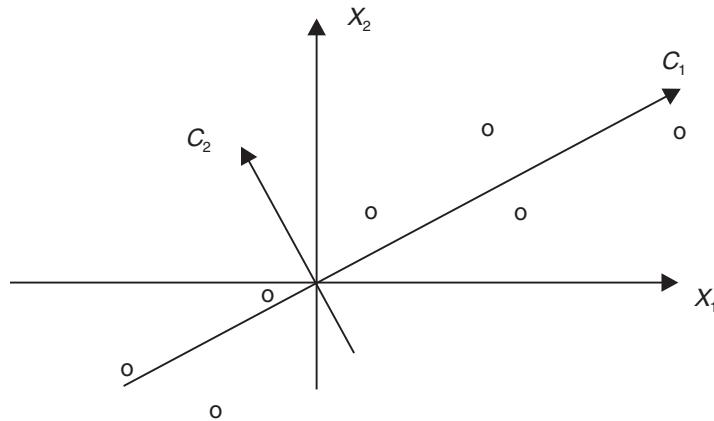


Figura 27.1 As Componentes Principais de um Conjunto de Observações.

As componentes principais C_1 e C_2 do conjunto de pontos estão representadas na figura e fornecem *as direções de maior dispersão* dos pontos observados. Trata-se de uma rotação ortogonal do sistema de referência original (dado pelas variáveis X_1 e X_2), onde C_1 é uma combinação linear de X_1 e X_2 na direção de maior dispersão (ou variância) dos pontos. C_2 é também uma combinação linear de X_1 e X_2 , ortogonal a C_1 .

Voltando à matriz $X(n \times p)$ representando n observações de p variáveis, podemos nos referir a uma nuvem de n pontos no espaço R^p . As componentes principais extraídas serão combinações lineares das p variáveis, tais que C_1 tenha variância máxima; C_2 deve ter variância máxima e ser ortogonal a C_1 (em termos estatísticos, isto equivale à não correlação entre as componentes); assim sucessivamente, até que C_p será uma combinação linear das p variáveis com variância máxima, e ortogonal a C_1, C_2, \dots, C_{p-1} . Então o modelo de análise de componentes principais pode ser escrito pelo seguinte conjunto de equações:

$$\begin{aligned} C_1 &= \sum_j a_{j1} X_j, \quad \text{var}(C_1) = \max \\ C_2 &= \sum_j a_{j2} X_j, \quad \text{var}(C_2) = \max \\ &\quad \text{corr}(C_1, C_2) = 0. \\ &\vdots \quad \vdots \\ &\vdots \quad \vdots \\ C_p &= \sum_j a_{jp} X_j, \quad \text{var}(C_p) = \max \\ \text{corr}(C_k, C_p) &= 0 \quad k = 1, 2, \dots, p-1. \end{aligned}$$

Na extração das componentes C_1, C_2, \dots, C_p , ocorre que as variâncias são decrescentes devido à restrição crescente de não correlação entre elas. Usualmente as últimas componentes têm variâncias tão pequenas que podem ser desprezadas (na Figura 27.1, a dispersão na direção da primeira componente é tão maior que na direção da segunda, que a configuração dos pontos pode ser quase totalmente descrita na direção de C_1). Frequentemente é possível uma redução na dimensão inicial do problema mantendo-se apenas as $q < p$ primeiras componentes, que devem descrever uma “grande percentagem” da variância total do conjunto de p variáveis.

27.4.1 DETERMINAÇÃO DAS COMPONENTES PRINCIPAIS

Considere a primeira componente principal:

$$\begin{aligned} C_1 &= a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p. \text{ Sendo} \\ X &= (X_1, \dots, X_p)' \text{ e } a_1 = (a_{11}, \dots, a_{p1})', \text{ podemos escrever:} \\ C_1 &= a_1'X. \end{aligned}$$

Dessa forma, o problema de determinar a primeira componente passa a ser o de determinar $a_1 \in \mathbb{R}^p$ cuja direção seja a da maior variância do conjunto de pontos. O problema se resolve encontrando a *direção de a_1* , portanto, podemos considerar sua norma unitária, isto é: $a_1'a_1 = 1$.

Sendo $C_1 = a_1'X$, é possível mostrar que: $Var(C_1) = a_1'Sa_1$, onde S é a matriz de covariância amostral de X . Assim o problema a ser resolvido é um problema de *maximização com restrição* que pode ser definido assim:

Determine $a_1 \in R^p$ tal que

$$\begin{aligned} Var(C_1) &= a_1'Sa_1 = \text{max, sujeito a:} \\ a_1'a_1 &= 1. \end{aligned}$$

A solução desse problema pode ser obtida utilizando-se o método dos multiplicadores de Lagrange. A função lagrangeana será:

$$L = a_1'Sa_1 - \lambda_1(a_1'a_1 - 1).$$

Derivando em relação a a_1 e igualando a zero:

$$\begin{aligned} 2(S - \lambda_1 I)a_1 &= 0 \\ Sa_1 &= \lambda_1 a_1. \end{aligned}$$

Esta é a equação característica da matriz S que permite extrair seus autovalores λ_1 e autovetores a_1 . Por outro lado, voltando à expressão para a variância da componente:

$$\begin{aligned} Var(C_1) &= a_1'Sa_1 = a_1'\lambda_1 a_1 = \lambda_1 a_1'a_1, \text{ pois } \lambda_1 \text{ é um escalar;} \\ \text{como } a_1'a_1 &= 1, \text{ então} \\ Var(C_1) &= \lambda_1. \end{aligned}$$

Assim, a variância da primeira componente é um autovalor da matriz S . Como essa variância deve ser máxima, λ_1 é o maior autovalor de S . O vetor a_1 será o autovetor correspondente, com norma unitária.

O procedimento para determinar as demais componentes é o mesmo, e a solução completa é obtida quando são extraídos todos os autovalores e autovetores da matriz S . Portanto, operacionalmente falando, determinar as componentes principais de um conjunto de dados é extrair os autovalores e autovetores de sua matriz de covariância S . Em aplicações do modelo de análise de componentes principais, muitas vezes as variáveis são padronizadas, isto é, são transformadas de tal forma que suas médias são iguais a zero, e suas variâncias são unitárias. Nesse caso, a matriz de covariância se transforma na matriz de correlação. Portanto, é possível resolver o modelo de componentes principais decompondo a matriz de covariância ou a matriz de correlação.

27.4.2 INTERPRETAÇÃO DA SOLUÇÃO DO MODELO DE COMPONENTES PRINCIPAIS

O modelo de análise de componentes principais permite diversas interpretações. Já mencionamos que a solução do modelo se traduz geometricamente como uma rotação ortogonal do sistema de coordenadas fornecido pelas variáveis iniciais X .¹ Aqui ressaltaremos as questões mais básicas do ponto de vista estatístico. Serão apontados os aspectos indispensáveis para se compreender o resultado do modelo de componentes principais. Assim sendo, focalizaremos o papel desempenhado pelos autovalores λ_k e pelos autovetores a_k ($k = 1, 2, \dots, p$), obtidos na solução.

27.4.3 INTERPRETAÇÃO DOS AUTOVALORES

Como já foi visto anteriormente na solução do modelo, a componente C_k terá variância igual ao autovalor λ_k da matriz de covariância (ou correlação). Quando são extraídas todas as p componentes, a variância do conjunto de dados é totalmente reproduzida:

$$\lambda_1 + \lambda_2 + \dots + \lambda_p = \text{variância total} = \text{tr } S,$$

onde $\text{tr } S$ é a soma dos elementos da diagonal da matriz S , isto é, a soma das variâncias de todas as variáveis iniciais X .

Portanto, a medida da importância descritiva de uma componente C_k qualquer é dada pela razão entre a sua variância e a variância total, isto é, a proporção da variância total descrita por ela:

$$\frac{\lambda_k}{\text{tr } S} = \frac{\lambda_k}{\sum_{k=1}^p \lambda_k}.$$

Na aplicação do modelo de componentes principais, frequentemente são desprezadas as últimas componentes devido à pequena participação destas na explicação da variância total. Supondo que são mantidas $q < p$ componentes, a proporção da variância explicada pelas q componentes será:

$$\frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^p \lambda_k} = \text{Proporção da variância explicada.}$$

Esta proporção serve de critério para a escolha do número de componentes mantidas na análise. Deve-se manter um número q de componentes que descrevem uma proporção “razoável” da variância total.

27.4.4 INTERPRETAÇÃO DOS AUTOVETORES

A combinação linear das variáveis que compõem a componente C_k

$$C_k = a_{1k} X_1 + \dots + a_{pk} X_p$$

tem como coeficientes as coordenadas do k -ésimo autovetor. O sinal e a grandeza de a_{jk} indicam o sentido e a contribuição da j -ésima variável para a k -ésima componente. É possível mostrar que a covariância entre a variável X_j e a componente C_k é:

$$\text{cov}(X_j, C_k) = \lambda_k a_{jk}.$$

Assim, quanto maior o coeficiente a_{jk} , maior a associação entre a variável X_j e a componente C_k . Esse resultado é muito importante para interpretação, pois permite identificar as variáveis que estão mais fortemente associadas a cada componente.

Como já foi mencionado, se as variáveis são padronizadas, a solução do modelo é obtida a partir de matriz de correlação. Nesse caso é possível mostrar que:

$$\text{corr}(X_j, C_k) = \sqrt{\lambda_k} \cdot a_{jk}.$$

Novamente, a associação entre as variáveis e as componentes é medida em função dos coeficientes a_{jk} .

Exemplo 3

Nos anos de 1995, 1996 e 1997, foi realizada, com apoio do BNDES, CNI e SEBRAE, uma pesquisa envolvendo os setores da indústria brasileira, contendo informações sobre produtividade, qualidade e recursos humanos, em diversas empresas. Neste exemplo, usaremos os dados relativos ao ano de 1997. Entre as 35 variáveis que compõem a pesquisa, selecionamos 14 delas:

- dias de produção perdidos por interrupções não previstas;
- percentual das exportações na receita operacional líquida (ROL);
- percentual dos salários no custo;
- produtividade = ROL – custo da matéria-prima / número de empregados;
- percentual de pedidos atendidos dentro do prazo de entrega;
- percentual de defeitos;
- dias de estoque do produto acabado;
- percentual da ROL obtido com novos produtos;
- percentual da ROL aplicado na aquisição de equipamentos;
- demissões por estabelecimentos;
- taxa de rotatividade;
- treinamento formal por empregado (horas);
- taxa de acidentes;
- taxa de absenteísmo.

Os setores da indústria incluídos na pesquisa foram: minerais não metálicos; metalurgia; mecânica; materiais elétricos e de comunicações; materiais de transportes; madeira; mobília; borracha; química; papel e papelão; produtos farmacêuticos e

veterinários; perfumaria, sabões e velas; produtos de matéria plástica; têxtil; vestuário, calçados, artefatos de tecido; produtos alimentares; bebidas; editorial e gráfica; diversos.

Considerando essa base de dados, é possível explorá-la segundo o ponto de vista das variáveis (os indicadores de desempenho e recursos humanos), ou do ponto de vista dos elementos (os setores da indústria). O modelo de componentes principais é adequado para a análise das relações entre variáveis, sendo portanto esse ponto de vista adotado agora. Na próxima seção, será desenvolvido um método (método de análise de agrupamento) que permitirá a análise dos setores da indústria.

Os resultados apresentados a seguir foram obtidos com auxílio do *SPSS V8.0*,² a partir das variáveis padronizadas. A Tabela 27.4 apresenta as variâncias das componentes e a proporção da variância total explicada por cada uma delas.

Para interpretar essa tabela, devemos lembrar que, trabalhando com variáveis padronizadas, a variância total das variáveis é igual ao número de variáveis envolvidas, que nesse exemplo é 14. Assim, cada variável é responsável por 1/14 da variância total, que corresponde a 7,1% da variância. Assim, podemos avaliar a agregação obtida pelo modelo, constatando que a primeira componente mantém 27,84% da variância e que apenas as três primeiras componentes mantêm quase 60% da variância total. Essas três primeiras componentes serão retidas para análise. A Tabela 27.5 fornece a matriz das componentes com os valores:

$$\sqrt{\lambda_k} a_{ik} = \text{corr}(X_i, C_k), k = 1, 2, 3; j = 1, 2, \dots, 14$$

Examinando esse resultado, vemos que a componente 1 tem forte correlação positiva com “Dias perdidos”, “Produtividade”, “%Entrega no prazo” e “Treinamento”. Observamos também, forte correlação negativa com “%Salário no custo”, “Defeitos”, “TxAbsenteísmo”. Por isso é possível interpretar essa componente como uma dimensão associada ao *desempenho do setor* (exceto para a variável “Dias perdidos”). Isto porque, setores que apresentarem *altos valores* para “Produtividade”, “%Entrega no prazo”, “Treinamento”, e apresentarem ao mesmo tempo *baixos valores* para “%Salário no custo”, “Defeitos”, “TxAbsenteísmo”, terão maiores valores associados à componente 1, e estes serão os setores com melhor desempenho. Os setores que se comportarem de maneira inversa terão os menores valores para a componente 1.

A componente 2 está associada às variáveis “%Exportação na receita”, “Demissões”, “TxRotatividade” e “TxAcidentes”. Com exceção da primeira variável, essa componente descreve a dimensão de *condições para o trabalho*. A componente 3 opõe “%Exportação” e “Treinamento” a “%Entrega no prazo”, “Estoque de produto” e “%Receita aplicada a equipamentos”, diferenciando dois tipos de setores: um caracterizado pelas exportações e que pratica o treinamento dos funcionários, e outro que dá prioridade à entrega no prazo, mantendo estoque e aplicando em equipamentos. Vale ressaltar que estas duas últimas componentes têm variâncias bem menores que a primeira (ver na Tabela 27.4), o que necessariamente torna essas associações e tendências mais fracas.

TABELA 27.4 Variância Explicada

Componente	$\lambda = \text{Variância}$	% Variância	% Acumulado
1	3,898	27,84	27,84
2	2,128	15,20	43,04
3	1,983	14,17	57,21
4	1,777	12,69	69,89
5	1,095	7,82	77,71
6	0,790	5,64	83,36
7	0,639	4,57	87,93
8	0,499	3,57	91,49
9	0,385	2,75	94,25
10	0,282	2,01	96,23
11	0,193	1,38	97,63
12	0,137	0,98	98,61
13	0,120	0,86	99,47
14	0,074	0,53	100

TABELA 27.5 Matriz das Componentes

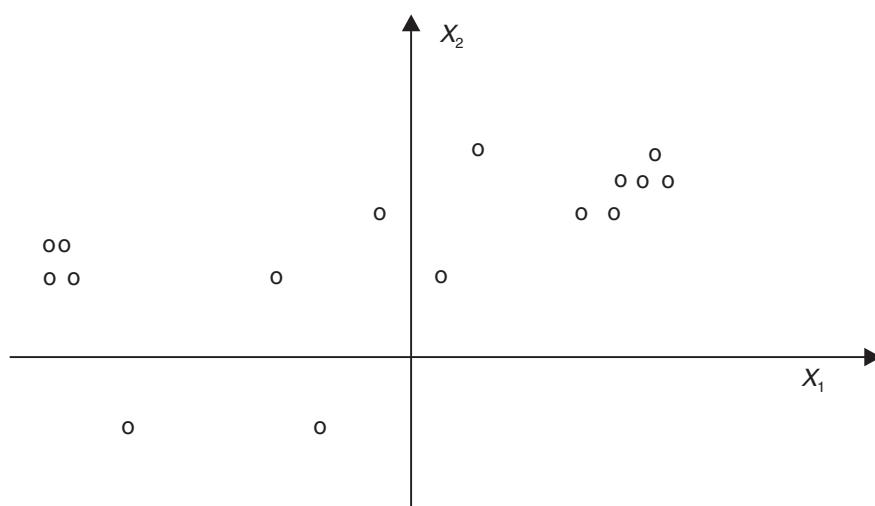
	Componente 1	Componente 2	Componente 3
Dias perdidos	0,777	0,077	-0,079
%Exp na receita	-0,205	0,576	-0,674
%Salário no custo	-0,595	-0,149	-0,064
Produtividade	0,847	-0,162	-0,296
%Entrega no prazo	0,411	0,035	0,534
%Defeitos	-0,757	-0,047	-0,274
Estoque de produto	0,331	0,210	0,453
%Rec c/ Nprodutos	0,112	0,299	0,227
%Rec ap/Equipam	-0,233	0,202	0,692
Demissões	0,145	0,656	0,011
TxRotat	-0,256	0,678	0,362
Treinamento	0,734	0,237	-0,400
TxAccidente	0,018	0,774	-0,183
TxAbsenteísmo	-0,803	0,136	-0,101

27.5 Análise de Grupamento

Antes de explicar a técnica aqui denominada de análise de grupamento, advertimos o leitor que essa mesma técnica pode aparecer na literatura com diversos nomes. Pode ser chamada de análise de *cluster*, análise de agrupamentos, ou ainda análise de conglomerados. Adotamos aqui o termo usado em 1982 por Lucas, L.C.S, onde é feita uma descrição bastante detalhada desta técnica. A seguir, introduziremos a ideia da análise de grupamento examinando um exemplo muito simples.

A Figura 27.2 representa as medidas das variáveis X_1 e X_2 sobre um conjunto de elementos. Observamos dois pequenos grupos de pontos próximos: um no primeiro quadrante com 6 pontos, e outro no segundo quadrante com 4 pontos. No que se refere às variáveis X_1 e X_2 , podemos dizer que os 6 pontos no primeiro quadrante são semelhantes. O mesmo podemos dizer dos 4 pontos no segundo quadrante. Além disso, podemos ainda afirmar que esses dois grupos são diferentes.

Essa figura permite observar a *configuração dos elementos no espaço gerado pelas variáveis X_1 e X_2* . Evidentemente, essa análise visual só foi possível por estarmos considerando duas variáveis. Considerando um conjunto de p variáveis, sendo $p > 2$,


Figura 27.2 Representação de um Conjunto de Elementos no Espaço das Variáveis X_1 e X_2 .

a configuração dos elementos no espaço p -dimensional não permite mais uma análise visual. A análise de agrupamento tem por objetivo identificar grupos de elementos semelhantes no espaço das variáveis, assim mesmo, como acabamos de fazer aqui, só que dispensando a visualização dos pontos, isto é, permitindo a utilização de $p > 2$ variáveis. Assim, a análise de agrupamento procura identificar a estrutura de grupos dos elementos, segundo um critério definido pelo conjunto de p variáveis observadas. O problema da análise de agrupamento pode ser colocado da seguinte forma:

Seja $X = \{X_1, \dots, X_p\}$ um conjunto de variáveis, e $E = \{e_1, \dots, e_n\}$ o conjunto de elementos que se deseja agrupar.

Com base no conjunto X , determinar uma partição de E em grupos g_i tal que:

se $e_r, e_s \in g_i \Rightarrow e_r, e_s$ são semelhantes;

se $e_r \in g_i, e_s \in g_j \Rightarrow e_r, e_s$ são distintos.

Para a solução desse problema é necessário uma medida de similaridade entre os elementos. Usa-se frequentemente a distância entre os elementos no espaço das variáveis como medida de similaridade. Pensando em um conjunto de n elementos, é possível construir uma matriz que fornece todas as distâncias entre eles. Essa *matriz de distância* entre os elementos é o ponto de partida de diferentes métodos utilizados para resolver o problema de análise de agrupamento.

A distância entre os elementos pode ser calculada de várias maneiras. Examinemos a métrica de Minkowski, que fornece uma forma geral de distância entre dois elementos.

Sejam dois elementos da matriz de dados:

$e_r = (X_{1r}, \dots, X_{pr})$, $e_s = (X_{1s}, \dots, X_{ps})$, a distância de Minkowski entre esses elementos é dada por:

$$d(e_r, e_s) = \left[\sum_{i=1}^p |X_{ir} - X_{is}|^m \right]^{1/m}.$$

Note que, se $m = 2$, a expressão anterior se transforma na distância euclidiana, e por isso podemos dizer que esta é uma forma generalizada de distância.

Uma vez escolhida uma medida de distância, é necessário um procedimento que, examinando essas distâncias, determine uma partição do conjunto dos elementos, resolvendo o problema de análise de agrupamento. Existem vários métodos para resolver o problema. A seguir focalizaremos alguns métodos hierárquicos aglomerativos. O leitor interessado em outros tipos de métodos deve procurar na Seção 27.6 deste capítulo as referências indicadas.

27.5.1 MÉTODOS HIERÁRQUICOS AGLOMERATIVOS

O método hierárquico aglomerativo, como a maioria dos métodos de análise de agrupamento, parte de uma matriz de distância entre os elementos. De um modo geral, são métodos iterativos, com número de iterações igual ao número de elementos. Na inicialização do método, cada elemento é considerado como um grupo separado, ou seja, inicialmente contamos com n grupos. A seguir, os dois grupos mais próximos, segundo a métrica escolhida, formarão um novo grupo, e nessa iteração contamos $n-1$ grupos. As distâncias são devidamente recalculadas, e novamente os dois grupos mais próximos formam um grupo, e nesse momento contamos com $n-2$ grupos. Esse procedimento se repete, até que tenhamos um único grupo contendo os n elementos. A seguir mostramos um pequeno exemplo.

Exemplo 4

Considere duas variáveis observadas sobre cinco elementos e seja a matriz de distância dada a seguir:

	E1	E2	E3	E4	E5
E1	0	2	5	7	10
E2		0	6	3	8
E3			0	7	4
E4				0	6
E5					0

Vemos que os elementos mais próximos são E1 e E2. No próximo passo esses dois elementos aparecem juntos, formando o primeiro grupo, e a nova matriz de distância será:

	(E1,E2)	E3	E4	E5
(E1,E2)	0	5	3	8
E3		0	7	4
E4			0	6
E5				0

Nesse passo forma-se o grupo (E1,E2,E4) e a nova matriz será:

	(E1,E2,E4)	E3	E5
(E1,E2,E4)	0	5	6
E3		0	4
E5			0

O próximo grupo será (E3,E5) e a matriz de distância será:

	(E1,E2,E4)	(E3,E5)
(E1,E2,E4)	0	5
(E3,E5)		0

A solução final pode ser representada num diagrama em árvore chamado *dendrograma* (Figura 27.3):

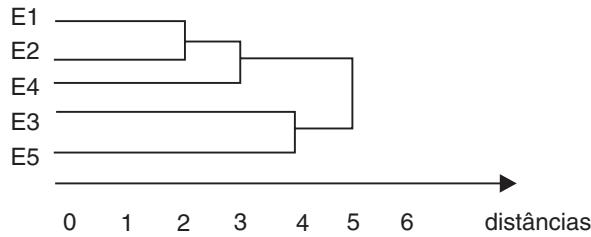


Figura 27.3 Solução da análise de agrupamento.

A escolha do número de grupos da solução final faz parte da interpretação do resultado. Isto é, o método fornece diversas partições do conjunto de elementos. A escolha da “melhor” partição é feita pelo analista. Nesse exemplo, percebe-se dois grupos distintos, um formado por E1, E2 e E4; o outro, com menor similaridade, formado por E3 e E5.

Na descrição de um método de análise de agrupamento é necessário definir como *atualizar as sucessivas matrizes de distância*. No exemplo anterior, a distância entre os grupos foi definida pela menor das distâncias entre os elementos dos grupos. Esse é um dos métodos para análise de agrupamento, chamado de “ligação simples”. Escolhemos esse método para o exemplo citado principalmente pela sua simplicidade no cálculo das novas distâncias. Outros métodos vêm se mostrando mais eficientes na identificação de estruturas de grupos. Neste capítulo, descreveremos brevemente alguns dos métodos hierárquicos mais usados.

O Método de Ligação Simples – Esse método consiste em atualizar a matriz de distância, considerando a distância entre grupos como a menor distância entre os elementos dos grupos. Isto é, se o grupo 1 é composto dos elementos a, b , e se o grupo 2 é composto dos elementos c, d , então a distância entre os dois grupos será o menor valor entre $d(a,c), d(a,d), d(b,c)$ e $d(b,d)$. Esse foi o método usado no exemplo citado.

O Método do Centroíde – Neste método, cada grupo é representado por um ponto (o centroíde), cujas coordenadas no espaço das variáveis é a *média das variáveis dos elementos do grupo*. Dessa forma, a cada iteração são calculadas as distâncias entre os centroides e serão aglomerados os dois grupos cujos centroides estão mais próximos.

O Método de Ward – A ideia do método de Ward é avaliar os grupos pela sua dispersão. A dispersão *dentro* de cada grupo é medida pela soma dos desvios quadráticos entre seus elementos. Isto é, se um grupo tem q elementos a dispersão dentro dele será:

$$SQD = \sum_i^q [d(e_i, o)]^2,$$

onde o é o centroíde do grupo.

O método de Ward, em cada iteração, busca juntar dois grupos tais que o acréscimo em SQD seja mínimo.

Exemplo 5

Nesse exemplo analisaremos a mesma matriz de dados utilizada na aplicação do modelo de componentes principais na Seção 27.4, só que agora buscaremos grupos de setores da indústria brasileira semelhantes quanto aos indicadores observados. Utilizaremos aqui o mesmo conjunto de 14 indicadores (variáveis) e os mesmos 19 setores da indústria. Foi escolhido o método do centroide com distância euclidiana e os resultados foram obtidos com o Software *SPSS V8.0*. Apresentamos a seguir a solução na forma de dendrograma.

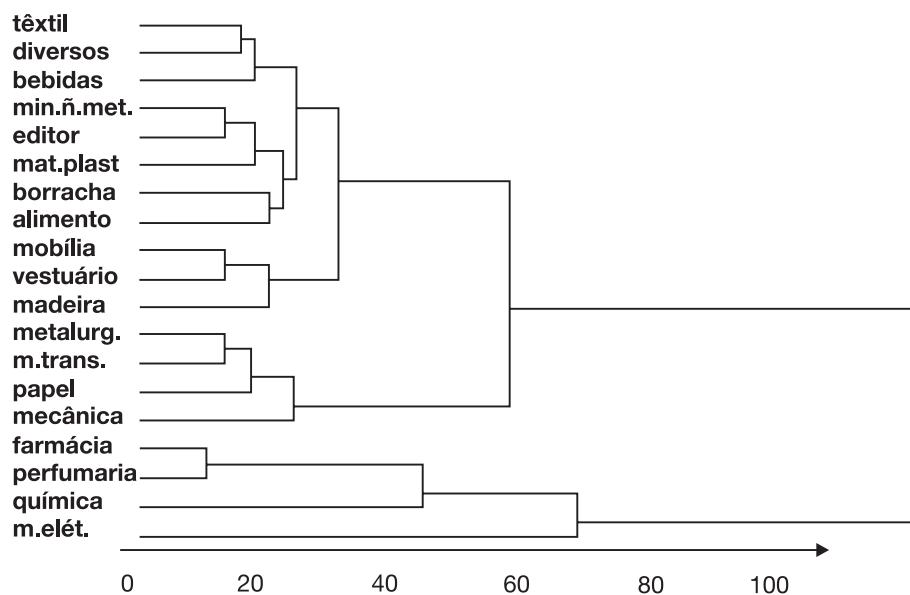


Figura 27.4 Análise de Grupamento para os Setores da Indústria Brasileira – 1997.

Analizando o dendrograma anterior, notamos nitidamente a formação de dois grandes grupos. O primeiro composto dos setores Farmácia, Perfumaria, Química e Material elétrico, estando este último um tanto separado dos demais. O segundo grande grupo é formado por todos os demais setores que compõem o universo de análise. Assim, chama a atenção a distinção dos setores de Farmácia, Perfumaria, Química e Materiais elétricos dos demais setores da indústria. É razoável supor que os indicadores usados na análise (as variáveis) têm um comportamento diferente nestes setores, se comparados com os outros. Uma possível hipótese que pode ser levantada diante desse resultado é que foram destacados os setores mais modernos da economia (Farmácia, Perfumaria, Química, Material elétrico). No entanto, seria necessário a utilização de outras técnicas de análise de dados para confirmar ou não essa conjectura.

Voltando agora nossa atenção para o segundo grande grupo formado, vemos que, dentro deste, outros grupos também são formados. Chama a atenção o destaque de Metalurgia, Material de Transporte, Papel e Mecânica dos demais setores. O próximo grupo, composto de 11 setores, mostra-se bastante homogêneo (pois se juntam em um nível de distância baixo comparado com a distância máxima do dendrograma), porém ainda podemos observar o alto grau de similaridade dos setores Mobiliário, Vestuário e Madeira.

27.6 Conclusão

Neste capítulo apresentamos brevemente alguns modelos estatísticos para associação de variáveis. Ao leitor interessado em aprofundar um pouco mais seu estudo, apresentamos algumas referências básicas e comentários por assunto aqui abordado. Assim recomendamos, para outras análises da tabela de contingência definida na Seção 27.2, os livros de Benzecri, J. P. et al. (1980), e Bourassa, J. M. e Saporta, G. (1982). Mais detalhes sobre o modelo de componentes principais podem ser encontrados no capítulo 3 do livro de Mingoti, S. A. (2007) ou no livro de Jolliffe, I. T. (1986), este último todo dedicado à análise de

componentes principais. Quanto aos métodos de análise de agrupamento, a abordagem mais completa está no livro de Anderberg (1973). Recomendamos também o capítulo 12 do livro de Johnson, R. A. e Wichern, D. W. (1992), o capítulo 6 do livro de Mingoti, S. A. (2007), e ainda o trabalho de Lucas, L. C. S. (1982), onde diversos métodos são detalhadamente apresentados. Especialmente para alunos de economia, recomendamos o trabalho de Barros, A. C. (1992), onde é apresentada uma análise da matriz de insumo-produto usando métodos de análise de agrupamento, e Kubrusly e Sabóia, 2006, onde as técnicas de análise de correlações, análise de componentes principais e análise de agrupamento são aplicadas sobre dados da PNAD/IBGE.

Notas

1. Outros aspectos geométricos desse modelo podem ser vistos em Jolliffe (1986).
2. Versão 8.0 do software Pacote Estatístico para as Ciências Sociais (*Statistical Package of Social Science – SPSS*).

Bibliografia

- Anderberg, M. R. (1973) *Cluster Analysis for Applications*. Academic Press.
- Barros, A. C. (1992) *Relações Intersetoriais em Matrizes de Insumo-Produto: uma abordagem da análise de agrupamento*. Dissertação de mestrado, COPPE/UFRJ.
- Benzecri, J. P. et al. (1980) *La Pratique de L'analyse des Données*, Dunod, Paris.
- BNDES, CNI, SEBRAE (1997) *Indicadores de qualidade e produtividade na indústria brasileira*.
- Bourouche, J. M.; Saporta, S. (1982) *Análise de Dados*, Zahar Editores.
- Bussab, W. O.; Morettin, P. A. (2005) *Estatística Básica*, Editora Saraiva.
- Johnson, R. A.; Wichern, D. W. (1992) *Applied Multivariate Statistical Analysis*, Prentice-Hall.
- Jolliffe, I. T. (1986) *Principal Components Analysis*, Springer-Verlag.
- Kubrusly, L. S.; Saboia, J. (2006) *Uma Análise Multivariada da População Ocupada nas Regiões Metropolitanas Brasileiras* Ensaios FEE – v. 27, n. 2.
- Lucas, L. C. S. (1982). *Revista Brasileira de Estatística*, n. 172, ano XLIII.
- Mingoti, S. A. (2007). *Análise de Dados Através de Métodos de Estatística Multivariada - uma abordagem aplicada*, Editora UFMG.
- SPSS V 8.0 (1998). User's Guide, SPSS Inc.
- Zambalde, A. L. (2000). *A Informática na Modernização do Sistema Agroindustrial do Café no Estado de Minas Gerais*, Tese D.Sc. COPPE/UFRJ.



Capítulo 28

Fontes de Informação sobre a Indústria Brasileira¹

Lia Haguenauer e Magdalena Cronemberger Góes

28.1 Introdução

A tradução empírica dos elementos teóricos abordados nos diversos capítulos deste livro nem sempre é viável. A teoria, em geral, diz respeito a comportamentos e relações estabelecidas *ex ante* entre agentes e variáveis específicas, ao passo que as estatísticas representam o resultado *ex post* da interação de todos os fatores que podem afetar o fenômeno em questão. As estatísticas correspondem a observações de variáveis em um ponto do tempo, possibilitando apenas a inferência de relações ou comportamentos, mediante a comparação de variações no tempo (séries temporais) e/ou entre diferentes agentes (comparações entre regiões geográficas, empresas de diferentes tamanhos, tipos de propriedade do capital etc.).

Essa tradução exige ainda a correspondência precisa entre o fenômeno que se deseja observar e a base de dados a ser utilizada, notadamente quanto a: agentes – definição e seleção das unidades de informação às quais se referem às estatísticas – e classificações – critérios utilizados para o agrupamento dos agentes e dos produtos –, além da escolha e conceituação das variáveis investigadas.

Neste capítulo apresentamos as principais fontes de informação atualmente disponíveis sobre a indústria brasileira, fazendo antes breve comentário sobre agentes e indústrias e mercados, conceitos já discutidos no Capítulo 2, e de particular importância na operacionalização das estatísticas industriais.

28.2 Conceitos Básicos nas Estatísticas Industriais

28.2.1 AGENTES

No que se referem aos agentes, os principais conceitos da Economia Industrial dizem respeito a empresas e plantas industriais.

Dada a variedade de formatos de organização e operação dos agentes produtores de bens e serviços, a construção de estatísticas consistentes e comparáveis internacionalmente da atividade econômica que realizam requer que sejam definidas de forma padrão as unidades a partir das quais serão levantadas as informações. São as chamadas unidades estatísticas de investigação ou de observação. A comparabilidade das estatísticas fica mais bem garantida quando estas unidades são definidas e classificadas de forma padronizada.

Atualmente, nas estatísticas industriais brasileiras, as unidades estatísticas de investigação são:²

1. Empresa – “unidade jurídica caracterizada por uma empresa ou razão social que engloba o conjunto de atividades econômicas exercidas em uma ou mais unidades locais” (IBGE, 2004c, p.16).
2. Unidade local – “espaço físico, geralmente uma área contínua, no qual uma ou mais atividades econômicas são desenvolvidas, correspondendo a um endereço de atuação da empresa” (*id. ibid.* p.16).³

As empresas são as unidades reais que operam no mercado. São as unidades de decisão, de existência jurídica, que assumem obrigações financeiras e estão à frente das transações de mercado e, além do mais, é sobre as empresas que recai a obrigatoriedade dos registros contábeis, balanços etc. A empresa constitui, assim, a unidade adequada tanto para análises de comportamento de agentes econômicos (estratégias produtivas, de acumulação e expansão) como para o levantamento de informações econômico-financeiras. Só a este nível é possível a avaliação precisa dos processos de geração e distribuição de renda, abrangendo despesas diretas e indiretas da produção, renda e lucro gerados.

Na maioria dos casos ocorre a coincidência física entre empresa e unidade local/estabelecimento e entre estes e os conceitos teóricos de empresa e de planta industrial. No entanto existe na realidade grande diversidade de tipos de unidades, resultantes de diferentes formas de organização da atividade industrial. Empresas com múltiplas localizações e/ou múltiplas atividades econômicas são bastante comuns no segmento das empresas com maior peso na produção industrial.⁴ A teoria contempla a existência de empresas multiplanta, bem como a ocorrência de diferentes estratégias de diversificação por parte das empresas. Assim, para estudos nos quais sejam relevantes a definição precisa de segmentos produtivos, características técnicas do processo produtivo (como por exemplo, atualização tecnológica e economias técnicas de escala, definidas no nível das plantas), ou ainda a ótica espacial, são necessárias informações de agentes definidos como partições de empresas, ou seja, informações no nível das unidades locais.

É importante notar também que a empresa nas estatísticas (entidade jurídica) muitas vezes não detém a autonomia decisória atribuída à empresa teórica. É crescente a importância de grupos econômicos que abrangem diversas empresas e cuja estrutura jurídica é particular a cada um. Participações cruzadas e percentagens variáveis do capital social determinam, na realidade, o poder de controle e decisão sobre o conjunto do grupo, o que torna esta categoria um agente de difícil identificação para levantamentos estatísticos.⁵ Ademais, a questão da diversificação de atividades também torna seu uso analítico problemático.

28.2.2 INDÚSTRIA E MERCADO

Quanto ao agrupamento dos agentes, os conceitos teóricos relevantes são os de indústria e mercado.

A teoria, em geral, pressupõe uma correspondência biúnivoca entre produtos (conjuntos de bens e serviços) e empresas (ou plantas), ou seja, cada produto é produzido por um conjunto univocamente determinado de empresas. As empresas reais, no entanto, quase sempre produzem mais de um produto, e de um mesmo processo produtivo podem resultar produtos distintos, inclusive característicos de outras atividades.

Para dar conta desta realidade, existem duas dimensões básicas na representação do processo produtivo e, consequentemente, na construção das estatísticas econômicas: de um lado, a atividade econômica levada pelo conjunto de agentes produtivos (empresas, unidades locais/estabelecimentos) e, de outro, os produtos na forma de bens e serviços resultantes do processo produtivo. Cada vertente trabalha com sistemas de classificações específicos e gera conjuntos de estatísticas específicas. Para a organização e análise das estatísticas construídas sob o enfoque do agente produtivo (empresa, unidades locais/estabelecimentos) – dados que permitem analisar a estrutura produtiva e a organização da produção – trabalha-se com a classificação de atividades econômicas. Na vertente dos produtos, as estatísticas referem-se aos fluxos dos mercados de bens e serviços – composição da oferta, fluxos do comércio externo, composição do consumo, índices de preços etc. – e são construídas com base em classificações de produtos.

As classificações servem como sistema de linguagem na produção de informações estatísticas. A uniformização desta linguagem é imprescindível para a articulação das informações no tempo (séries temporais), no espaço (comparações regionais e internacionais) e entre fontes distintas, o que ressalta a importância da padronização tanto no nível nacional como no internacional.

A classificação internacional de referência para as atividades industriais é a *International Standard Industrial Classification –ISIC*,⁶ que abrange todas as atividades econômicas e é elaborada e atualizada periodicamente sob a coordenação da Divisão de Estatísticas da Organização das Nações Unidas - ONU. Deve-se ressaltar que os sistemas de classificação de atividades procuram retratar a organização real das estruturas produtivas, sendo, nesse sentido, historicamente determinados. Assim, precisam ser periodicamente atualizados e revisados em função de mudanças na organização produtiva e, também, da demanda por novas abordagens analíticas.

Da mesma forma, particularidades regionais das estruturas produtivas exigem adaptações da classificação internacional. O Brasil sempre procurou, em linhas gerais, seguir as classificações da ONU, mas apenas a partir da adoção da Classificação Nacional de Atividades Econômicas - CNAE como sistema oficial passou a trabalhar com um padrão predefinido de relacionamento com a classificação internacional que efetivamente garante a comparabilidade internacional das estatísticas econômicas brasileiras. Tal como a ISIC, a CNAE está organizada em uma estrutura hierárquica com vários níveis: seção, divisão, grupo, classe. Os dois primeiros níveis são idênticos aos da classificação internacional e, nos dois seguintes, a CNAE introduz maior detalhamento sempre que necessário para refletir a estrutura da economia brasileira, em princípio possibilitando a reconstituição das categorias da classificação internacional.⁷

Na versão original, a CNAE é uma classificação derivada da ISIC, revisão 3. À versão original seguiram-se duas outras: a CNAE 1.0, atualização ocorrida em 2002, e a CNAE 2.0,⁸ resultado de revisão mais ampla em 2006, ambas acompanhando os movimentos da classificação internacional (ISIC 3.1 e revisão 4, respectivamente).

Atualmente, a CNAE é usada na produção e disseminação de informações por tipo de atividade econômica em todas as estatísticas econômicas, inclusive as Contas Nacionais a partir da série ano-base 2000, e socioeconômicas.

Os agentes (empresa, unidade local/estabelecimento) são classificados segundo a atividade associada aos produtos de maior valor em sua produção.

Os critérios subjacentes às classificações de atividades procuram manter em um mesmo grupo as atividades das quais resultam produtos semelhantes quanto aos processos produtivos (tecnologia, insumos, organização da produção) – caracterizando indústrias segundo a *base técnica* – e também quanto ao destino da produção (agrupando produtores de bens que sejam substitutos próximos), procurando caracterizar *mercados*, conforme discutido no Capítulo 2. A importância conferida aos critérios de base técnica e mercado varia entre classificações e também entre as atividades no interior de um mesmo sistema, podendo assim cada atividade definida estar mais associada ao conceito de mercado ou ao de indústria.

Considerando os grupos (três dígitos) da Classificação Nacional de Atividades Econômicas (CNAE), tem-se, por exemplo, a *fabricação de produtos de material plástico*, atividade definida pela tecnologia e insumos, abrangendo artigos de consumo pessoal, material de embalagem e outros, ao lado da *fabricação de calçados*, que agrupa plantas com insumos e processos distintos, abrangendo a fabricação de calçados de couro, plástico, borracha, têxteis etc. No primeiro caso o critério subjacente à classificação é o da base técnica; no segundo, o mercado.

As variáveis referentes às atividades (conjuntos organizados de agentes produtivos) permitem analisar a estrutura industrial e a organização da produção. As principais variáveis associadas às atividades são: valor da produção, valor adicionado (ou valor da transformação industrial, aproximação do valor adicionado, obtido deduzindo-se do valor da produção industrial apenas os custos diretos), emprego, características dos processos produtivos, formação de capital, custos, receita operacional líquida, *mark up* (receitas operacionais/custos diretos da produção), propriedade do capital.

As estatísticas dos fluxos de bens e serviços são levantadas e disseminadas por meio de classificações de produtos, cujos princípios de agregação variam de acordo com o fluxo enfocado. Assim, são definidas classificações de produtos específicas para a elaboração e análise dos fluxos do comércio externo, dos fluxos de produtos resultantes do sistema produtivo e dos fluxos do consumo das famílias, entre outros. Por outro lado, também para o sistema de classificações de produtos prevalecem os princípios de padronização/compatibilização nos planos tanto nacional como internacional, sendo importante a definição de correspondências inequívocas entre os vários tipos de classificações de produtos.

No Brasil, os fluxos de mercadorias do comércio externo (importação e exportação de mercadorias) são registrados e suas estatísticas geradas com o uso da Nomenclatura Comum do Mercosul – NCM, adotada desde janeiro de 1995 pelos países da Região. A NCM é derivada do Sistema Harmonizado de Designação e de Codificação de Mercadorias, ou simplesmente Sistema Harmonizado – SH, de uso internacional. Dos oito dígitos que compõem os códigos NCM, os seis primeiros são idênticos aos códigos do SH e os dois últimos correspondem a desdobramentos específicos para o Mercosul.

Na organização e disseminação de estatísticas de produção de produtos industriais, o IBGE usa a PRODLIST-Indústria, uma nomenclatura abrangente e detalhada de produtos e serviços industriais produzidos pelas indústrias de mineração e de transformação. Os produtos PRODLIST-Indústria são definidos a partir da NCM, por agregações ou desagregações e são organizados segundo classes CNAE de origem (os quatro primeiros dígitos do código de oito dígitos dos produtos PRODLIST são os da classe CNAE de origem). A PRODLIST-Indústria, tal como a NCM, é atualizada anualmente. A partir da versão 2007, a PRODLIST-Indústria está referenciada à versão 2.0 da CNAE. Tabelas de correspondências entre as versões anuais da PRODLIST-Indústria e destas com a NCM são disponibilizadas pelo IBGE.⁹

As estatísticas de produtos focalizam basicamente o mercado, ou seja, os fluxos de oferta (produção e importação) e de demanda (consumo, investimento, exportações) e as variáveis mais importantes associadas a produtos são: quantidade, valor e preços da produção interna, exportações e importações (em valores, quantidades ou índices).

28.3 Fontes de Informação sobre a Indústria Brasileira

28.3.1 INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

O IBGE é o órgão responsável pelo Sistema Estatístico Nacional e é a instituição que realiza os levantamentos mais abrangentes sobre a indústria brasileira.

O subsistema de pesquisas industriais do IBGE cobre as indústrias extractivas e de transformação e organiza-se dentro de padrão comum às pesquisas do sistema integrado de pesquisas econômicas, implementado a partir de meados dos anos 1990, com destaque dos seguintes parâmetros:

- as pesquisas são desenhadas por amostra levando em conta a concentração da atividade produtiva nos segmentos de maior porte e, consequentemente, dando maior peso à sua representação; as amostras são definidas com base no mapeamento atualizado do universo das entidades empresariais do país do Cadastro Central de Empresas – Cempre, do IBGE, que para tanto articula dados recebidos de registros administrativos e das pesquisas econômicas correntes;
- as pesquisas são organizadas em três conjuntos com funções específicas: as pesquisas estruturais, as pesquisas conjunturais e as pesquisas satélites.

As pesquisas estruturais formam o núcleo central do sistema, com o duplo papel de fornecer dados para a caracterização da atividade produtiva das empresas, fortemente apoiados nos conceitos das Contas Nacionais, e de servir de núcleo de articulação das demais pesquisas. Substituem os censos econômicos. As pesquisas conjunturais estão focadas no acompanhamento do desempenho no curto prazo e elas são a base para a construção dos indicadores de acompanhamento da conjuntura. As pesquisas satélites são temáticas, voltadas ao tratamento de questões relevantes da atividade produtiva.

No subsistema das pesquisas industriais, a Pesquisa Industrial Anual – PIA, reformulada a partir de 1996 para se adequar aos novos parâmetros, desdobra-se em duas pesquisas:

1. A *PIA-Empresa* objetivando a caracterização da atividade das empresas industriais como agentes organizadores da produção e levantando informações econômico-financeiras;
2. A *PIA-Produto* respondendo pelas informações de quantidade e valor dos produtos produzidos nas unidades locais industriais.

A PIA-Empresa é a pesquisa central estrutural do subsistema das pesquisas industriais. Propicia informações relativas à atividade industrial em substituição aos censos industriais – a identificação das características estruturais e acompanhamento de suas transformações no tempo – e constitui o núcleo em torno do qual se articulam as demais pesquisas da indústria, tanto as de acompanhamento conjuntural como as de aprofundamento temático, como é o caso da Pesquisa de Inovação Tecnológica.

As principais características metodológicas da PIA-Empresa são:¹⁰

1. Âmbito da pesquisa: a PIA-empresa cobre as empresas com atividade principal nas indústrias extractivas e de transformação, dentro da definição da CNAE.
2. Unidade estatística: a unidade central de informação/observação e principal foco de atenção é a empresa, o que significa que os quantitativos e valores mais completos informados na pesquisa referem-se à empresa como um todo. Para as empresas industriais com múltiplas unidades, o enfoque centralizado na empresa é complementado com a identificação das unidades locais, por localização geográfica e atividade principal, e o levantamento de um conjunto limitado de variáveis referentes às atividades nelas exercidas, cujos quantitativos e valores são usados para o rateio de variáveis informadas no nível da empresa. As estatísticas regionais ou por atividade econômica com base nas informações no nível da unidade local são, portanto, construções a partir dos dados efetivos no nível das empresas, o que, por outro lado, garante a coerência entre as informações das empresas e sua leitura por atividade industrial e/ou por localização geográfica. Na PIA-Empresa de 1996 em diante, prevalece a lógica de construção dos agregados econômicos (valor da produção, consumo intermediário, valor adicionado, valor de transformação industrial, remuneração de assalariados, contribuições sociais etc.) de cima para baixo: os valores definidos para a empresa como um todo são redistribuídos pelas unidades locais com base no peso das variáveis informadas neste nível. Em relação às pesquisas anteriores – censos e PIAs até 1995 – a mudança na definição da unidade estatística de informação/observação é a alteração metodológica de maior impacto.¹¹
3. Classificação: CNAE: versão original, nas PIAs-Empresa de 1996 a 2002; versão 1.0, nas PIAs 2003 a 2007; versão 2.0, a partir de 2007. Dada a extensão das mudanças na versão 2.0, a transição para a nova versão exigiu cuidados para evitar descontinuidades nas séries históricas. Os resultados da PIA empresa referentes a 2007, ano de transição para a nova versão, foram divulgados nas duas versões 1.0 e 2.0.¹²

4. Desenho da pesquisa: o desenho amostral da PIA-Empresa com base no Cadastro Central de Empresas do IBGE comporta dois estratos com tratamento diferenciado:
- Empresas com trinta ou mais pessoas ocupadas e/ou receita de vendas e serviços industriais superior a determinado valor atualizado anualmente (R\$ 8,8 milhões, na pesquisa de 2009): compõem o estrato certo, entrando com probabilidade 1, o que significa que são todas pesquisadas anualmente. Em 2009, corresponderam a um universo em torno de 36,1 mil empresas, respondendo por mais de 90% do valor adicionado das empresas industriais no âmbito da pesquisa.
 - Empresas com 1 a 29 pessoas ocupadas: são selecionadas por critérios de amostragem probabilística, garantindo representatividade nos níveis Brasil, Unidades da Federação, e segmentos produtivos CNAE predefinidos. Em 2009, para representar o universo de 299 mil empresas industriais com 1 a 29 pessoas ocupadas, foram selecionadas mil empresas.
5. Variáveis: as principais variáveis pesquisadas referem-se a pessoal ocupado, receitas auferidas, impostos sobre a produção, valor das compras, estoques, composição dos custos incorridos, inclusive o de mão de obra, variações do ativo imobilizado; a PIA-Empresa prioriza o levantamento de informações econômico-financeiras voltadas a subsidiar o Sistema de Contas Nacionais nas estimativas do volume do valor adicionado (valor da produção menos consumo intermediário) e sua composição (remuneração de assalariados, encargos sociais) e da formação de capital pelas empresas industriais.

Os resultados da PIA-Empresa são disponibilizados no nível Brasil e das Unidades da Federação. Para os Estados com maior peso da indústria (São Paulo, Rio de Janeiro, Minas Gerais, Paraná, Santa Catarina e Rio Grande do Sul), são divulgados resultados por segmento industrial definido a três dígitos da CNAE; e, para os demais, a dois dígitos. Para as empresas com 30 ou mais pessoas ocupadas, cujo levantamento é censitário, é possível a obtenção de informações no detalhe do questionário, no nível mais detalhado da classificação de atividades (4 dígitos da CNAE) e para qualquer recorte geográfico (unidade da federação, municípios).

A série atual da PIA-Empresa inicia-se em 1996. Os resultados são divulgados pelo IBGE com dezoito meses de defasagem em relação ao ano de referência (no final do primeiro semestre do ano $n+2$), com perspectivas de redução deste prazo. A data de divulgação é anunciada no portal do IBGE no início do ano $n+2$. Os resultados são disseminados por meio da publicação Pesquisa Industrial Anual – Empresa, de periodicidade anual, que inclui um CD-ROM, e pela Internet no portal do IBGE <http://www.ibge.gov.br> e no Sidra – Sistema IBGE de Recuperação Automática <http://sidra.ibge.gov.br/bda/indust>.

Para os anos anteriores a 1996, as fontes mais abrangentes sobre a estrutura da indústria brasileira – indústrias extrativas e de transformação são os censos industriais, referidos aos anos 1920, 1940, 1950, 1960, 1970, 1975, 1980 e 1985 e a Pesquisa Industrial Anual - PIA que teve início em 1966, com o objetivo de acompanhar a evolução da estrutura industrial nos períodos intercensitários. As PIAs referentes ao período 1966-1995, no entanto, passaram por diversas fases com diferenças nas abordagens metodológicas e nos desenhos amostrais, com prejuízo para a coerência das séries temporais. Por outro lado, as diferenças metodológicas em questões fundamentais da produção de estatísticas industriais, tais como a definição da unidade estatística de informação e da classificação de atividades econômicas, entre as PIAs Empresa de 1996 em diante e as bases de dados anteriores (censos e PIAs até 1995), como já mencionado, introduzem fortes complicadores no encadeamento das séries das estatísticas industriais pré e pós 1996.¹³

Completando a caracterização da estrutura industrial, a PIA-Produto, cuja série tem início em 1998, levanta informações de valor e quantidade dos produtos e serviços industriais produzidos e/ou vendidos no ano, nas unidades locais industriais de empresas industriais. A PIA-Produto é desenhada como subamostra intencional da PIA-Empresa. Após sucessivos ajustes,¹⁴ a partir de 2005 a amostra da pesquisa passa a ser formada por todas as unidades produtivas industriais pertencentes ao estrato certo da PIA-Empresa, ou seja, as empresas industriais com 30 ou mais pessoas ocupadas. As informações são solicitadas a partir da PRODLIST-Indústria, com 3.500 denominações compatibilizadas com a Nomenclatura Comum da Mercosul – NCM, o que possibilita a articulação entre informações da produção nacional e dos fluxos de importação e exportação de produtos industriais, além da associação às classes de atividades industriais da CNAE. Os resultados da PIA-Produto são divulgados nos mesmos prazos e mídias que a PIA-Empresa, sendo objeto de uma publicação específica. Anteriormente à PIA-Produto, os censos industriais de 1960, 1970, 1975 e 1980 e as PIAs da década 1970 levantaram informações no nível de produtos que foram objeto de publicações do IBGE. A nomenclatura usada, desenvolvida pelo IBGE, não era compatibilizada com a do comércio externo.

Na área de estatísticas conjunturais da indústria, o IBGE produz indicadores de acompanhamento do desempenho no curto prazo da indústria brasileira baseados em dois levantamentos: a Pesquisa Industrial Mensal – Produção Física (PIM-PF) e a Pesquisa Industrial Mensal de Emprego e Salários (Pimes).

Desde a década de 1970, o IBGE divulga índices mensais relativos à produção industrial mensal, calculados a partir da PIM-PF. A variável levantada pela pesquisa é a produção física de uma série de produtos. Os produtos e os seus respectivos informantes são previamente selecionados pelo critério geral de participação no mercado, segundo os segmentos industriais que se pretende representar pelo sistema de indicadores. O sistema de índices da produção industrial passou por reformulação para ajustar-se aos parâmetros da produção das estatísticas econômicas vigentes a partir de 1996, com a nova série iniciando-se em 2002. A base atual para a seleção dos produtos e informantes da PIM-PF e para a definição do sistema de ponderação dos índices de quantum construídos com as informações da pesquisa passa a ser a PIA-empresa e a PIA-Produto e as classificações de referência, a CNAE e a PRODLIST-Indústria.¹⁵

A PIM-PF, no desenho atual, coleta informações mensais sobre a quantidade produzida de cerca de 830 produtos em cerca de 3700 unidades locais de empresas das indústrias extrativas e de transformação. O painel de produtos e informantes consiste de uma amostra de seleção intencional para representar no mínimo 50% do conjunto das atividades selecionadas em cada detalhamento geográfico. No conjunto, a PIM-PF cobre 63% do Valor de Transformação Industrial da PIA-Empresa do período 1998/2000.

São divulgados índices de *quantum* no nível nacional, desagregado em 27 atividades, e no nível regional, detalhado para as atividades mais representativas da indústria da Grande Região ou Unidade da Federação.

As atividades são definidas segundo categorias da CNAE (versão original). No caso das indústrias extrativas, os resultados são apresentados no nível agregado de *seção*. No caso das indústrias manufatureiras, além do nível agregado de *seção*, os resultados são apresentados em 27 atividades, das quais 19 correspondem no nível de *divisão* (dois dígitos da CNAE) e as oito restantes, a detalhamento de *grupos* (três dígitos da CNAE) ou agregações de *grupos*, dentro das divisões de Fabricação de Produtos Alimentícios, Produtos Químicos e Produtos Diversos. A atividade de reciclagem (Divisão 37 da CNAE) não está incluída. No nível nacional, são ainda divulgados índices por categorias de uso (bens de capital, intermediários, de consumo durável e de consumo não durável), para o conjunto de produtos identificados como consumo da construção civil e segundo o detalhamento de 26 subsetores industriais, compatíveis com a classificação nível 100 da série anterior.

No nível regional, os índices são divulgados para 13 Unidades da Federação – Amazonas, Pará, Ceará, Pernambuco, Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul e Goiás - segundo critério de representação de no mínimo 1% do Valor de Transformação Industrial – VTI, e para o conjunto da Região Nordeste.

Os índices são calculados pelo método de Laspeyres-base fixa em cadeia, com atualização de pesos. A base atual de ponderação (revisão 2002) é fixa e tem como referência a estrutura média do Valor de Transformação Industrial do período 1998/2000 e a composição de produtos da PIA-Produto de 1998. São divulgados quatro tipos de índices: base fixa mensal (mês/média de 2002=100), a partir do qual se pode construir qualquer outro, mensal (mês/mês do ano anterior), acumulado no ano (de janeiro até o mês/igual período do ano anterior) e acumulado dos últimos 12 meses (12 meses até o mês/12 meses anteriores). São apresentadas séries dessazonalizadas. Os índices de Produção Física, Brasil são divulgados em cerca de 34 dias corridos depois do mês de referência, pela internet, no portal do IBGE <http://www.ibge.gov.br>.

A série reformulada dos índices de produção física com início em janeiro de 2002 está encadeada à série anterior com início em janeiro de 1991. As séries históricas antigas (de janeiro de 1985 até janeiro de 2004) e a série atual (de janeiro de 1991 em diante) podem ser consultadas no Sistema IBGE de Recuperação de Dados – Sidra, no endereço <http://sidra.ibge.gov.br/bda/indust/>.

A Pesquisa Industrial Mensal de Emprego e Salários (Pimes) é, a partir de dezembro de 2000, a base de informações para a elaboração dos índices de emprego e salário voltados ao acompanhamento no curto prazo do mercado de trabalho da indústria brasileira. A pesquisa levanta informações mensais do nível do emprego, do movimento de admissões e desligamentos, da intensidade de uso do trabalho (número de horas pagas) e dos valores recebidos pelos assalariados (folha de pagamento). A Pimes segue os parâmetros do sistema de produção das estatísticas econômicas, implementado a partir de meados da década de 1990. A unidade estatística de investigação é a unidade local produtiva de empresas industriais, o âmbito é o universo das empresas com cinco ou mais pessoas ocupadas nas atividades das indústrias extrativas e de transformação, a classificação é a CNAE e a seleção da amostra é feita com base no Cadastro Central de Empresas do IBGE.

São divulgados índices de emprego e salário com base na Pimes para 18 grupamentos de atividades econômicas, correspondendo a divisões ou agregações de divisões das seções da CNAE referentes às indústrias extrativas e de transformação. Os índices são divulgados no nível Brasil, de 10 Unidades da Federação (Ceará, Pernambuco, Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul), e de agregados de Unidades da Federação (regiões Norte e Centro Oeste; e Nordeste, exceto Ceará, Pernambuco e Bahia. Esses índices, no nível Brasil, são divulgados, com defasagem de cerca de 40 dias em relação ao mês de referência, pela internet no endereço <http://www.ibge.gov.br>. As séries históricas dos índices estão disponíveis no portal do IBGE, no Sistema IBGE de Recuperação Automática – SIDRA <http://www.sidra.ibge.gov.br/bda/indust/>.¹⁶

Além das pesquisas anuais de natureza estrutural - PIA-Empresa e PIA-Produto - e do sistema de indicadores conjunturais da indústria, o IBGE realiza, desde 2000, a Pesquisa de Inovação Tecnológica - Pintec.

A Pintec visa caracterizar e mensurar as atividades inovativas das empresas brasileiras e acompanhar sua evolução no tempo, dentro de padrão metodológico e conceitual acordado internacionalmente¹⁷ para garantir a construção de indicadores setoriais, nacionais e regionais compatíveis com as demandas nacionais de informação e comparáveis internacionalmente. A atividade inovativa abrange a inovação de produto – novo produto que a empresa coloca no mercado – e de processo – introdução de tecnologias de produção ou de métodos de oferta de serviços, manuseio e entrega de produtos, novos ou significativamente aperfeiçoados. Entre outras, as informações mais relevantes referem-se a: número de empresas inovadoras, gastos com as atividades inovativas, fontes de financiamento destes gastos, número, nível de qualificação e tempo de dedicação das pessoas ocupadas em P&D interno às empresas. Permite, ainda, mensurações do tipo impacto das inovações sobre o faturamento e as exportações, efeitos sobre o desempenho das empresas, papel dos mecanismos de apoio do governo. Pesquisa também questões como a de fontes de ideias e arranjos cooperativos estabelecidos, métodos de proteção à inovação, obstáculos encontrados à atividade inovativa e inovações não tecnológicas implementadas.

A série da pesquisa inicia-se com a Pintec 2000, levantando informações relativas ao triênio 1998-2000, seguida pelas pesquisas de 2003, 2005 e 2008, sempre referidas ao triênio anterior. Nas duas primeiras pesquisas, 2000 e 2003, o âmbito foi o universo das empresas com 10 ou mais pessoas ocupadas nas indústrias extrativas e de transformação. A partir da PINTEC 2005, o âmbito da pesquisa foi alargado para incorporar as empresas com 10 ou mais pessoas atuantes em atividades de serviços de alta intensidade tecnológica de telecomunicações, informática e pesquisa e desenvolvimento.¹⁸

A unidade estatística de investigação e observação é a empresa, o que significa que as informações se referem à empresa como um todo. A classificação de referência é a CNAE, com a transição para a versão 2.0 na Pintec 2008.

Como as demais pesquisas a empresas do IBGE, a Pintec é uma pesquisa por amostra com base no Cadastro Central de Empresas – Cempre. No desenho da Pintec são usadas técnicas de amostragem apropriadas para situações de eventos raros, como é o caso das atividades inovativas. A amostra é desenhada para produzir resultados no nível Brasil, Grandes Regiões e Unidades da Federação selecionadas em função da representatividade na produção. No nível Brasil, produzem-se resultados para as indústrias extrativas como um todo (nível seção da CNAE) e para as indústrias de transformação, no nível de dois dígitos da CNAE (divisões), no geral, e de três dígitos (grupos) selecionados.

O banco de dados da Pintec inclui, além dos resultados da pesquisa, cruzamentos com dados de outras fontes, em particular dados da PIA-Empresa (variáveis econômico-financeiras e o volume das pessoas ocupadas da empresa), dados sobre quantidade e valor de produtos importados e exportados pelas empresas da pesquisa (fonte Secretaria de Comércio Exterior – SECEX), identificação da origem do capital nacional ou estrangeiro (fonte banco central).

A metodologia completa da pesquisa está documentada em publicação específica (IBGE. 2004 d) e nas notas técnicas que acompanham a divulgação dos resultados, o que é feito por meio de publicações específicas, acompanhadas de CD-ROM, e pela internet no endereço www.ibge.gov.br/home/estatistica/economia/industria/pintec.

Ainda no tocante a informações relativas à indústria brasileira, a partir de 2010, o IBGE disponibiliza mensalmente e no nível nacional o Índice de Preços ao Produtor –IPP das atividades das indústrias de transformação. Em etapas posteriores, o IPP será estendido às demais atividades produtivas do País. O IPP mede as variações médias nos preços dos produtos vendidos pelos produtores domésticos no mercado interno e externo (preços na porta da fábrica). O IPP é estruturado na CNAE e toma como referência a PROLIST-Indústria, o que facilita sua comparação com congêneres no nível internacional. Por outro lado, os produtos especificados para o acompanhamento de preços são um subconjunto dos produtos da PIM-PF, favorecendo a articulação interna do sistema das estatísticas industriais do IBGE. São produzidos índices para os grupamentos de atividades definidas na CNAE a dois dígitos (nível divisão) e, em alguns casos, a três dígitos (nível grupo) e para grupamentos seguindo a ótica de categorias de uso (bens de consumo, bens de capital e bens intermediários). Entre outros aportes analíticos, o IPP possibilitará um melhor mapeamento da transmissão de aumento de preços em cadeia. As informações sobre o IPP estão disponíveis na Internet, no endereço www.ibge.gov.br/home/estatistica/indicadores/precos/ipp. O IBGE é também responsável pela elaboração e divulgação das Contas Nacionais do Brasil, sistema de informações indispensável à análise da indústria, como quadro de referência completo e integrado de toda a economia. Nas séries atuais das Contas Nacionais, com base no ano 2000, entre outros avanços metodológicos e operacionais destacam-se a adoção da CNAE como classificação de referência e o uso das pesquisas econômicas anuais como fonte principal dos dados para mensuração dos agregados relativos às atividades produtivas. São produzidas Contas Nacionais anuais e trimestrais e Contas Regionais, por Unidade da Federação.

Em relação ao acervo de dados das Contas Nacionais, são particularmente úteis à análise da indústria brasileira, as tabelas de Usos e Recursos de Bens e Serviços que combinam as óticas de atividades e produtos, fornecendo dados de valor da produção, consumo intermediário, estrutura do valor adicionado, pessoal ocupado, impostos e destino da produção por setor de atividade

e categoria da demanda final, além de exportações e importações. No maior detalhamento, estes dados são apresentados para 55 atividades compatíveis com categorias da CNAE, das quais 36 são atividades das indústrias extrativas e de transformação, e 100 grupamentos de produtos, dos quais 70 são de origem em atividades industriais.

A partir das Tabelas de Usos e Recursos, são construídas matrizes de Insumo-Produto, envolvendo trabalho de desagregação dos dados para identificação da origem nacional e importada dos produtos consumidos e de revaloração dos fluxos de consumo a preços básicos. Com base em hipóteses sobre comportamentos do mercado (*market share* e outras), são construídas tabelas de coeficientes técnicos diretos e de coeficientes de impactos, segundo o modelo Leontieff, que possibilitam calcular o impacto sobre a produção no nível das atividades de um acréscimo exógeno na demanda final.

Os resultados das Contas Nacionais, anuais e trimestrais, das Contas Regionais e das Matrizes de Insumo Produto 2000 e 2005 são objeto de publicações específicas e podem ser acessados no portal do IBGE na Internet www.ibge.gov.br/home/estatistica/economia/contasnacionais.

28.3.2 CONFEDERAÇÃO NACIONAL DA INDÚSTRIA (CNI) E FEDERAÇÕES ESTADUAIS

A CNI também elabora e divulga um sistema de indicadores conjunturais da indústria. A unidade de investigação é o conjunto de estabelecimentos pertencentes à mesma atividade de cada empresa selecionada nos diversos estados. O levantamento é realizado pelas Federações Estaduais da Indústria (FIESP, FIRJAN, FIEMG etc.) e são pesquisadas cerca de 3.700 empresas de médio e grande porte em 12 estados.

As variáveis levantadas são: valor das vendas, pessoal empregado, salários líquidos, horas trabalhadas na produção e nível de utilização da capacidade. Os indicadores nacionais consolidam dados obtidos pelas federações de indústria estaduais, ponderando indicadores de cada estado pela sua participação na indústria nacional segundo o censo industrial de 1985.

Além disso, produzem estudos e avaliações sobre a conjuntura econômica e realizam pesquisas especiais não sistemáticas, com foco na atividade industrial (sobre modernização, investimento e outras). A FIRJAN, por exemplo, disponibiliza periódicos e estudos que abordam temas tais como: infraestrutura, expectativas, competitividade, meio-ambiente, inovação, comércio exterior e educação. A FIESP, por sua vez, além de divulgar diversos indicadores da atividade industrial, realiza o *Sensor FIESP*, caracterizado por ser uma pesquisa qualitativa de conjuntura econômica, trazendo resultados sobre mercado, vendas, estoque, emprego e investimentos industriais.

28.3.3 FUNDAÇÃO GETÚLIO VARGAS (FGV)

Os principais índices de preços industriais são produzidos pela FGV. Na revista *Conjuntura Econômica* são divulgados mensalmente índices de preços por atacado (IPA), segundo o conceito de disponibilidade interna (IPA-DI) e de oferta global (IPA-OG). Este último apresenta informações para cerca de quarenta atividades industriais. Além desses índices de preços industriais, são disponibilizados também índices de preços agropecuários, índices de confiança, taxas de câmbio, índices de ações, indicadores financeiros, indicadores sociais, mercado de trabalho, indicadores industriais, PIB, setor público e comércio exterior.

A FGV também realiza a *sondagem conjuntural* do setor industrial, investigando a cada trimestre a utilização de capacidade e variáveis qualitativas que sinalizam a expectativa das empresas para o trimestre seguinte. Há também a *sondagem de investimentos da indústria*, que fornece sinalizações sobre o rumo dos investimentos produtivos no setor e informações sobre objetivos dos investimentos, sua evolução e seus fatores limitadores. Por fim, pode-se destacar também a *sondagem da indústria* (somente disponível para venda), a qual apresenta previsões para o trimestre que se inicia no mês da pesquisa (janeiro, abril, julho e outubro); observações sobre o trimestre imediatamente anterior ao da pesquisa; e avaliações relativas à situação atual dos negócios da empresa. Há ainda uma avaliação de como o empresariado nacional está prevendo a situação dos negócios durante os seis meses seguintes ao de realização da pesquisa.

28.3.4 SECRETARIA DE COMÉRCIO EXTERIOR (SECEX)

A SECEX, por meio da base Aliceweb, apresenta dados sobre o comércio externo – valores, toneladas, países de origem e destino, portos de entrada e saída de importações e exportações – cuja fonte básica e censitária é a própria SECEX, a partir do registro obrigatório dos importadores e exportadores para realizarem as operações de comércio. São disponíveis dados mensais e anuais para quase 9.000 produtos.

28.3.5 FUNDAÇÃO CENTRO DE ESTUDOS DO COMÉRCIO EXTERIOR (FUNCEX)

A FUNCEX elabora índices de *quantum* e valor do comércio externo para cerca de trinta atividades industriais com base nos dados da SECEX. Já a base de dados Funcex Data é dividida em dois módulos, sendo apenas um de acesso gratuito. Porém, para seu uso é necessário cadastramento. A base de acesso gratuito contém os seguintes dados: valores de exportação e importação; índices de preço e *quantum* das exportações e importações; índices de rentabilidade das exportações; índice de demanda externa efetiva; termos de troca; coeficientes de exportação; coeficientes de penetração das importações; estatísticas de empresas exportadoras e importadoras por porte da firma; estatísticas de exportação e importação segundo intensidade tecnológica; estatísticas de exportação e importação por grupos de produtos segundo a origem setorial e a intensidade no uso de fatores; estatísticas de exportação e importação segundo dinamismo dos produtos; estatísticas de exportação segundo categorizações setoriais específicas; taxa de câmbio.

28.3.6 RELAÇÃO ANUAL DE INFORMAÇÕES SOCIAIS (RAIS)

A RAIS é a fonte mais abrangente e detalhada para estudos do pessoal ocupado na indústria. É uma fonte gratuita, mas há a necessidade de cadastro prévio. Toda atividade econômica formal é obrigada a prestar informações ao Ministério do Trabalho, que divulga os resultados da RAIS para a indústria e demais setores, utilizando a CNAE (Classificação Nacional das Atividades Econômicas). Dessa forma, a RAIS é a fonte básica de dados não só quanto ao emprego formal como quanto a informações referentes a pequenas e médias empresas no país.

As informações sobre o pessoal ocupado abrangem, entre outras: tipos de ocupação, escolaridade, faixa etária, gênero, área geográfica, tipo de vínculo empregatício, tempo de serviço, remunerações, horas contratuais dos empregados. Embora não inclua dados sobre valor da produção ou receita, a RAIS fornece informações sobre o tamanho (segundo o número de empregados) dos estabelecimentos.

28.3.7 CADASTRO GERAL DE EMPREGADOS E DESEMPREGADOS (CAGED)

Dados sobre a admissão e demissão de trabalhadores com carteira assinada mensalmente e o acumulado do ano estão disponíveis nesta base de dados. São disponibilizados pelo MTE (como a RAIS é gratuito, mas há a necessidade de cadastro prévio). Os dados são desagregados por setor e traz os mesmos tipos de informações disponíveis na RAIS para empregos: qualificação, remuneração, faixa etária, horas trabalhadas e tempo no emprego.

28.3.8 GAZETA MERCANTIL

A fonte mais abrangente para informações financeiras das maiores empresas do país é o *Balanço Anual* elaborado pela Gazeta Mercantil. No *Balanço Anual* publicado em 2000 (referente a dados do ano anterior) constam informações referentes à cerca de 8.500 empresas e trezentos grupos econômicos.

Os dados são obtidos dos balanços das empresas e incluem: receita operacional líquida, lucro, patrimônio líquido e dívidas financeiras. As empresas são apresentadas em ordem decrescente de receita e classificadas em mais de duzentas atividades. É identificada a unidade da federação da sede da empresa e, para as maiores empresas e grupos econômicos, a propriedade do capital (privado nacional, estrangeiro e estatal).

O *Balanço Anual* constitui a principal fonte para a análise de empresas específicas e para estudos de liderança empresarial.

Havia também um jornal denominado *Gazeta Mercantil*, que publicava os balanços das empresas e trazia notícias sobre a indústria, comércio e serviços. O jornal deixou de circular em 1º de junho de 2009 e o site saiu do ar.

28.3.9 ASSOCIAÇÕES DE INDÚSTRIA

Diversas atividades industriais contam com Associações que divulgam sistematicamente informações sobre o desempenho do setor. Metodologia, escopo e abrangência de informantes são diferenciados, mas predominam dados sobre produção física. Dentre as principais associações tem-se: ABIQUIM (cobrindo quase toda a indústria química), ANFAVEA (automobilística), IBS (siderurgia), ANFPC (papel e celulose), SNIC (cimento), ABIT/Sinditêxtil (têxtil), ABIMAQ (máquinas e equipamentos), ABIOVE (óleos vegetais) e diversas outras.

28.3.10 INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA)

A partir do IPEAdata pode se obter informações: macroeconômicas (dados econômicos e financeiros em séries anuais, mensais e diárias), regionais (dados econômicos, demográficos e geográficos) e sociais (indicadores de distribuição de renda, pobreza, educação, saúde, previdência social e segurança pública). Trata-se de um banco de dados federal onde estão reunidas diversas outras bases de dados federais. Contem dados federais, estaduais e setoriais em séries históricas. A partir do site do IPEA estão disponíveis estudos, pesquisas e textos para discussão. Entre as demais estatísticas, podem ser encontradas o Índice de Qualidade do Desenvolvimento.

28.4 Outras Fontes

Cabe ainda citar algumas instituições que realizam pesquisas não sistemáticas ou divulgam estudos sobre a indústria brasileira que se constituem em importantes fontes adicionais.

O BNDES (Banco Nacional de Desenvolvimento Econômico e Social) é praticamente a única fonte de financiamento de longo prazo no país. Registros de suas operações são relevantes para a análise do investimento, variável-chave na análise da indústria. Esse órgão também produz e divulga continuamente estudos setoriais cobrindo diversas atividades.

Duas categorias de empresas vêm assumindo importância crescente na indústria brasileira: empresas transnacionais, no valor da produção e investimentos, e, numericamente, pequenas e microempresas. A SOBET (Sociedade Brasileira de Estudos de Empresas Transnacionais) fornece publicação mensal voltada para temas de conjuntura relativos ao setor externo, tais como fluxos de investimentos estrangeiro e de balança comercial. Incluem a avaliação das contas externas brasileiras, os movimentos gerais da conjuntura econômica internacional que repercutem na solvência externa do Brasil, além de eventuais análises sobre o desempenho externo de alguns países que traga potenciais desdobramentos sobre a economia brasileira. O SEBRAE (Serviço de Apoio às Micro e Pequenas Empresas) constitui fonte relevante sobre as últimas, realizando também pesquisas sobre a indústria em geral.

O Estado de São Paulo é responsável por quase 50% do valor da produção industrial do Brasil (IBGE, 2000), o que torna relevante a consulta ao órgão oficial de estatísticas do estado, o SEADE, que realiza pesquisas e análises sobre a indústria local. Merece menção a Pesquisa da Atividade Econômica Paulista (PAEP), produzida desde 1996.

O Banco Central produz todas as estatísticas financeiras do país e também realiza pesquisas específicas, como o censo do capital estrangeiro, com informações importantes para a análise industrial. São disponibilizadas informações sobre juros, spread bancário, balanço de pagamentos, balança comercial, taxa de câmbio, investimento direto externo, dívida pública, nível de atividade, reservas internacionais, indicadores monetários, de crédito, do mercado financeiro e de capitais. O censo de capital estrangeiro é realizado quinquenalmente, desde 1996. A série teve início com a publicação dos resultados do ano-base 1995, seguida pela divulgação dos resultados referentes aos anos-base 2000, 2005 e 2010.

Pode-se citar também a Agência Brasileira de Desenvolvimento Industrial (ABDI), que realiza estudos setoriais e conjunturais, além da *Sondagem de Inovação*, pesquisa que acompanha trimestralmente a evolução da inovação tecnológica na indústria brasileira.

Cabe ainda mencionar a Associação Nacional de Pesquisa, Desenvolvimento e Engenharia das Empresas Inovadoras (ANPEI), instituição que realiza pesquisas sistemáticas sobre outra importante variável em estudos industriais, a tecnologia.

Finalmente, constituem importante fonte de informações sobre a indústria estudos e pesquisas realizados por Universidades, como os institutos de Economia da Universidade Federal do Rio de Janeiro (IE/UFRJ) e da Universidade Estadual de Campinas (IE/UNICAMP). Essas instituições divulgam os resultados de suas pesquisas na forma de textos para discussão, monografias, e teses de mestrado e doutorado, que podem ser consultados nessas instituições. O IE/UFRJ disponibiliza ainda dados sobre os setores de petróleo, energia elétrica e sucroalcooleiro a partir dos sites dos respectivos grupos de pesquisa.

É importante frisar que foram mencionadas apenas as principais fontes de informação atualmente disponíveis sobre a indústria brasileira, lembrando ainda que esse tipo de levantamento é sempre datado, refletindo a situação vigente nesta data. Para cada atividade ou tema específico é necessário pesquisar fontes específicas, sendo sempre conveniente a realização de entrevistas com empresas líderes do segmento e com especialistas.

Notas

1. Os organizadores agradecem a Magdalena Cronemberger, Rodrigo Lopes e Henrique Cavalieri, a primeira coautora do capítulo e os dois últimos, respectivamente, alunos de mestrado e doutorado do IE/UFRJ, pela atualização do capítulo na ausência de Lia Haguenauer, falecida em 24 de maio de 2002.

2. Até a revisão do Programa das Estatísticas Econômicas iniciado em meados dos anos 1990, as principais estatísticas industriais brasileiras foram construídas tendo como unidade estatística de referência o estabelecimento definido como “empresa ou parte de uma empresa que exerce, de forma independente, predominantemente *um único tipo de atividade econômica em uma única localização*, para a qual existem ou podem ser compiladas informações que permitam o cálculo do excedente operacional”. Tal definição de estabelecimento, ao contrário do que ocorre em relação à empresa e à unidade local, não é inequívoca, exigindo a especificação de critérios operacionais para sua implementação prática. Em particular, nos censos industriais realizados pelo IBGE até 1985, ocorreu ao longo dos anos um desmembramento crescente de unidades locais em estabelecimentos, principalmente devido a procedimentos práticos que levaram à fragmentação cada vez maior das etapas de processos produtivos em plantas integradas. Como resultado, as informações para os diversos anos apresentam inconsistências entre si, o que levou a instituição a reformular suas estatísticas passando a dar prioridade à empresa e às unidades locais (IBGE, 1991). A mudança na unidade estatística de referência, necessária ao realinhamento das estatísticas industriais brasileiras, implicou, por sua vez, descontinuidades entre as séries pré e pós-revisão dos anos 1990.
3. Atualmente, as definições de empresa e unidades locais nas estatísticas industriais brasileiras coincidem com as definições do Cadastro Nacional de Pessoa Jurídica – CNPJ, da Secretaria da Receita Nacional – SRN. Assim, a cada empresa corresponde um específico CNPJ, cujos sufixos identificam os vários estabelecimentos/unidades locais que a compõem. A identidade entre as unidades estatísticas e as definições administrativas facilita a articulação do sistema estatístico com os registros administrativos, o que é crucial para que o sistema estatístico possa responder de forma eficiente e em menor tempo à demanda por informações econômicas mais abrangentes e diversificadas.
4. De acordo com a PIA – Empresa de 2007 (IBGE, 2009a), 99% das empresas industriais atuavam em apenas uma divisão de atividade. As empresas que operavam em mais de uma divisão de atividade (1%), entretanto, foram responsáveis por 31% do total da receita líquida de vendas das empresas industriais no mesmo ano.
5. Alguns países, como o Canadá, têm avançado na identificação dos grupos econômicos e, portanto, na possibilidade de leitura das estatísticas neste enfoque, ainda que na organização das pesquisas a unidade de referência continue sendo a empresa (preferencialmente) ou o estabelecimento.
6. Ver ONU 2008.
7. O quinto nível da CNAE – as subclasses – é um detalhamento para uso da Administração Pública.
8. Ver IBGE, 2007 também acessível no portal do IBGE www.ibge.gov.br.
9. Ver IBGE, 2008a também acessível no portal do IBGE www.ibge.gov.br.
10. Para a descrição completa da metodologia da PIA-Empresa ver IBGE, 2004c e, para ajustes ocorridos posteriormente, ver as notas técnicas que constam da publicação anual dos resultados da pesquisa (IBGE, 2009a), também acessível no portal do IBGE www.ibge.gov.br.
11. A prática nas pesquisas anteriores, censos e PIAs foi de considerar o estabelecimento como *locus* central de informações, onde se aplicava o questionário mais detalhado, prevalecendo alógica inversa: a construção dos agregados econômicos a partir de informações levantadas no nível do estabelecimento, sem a garantia de coerência com os dados econômico-financeiros da empresa como um todo.
12. As principais alterações introduzidas na versão 2.0 em atividades no âmbito das pesquisas industriais foram: a) na definição de âmbito, a passagem das atividades de edição (inclusive a impressão quando integrada na mesma empresa) para a nova seção de Informação e Comunicação e, das atividades de reciclagem, para a seção de Água, esgoto e atividades de gestão de resíduos e descontaminação; b) na definição de novas divisões, por desmembramentos ou rearranjo das categorias anteriores: a divisão Fabricação de produtos farmaquímicos e farmacêuticos (antes parte da divisão Fabricação de produtos químicos); a divisão Fabricação de equipamentos de informática, produtos eletrônicos e ópticos (rearranjo das divisões anteriores que comprendiam a produção de equipamentos e de material eletrônico e de comunicação, tendo como motivação a definição de uma melhor ferramenta para as estatísticas da produção de alta tecnologia); a divisão Fabricação de bebidas (antes parte da divisão Produtos alimentares); a divisão Fabricação de móveis (antes parte da divisão Fabricação de produtos diversos) e a divisão Manutenção, reparação e instalação de máquinas e equipamentos (antes as unidades especializadas nestas atividades estavam tratadas em grupos específicos dentro da divisão que compreendia os fabricantes dos respectivos equipamentos). Ver também as principais alterações na versão 2.0 IBGE, 2007, acessível no portal do IBGE www.ibge.gov.br.
13. Para os resultados das PIAs referentes aos anos 1966 a 1969, 1973 a 1974, 1976 a 1979, 1981 a 1984, 1986 a 1990, 1992 a 1995 acessar www.sidra.ibge.gov.br/bda/indust.
14. Para o histórico destes ajustes informado nas Notas Técnicas que acompanham os resultados da PIA-Produto de 2007 ver IBGE, 2009b, também disponível no portal do IBGE <http://www.ibge.gov.br>.
15. Para informações completas da metodologia da PIM-PF e do sistema de índices da produção industrial ver IBGE, Série relatórios Metodológicos, v. 31, Indicadores Conjunturais da Indústria - Produção, 2004.
16. Desde a década de 1970, o IBGE divulga dados de acompanhamento conjuntural de emprego e salário na indústria. Até 1984, as informações provinham da pesquisa voltada ao acompanhamento do volume físico da produção industrial. A partir de 1985, é implementada a Pesquisa Industrial Mensal – Dados Gerais – PIM-DG, de desenho amostral com base no Censo Industrial de 1980, voltada especificamente à construção de índices de emprego e salário na indústria brasileira. A série de indicadores com base na PIM-DG se inicia em janeiro de 1985 e se estende até abril de 2001 (IBGE, 1996).
17. A referência conceitual e metodológica da Pintec é o Manual de Oslo; e o modelo de referência, a pesquisa proposta pela Oficina Estatística da Comunidade Europeia – Eurostat, a terceira versão da *Community Innovation Survey – CIS III 1998-2000*.
18. Com o alargamento do âmbito da pesquisa alterou-se sua denominação, anteriormente Pesquisa Industrial de Inovação Tecnológica.

Bibliografia

- IBGE (2007) Classificação Nacional de Atividades Econômicas- CNAE: versão 2.0, Rio de Janeiro, IBGE.
- IBGE (1991) Ensaios sobre Estatísticas do Setor Produtivo, Texto para Discussão n. 42, Rio de Janeiro: IBGE.
- IBGE (1996) Indicadores Conjunturais da Indústria: Produção, Emprego e Salário, 2.ed. Série Relatórios Metodológicos, v. 11, Rio de Janeiro, IBGE.
- IBGE (2004A) Indicadores Conjunturais da Indústria – Emprego e Salários, Série Relatórios Metodológicos, v. 27, Rio de Janeiro, IBGE.
- IBGE (2004b) Indicadores Conjunturais da Indústria - Produção, Série Relatórios Metodológicos, v. 31, Rio de Janeiro, IBGE.

- IBGE (2011) Índices de Preços ao Produtor – Industrias de Transformação, Série Relatórios Metodológicos, v. 38, Rio de Janeiro, IBGE.
- IBGE (2011a) Lista de Produtos da Indústria PRODLIST – Indústria 2010, Rio de Janeiro, IBGE.
- IBGE (2008b) Matriz de Insumo-Produto Brasil 2000/2005, série Contas Nacionais n. 23, Rio de Janeiro, IBGE.
- IBGE (2011b) Pesquisa Industrial v. 28, n. 2, 2009 – Produto, Rio de Janeiro, IBGE.
- IBGE (2004c) Pesquisa Industrial Anual – Empresa, Série Relatórios Metodológicos, v. 26, Rio de Janeiro, IBGE.
- IBGE (2004d) Pesquisa de InovaçãoTecnológica, Série Relatórios Metodológicos, v. 30, Rio de Janeiro, IBGE.
- ONU (2008) International Standard Industrial Classification of all Economic Activities -ISIC, rev. 4, Statistical Paper Series M, n. 4, Nova York: ONU, 2008.

Fontes de informação acessíveis na Internet

<http://www.ibge.gov.br>

<http://www.sidra.ibge.gov.br/bda/indust>

http://www.cni.org.br

http://www.fgv.br