

# Métodos quantitativos em defesa da concorrência e regulação econômica

**TOMO 1**

Eduardo P. S. Fiúza • Ronaldo Seroa da Motta  
coordenadores técnicos

Imperfeições de mercado podem gerar concentrações de poder econômico ou condutas anticompetitivas, que reduzem a eficiência da economia e restringem o bem-estar social. Essa evidência é considerada nas legislações de quase todos os países. No Brasil, as atividades de defesa da concorrência têm passado por um aprendizado contínuo desde o advento da Lei 8.884/94. O processo envolve mudanças de mentalidade e de hábitos dos agentes econômicos, reformas das instituições oficiais habituadas a exercer controles, atitude política e, principalmente, capacitação nos órgãos que compõem o Sistema Brasileiro de Defesa da Concorrência (SBDC).

Esses órgãos, a saber, a Secretaria de Direito Econômico do Ministério da Justiça (SDE/MJ), a Secretaria de Acompanhamento Econômico do Ministério da Fazenda (Seae/MF) e o Conselho Administrativo de Defesa Econômica (Cade), já desenvolveram um corpo técnico habilitado a analisar atos de concentração e processos de conduta anticompetitiva. Consolidou-se também no país uma comunidade de advogados e economistas especializados na defesa da concorrência.

Toda essa comunidade se beneficia da publicação destes dois volumes. A obra resulta de ampla pesquisa teórica e empírica envolvendo especialistas de diversas instituições, selecionadas pela Associação Nacional dos Centros de Pós-Graduação em Economia (Anpec) sob a coordenação técnica do Instituto de Pesquisa Econômica Aplicada (Ipea). Os estudos apresentados identificam e analisam a potencialidade e a limitação de diversos métodos quantitativos na análise de casos antitruste em atos de concentração e conduta. A sofisticação das análises na área da defesa da concorrência já vem ocorrendo no Brasil, combinando modelos econômétricos e teoria dos jogos com aplicação de algoritmos de simulação de mercado. Esse nível de conhecimento ainda precisa ser absorvido de forma plena e simétrica pelos diversos atores do SBDC. Entretanto, como salientamos capítulos deste livro, nem sempre os modelos sofisticados são os mais recomendáveis e, considerando as características microeconômicas dos casos, o analista vezes, com técnicas e de execução rápida, desde que aplicadas no contexto adequado.

Os dois tomos trazem um total de nove capítulos, distribuídos em quatro partes. O Tomo 1 apresenta manuais e estudos de demanda e o Tomo 2 se dedica às análises complementares de atos de concentração e às análises de condutas.

A primeira parte do Tomo 1 divide-se em dois capítulos, que analisam os modelos econômétricos estruturais de oferta (Capítulo 1) e margens de erro nas previsões e mensuração de eficiências em fusões (Capítulo 2). Ainda no Tomo 1, a Parte 2 apresenta estudos de demanda por serviços bancários (Capítulo 3) e de saúde (Capítulo 4), que são objeto de regulação prudencial.

O Tomo 2 é todo dedicado a análises de atos de concentração e conduta. A Parte 3, com os Capítulos 5, 6 e 7, cobre as concentrações verticais, o poder de compra em concentrações horizontais, o conluio tácito, os efeitos coordenados, os preços predatórios e as restrições verticais. Esses temas, são retomados na quarta e última parte do livro, porém agora com foco voltado especificamente para os setores bancário (Capítulo 8) e de saúde suplementar (Capítulo 9).

O canal entre a literatura científica e a prática da análise antitruste está aberto. Muito ainda há por ser feito, mas acreditamos que esta publicação seja um grande passo no processo de aperfeiçoamento da regulação econômica e da defesa da concorrência no Brasil.

# Métodos quantitativos em defesa da concorrência e regulação econômica

## **Governo Federal**

**Presidente** – Luiz Inácio Lula da Silva

## **Ministério da Justiça**

**Ministro** – Márcio Thomaz Bastos

**Secretário-Executivo** – Luiz Paulo Teles Ferreira Barreto

**Secretário de Direito Econômico** – Daniel Krepel Goldberg

**Diretora do Departamento de Proteção e Defesa Econômica** – Mariana Tavares de Araújo

## **Ministério do Planejamento, Orçamento e Gestão**

**Ministro** – Paulo Bernardo Silva

**Secretário-Executivo** – João Bernardo de Azevedo Bringel

## **Instituto de Pesquisa Econômica e Aplicada**

**Presidente** – Glauco Arrix

**Chefe de Gabinete** – Persio Marco Antonio Davison

**Diretor de Estudos Macroeconômicos** – Paulo Mansur Levy

# Métodos quantitativos em defesa da concorrência e regulação econômica

Eduardo P. S. Fiuza • Ronaldo Seroa da Motta  
coordenadores técnicos

---

Métodos quantitativos em defesa da concorrência e regulação econômica/Eduardo P. S. Fiúza e Ronaldo Seroa da Motta, coordenadores técnicos – Rio de Janeiro: Ipea, 2006.

2v.

932 p.

1. Concorrência 2. Modelos Econométricos 3. Regulação Econômica  
I. Fiúza, Eduardo Pedral Sampaio II. Motta, Ronaldo Seroa da III. Instituto de Pesquisa Econômica Aplicada.

---

ISBN - 85-86170-79-8

CDD 330.015195

*As opiniões e recomendações emitidas nos estudos desta publicação não representam a posição oficial da Secretaria de Direito Econômico do Ministério da Justiça (SDE/MJ) nem exprimem, necessariamente, os pontos de vista do Instituto de Pesquisa Econômica Aplicada (Ipea), do Ministério do Planejamento, Orçamento e Gestão (MPOG) ou da Associação Nacional dos Centros de Pós-Graduação em Economia (Anpec). As idéias aqui apresentadas são de exclusiva e inteira responsabilidade dos autores e não têm qualquer caráter normativo, mas sim analítico e propositivo.*

*É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.*

## **APRESENTAÇÃO**

### **SDE: em busca de novas ferramentas para a análise antitruste**

As atividades de defesa da concorrência no Brasil têm passado por um processo de aprendizado contínuo desde o advento da Lei 8.884/94. A transição de uma economia fechada e cheia de controles para uma economia de mercado aberta e competitiva requer mudanças de mentalidade e de hábitos dos agentes econômicos, reformas das instituições oficiais habituadas a exercer os controles, atitude política e, principalmente, muita capacitação para os órgãos que aplicam aquela lei, compondo o Sistema Brasileiro de Defesa da Concorrência (SBDC).<sup>1</sup>

Malgrado a ausência de uma carreira específica de analista de concorrência, a baixa remuneração e a alta rotatividade do corpo técnico dos órgãos do SBDC, foi sendo criado ao longo do tempo um corpo de funcionários, recrutados em diversas carreiras, habilitado a efetuar as análises de atos de concentração e processos de conduta anticompetitiva. Em torno desses quadros também foi se desenvolvendo uma comunidade de advogados e economistas especializados no ramo. A cultura da defesa da concorrência foi criando raízes e o arcabouço de análise econômica foi pouco a pouco sendo construído, passando pela criação dos guias de análise e de certos parâmetros que, editados pelos órgãos antitruste, auxiliam a fixar regras para o diálogo entre administradores e administrados.

A responsabilidade do SBDC também foi crescendo, acompanhando a complexidade dos casos analisados. A experiência antitruste norte-americana e européia tinha de ser aproveitada, e com ela vieram as ferramentas de análise cada vez mais sofisticadas. A comunidade de economistas foi buscar essas ferramentas para responder a perguntas que inquietavam os conselheiros do Cade diante de operações de grande vulto, como a fusão entre as cervejarias Brahma e Antarctica, dando origem à Ambev, ou a aquisição da Chocolates Garoto pelo grupo Nestlé.

Era imperiosa a necessidade de o SBDC se capacitar para absorver essas ferramentas e avaliar com autoridade técnica as proposições, alegações e estimativas

---

1. O SBDC é integrado pela Secretaria de Direito Econômico do Ministério da Justiça (SDE/MJ), pela Secretaria de Acompanhamento Econômico do Ministério da Fazenda (Seae/MF) e pelo Conselho Administrativo de Defesa Econômica (Cade).

das partes envolvidas nas operações de atos de concentração e processos de conduta. Essa capacitação, que ainda está em curso e cujo sucesso depende da consolidação de um corpo técnico dedicado e ainda mais treinado, certamente deu um passo significativo com o Projeto do Centro de Métodos Quantitativos da SDE. A contratação da Associação Nacional de Centros de Pós-Graduação em Economia (Anpec), sob a supervisão técnica do Instituto de Pesquisa Econômica Aplicada (Ipea), para a realização de Estudos em Métodos Quantitativos Aplicados à Defesa da Concorrência e à Regulação, tentou mobilizar a academia brasileira para a pesquisa empírica rigorosa nos temas da Organização Industrial e Regulação Econômica e, assim, munir o SBDC das ferramentas de que ele tanto necessita no seu dia-a-dia.

É bom lembrar, no entanto, que a sofisticação cobra seu preço, geralmente em termos de agilidade dos órgãos instrutórios. É por isso que a aplicabilidade e uma certa dose de pragmatismo foram alguns dos critérios para a resenha das técnicas aqui apresentadas. Técnicos da SDE, da Seae e do Cade foram treinados pelos consultores da Anpec para que possam executar pessoalmente ou avaliar criticamente os exercícios econometrivos propostos pelas partes dos processos. Além desses exercícios econometrivos e do complemento ao *Guia de Análise de Atos de Concentração Horizontal* para levar em conta o poder de compra das firmas fusionadas, também foram propostos guias de análise para a instrução de processos de conduta e para a análise de atos de concentração vertical.

A aplicação de métodos quantitativos na defesa da concorrência, ao contrário de desonrar os juristas, requer-lhes que repensem uma vasta gama de regras e princípios de direito: que disciplina outorgar a esses estudos econometrivos do ponto de vista do direito da prova? Em que momento do processo administrativo, e sob quais regras, tais estudos devem ser introduzidos, quer pelos órgãos instrutórios, quer pelas partes ou terceiros interessados no processo?

Neste compêndio, os economistas encontrarão um primeiro passo na tentativa de sistematizar o ferramental que se tornará, inexoravelmente, parte do cotidiano da análise antitruste. De outro lado, os juristas hão de vislumbrar, aqui, uma nova ordem de desafios, instigantes e complexos.

Aos técnicos do Ipea e do Departamento de Proteção e Defesa Econômica (DPDE) da SDE, assim como aos consultores da Anpec envolvidos neste projeto, que tanto empenho e zelo demonstraram na busca incessante da qualidade deste produto final, nossos sinceros agradecimentos.

Daniel K. Goldberg

Secretário de Direito Econômico do Ministério da Justiça

## **Ipea: aproximar políticas da fronteira da teoria econômica**

A participação do Instituto de Pesquisa Econômica Aplicada (Ipea) na coordenação técnica dos Estudos em Métodos Quantitativos Aplicados à Defesa da Concorrência e à Regulação Econômica, atendendo à demanda da Secretaria de Direito Econômico do Ministério da Justiça (SDE/MJ), reflete o duplo compromisso da instituição de aproximar da formulação de políticas públicas os avanços da fronteira da teoria econômica e de apoiar a ação dos órgãos de governo na sua implementação. Em particular, em sintonia com as tendências mais recentes, uma das prioridades do Ipea tem sido a investigação das reformas microeconômicas que propiciam um ambiente de eficiência e concorrência para os negócios — essencial para que prevaleçam os incentivos aos investimentos necessários à consolidação de uma trajetória de crescimento sustentável, com todos os seus benefícios em termos de geração de emprego e renda.

A execução deste projeto, que esteve a cargo da Coordenação de Estudos de Mercado e Regulação do Ipea, contou também com a colaboração de outros especialistas de nossa Diretoria de Estudos Macroeconômicos (Dimac). E a parceria com a Associação Nacional dos Centros de Pós-Graduação em Economia (Anpec) foi fundamental para aglutinar os melhores centros de pesquisa no cumprimento desta complexa tarefa. A interação com os técnicos da SDE não só foi determinante para a qualidade acadêmica do trabalho, como também para a adequação dos estudos à realidade do exercício da defesa da concorrência no Brasil.

Assim como em outras iniciativas da SDE, temos certeza de que os resultados dessas pesquisas serão uma referência para o contínuo aprimoramento das análises econômicas no âmbito do Sistema Brasileiro de Defesa da Concorrência (SBDC).

Agradecemos a todos os que, com elevado zelo profissional, colaboraram na realização desta obra, desde os autores dos capítulos até aqueles que participaram na gestão do projeto na SDE, na Anpec e no Ipea, assim como na editoração destes dois tomos.

Glauco Arrix  
**Presidente do Ipea**

## **Anpec: incentivo aos estudos sobre defesa da concorrência**

*Métodos quantitativos em defesa da concorrência e regulação econômica* vem preencher uma importante lacuna na literatura econômica brasileira. Surge, sem dúvida, como o manual de referência do país na área de defesa da concorrência e análise antitruste. Combina o estado da arte da teoria econômica e dos métodos quantitativos à didática de uma prática, voltada para subsidiar os técnicos e tomadores de decisão direta e indiretamente ligados ao Sistema Brasileiro de Defesa da Concorrência (SBDC).

Coordenado por Eduardo P. S. Fiuza e Ronaldo Seroa da Motta, o livro contou com o apoio institucional da Associação Nacional de Centros de Pós-Graduação em Economia (Anpec). Cumprindo sua missão estatutária de estimular o desenvolvimento das Ciências Econômicas no país, a Anpec contratou especialistas de renome nacional e internacional para a elaboração dos estudos que compõem o presente volume. Dois editais de chamada foram realizados. O primeiro em junho de 2004 e o segundo em agosto do mesmo ano. Uma comissão de avaliação formada por economistas de notório saber analisou as 26 propostas e selecionou os estudos a serem contratados.

Pelo seu pioneirismo, estamos certos de que o livro traz uma grande contribuição para disseminar o conhecimento da economia da defesa da concorrência junto ao aparato institucional de regulação do país, ainda nos primórdios de seu desenvolvimento.

Mauro Borges Lemos  
**Secretário-Executivo da Anpec**

# **SUMÁRIO**

## **TOMO 1**

### **INTRODUÇÃO 1**

Eduardo P. S. Fiúza e Ronaldo Seroa da Motta

### **PARTE 1**

#### **MODELOS ECONOMÉTRICOS ESTRUTURAIS**

##### **CAPÍTULO 1**

##### **ESTIMAÇÃO E IDENTIFICAÇÃO DE DEMANDA E DE OFERTA 21**

Cristian Huse e Alberto Salvo

##### **CAPÍTULO 2**

##### **MARGENS DE ERRO E EFICIÊNCIAS EM FUSÕES 153**

Heleno Martins Pioner e Maurício Canêdo-Pinheiro

### **PARTE 2**

#### **ESTIMAÇÃO DE DEMANDA EM SETORES REGULADOS**

##### **CAPÍTULO 3**

##### **DEMANDA POR SERVIÇOS BANCÁRIOS 255**

Márcio I. Nakane, Leonardo S. Alencar e Fabio Kanczuk

##### **CAPÍTULO 4**

##### **DEMANDA NO SETOR DE SAÚDE SUPLEMENTAR NO BRASIL 295**

Mônica Viegas Andrade, Ana Carolina Maia, Daniela Goes Paraiso e Rubens José Amaral de Brito

###### **4A. Regulamentação do setor 297**

###### **4B. A sobreutilização dos serviços 317**

###### **4C. A seleção adversa no sistema 341**

###### **4D. Estudo de caso da demanda 365**

## **TOMO 2**

### **PARTE 3**

#### **MÉTODOS QUANTITATIVOS NA ANÁLISE DE ATOS DE CONCENTRAÇÃO E CONDUTA**

##### **CAPÍTULO 5**

###### **CONCENTRAÇÕES VERTICais E PODER DE COMPRA EM ATOS DE CONCENTRAÇÃO HORIZONTAL 461**

Maurício Canêdo-Pinheiro e Heleno Martins Pioner

**5A. Concentrações verticais 463**

**5B. Fusões e poder de mercado na compra de insumos 503**

**5C. Aplicações de métodos quantitativos 534**

##### **CAPÍTULO 6**

###### **CONLUIO TÁCITO, EFEITOS COORDENADOS E PREÇOS PREDATÓRIOS 579**

Emílio H. Matsumura e Marina Figueira de Mello

**6A. Conluio tácito 581**

**6B. Efeitos coordenados 653**

**6C. Preços predatórios 676**

##### **CAPÍTULO 7**

###### **RESTRIÇÕES VERTICais 715**

Marina Figueira de Mello e Emílio H. Matsumura

### **PARTE 4**

#### **REGULAÇÃO PRUDENCIAL E DEFESA DA CONCORRÊNCIA**

##### **CAPÍTULO 8**

###### **ATOS DE CONCENTRAÇÃO E CONDUTA NO SETOR BANCÁRIO 793**

João Manoel Pinho de Mello

##### **CAPÍTULO 9**

###### **ATOS DE CONCENTRAÇÃO, PODER DE MONOPSÔNIO E RESTRIÇÕES VERTICais NO MERCADO DE SAÚDE SUPLEMENTAR 855**

Maria Teresa Marins Duclos

**9A. Atos de concentração horizontal 857**

**9B. Poder de monopsônio e condutas verticais 888**

# INTRODUÇÃO

Eduardo P. S. Fiúza

Do Departamento de Proteção e Defesa Econômica da SDE

Ronaldo Seroa da Motta

Da Coordenação de Estudos de Mercados e Regulação do Ipea

Imperfeições de mercado podem resultar em concentrações de poder econômico ou condutas anticompetitivas que reduzem a eficiência da economia e restringem o bem-estar social. Essa percepção já é comumente aceita e está inserida nas legislações de quase todos os países. No Brasil a Constituição Federal de 1988 elegeu a concorrência como uma característica-chave da ordem econômica e, portanto, a defesa da concorrência se consolidou como uma das funções reguladoras do Estado. Função essa que se alterou drasticamente a partir da década de 1990 com a liberação das importações e, em seguida, com as privatizações, inclusive aquelas dos serviços públicos e de infra-estrutura.

Enquanto a presença do Estado na esfera produtiva se reduzia, o seu papel regulador tinha de crescer e se transformar, e uma nova forma de governança tinha de ser construída. Assim, foram criadas as agências reguladoras autônomas, que implementariam os marcos regulatórios dos ditos serviços públicos que fossem geradores de externalidades significativas, seja na forma de monopólio natural ou risco sistêmico. Complementarmente, aprimoravam-se os mecanismos de defesa da concorrência.<sup>1</sup>

Como parte das reformas de 1994 que acompanharam o Plano Real, uma nova lei de concorrência (Lei 8.884) foi promulgada, introduzindo controle de concentrações e importantes mudanças institucionais. Com a lei, o Conselho Administrativo de Defesa Econômica (Cade), criado em 1962, ganhou autonomia como parte de um projeto de Estado que abandonava definitivamente os controles de preços como instrumentos de política microeconômica de oferta e abraçava a defesa da concorrência como linha-mestra.<sup>2</sup>

Para dar suporte às atividades do Cade, foram criados dois outros órgãos: a Secretaria de Direito Econômico (SDE) do Ministério da Justiça e a Secretaria

1. Para uma análise detalhada da regulação econômica dos serviços públicos e de infra-estrutura, ver Salgado e Seroa da Motta (2005).

2. Como salientam o Banco Interamericano de Desenvolvimento (BID) e a Organização para Cooperação e Desenvolvimento Econômico (OCDE) (2005), a era moderna da política de concorrência no Brasil começou em 1994. O mesmo estudo indica uma evolução significativa em desempenho desde então. Ver também a Introdução de Fiúza (2001) para uma análise do período anterior.

de Acompanhamento Econômico (Seae) do Ministério da Fazenda, formando com o Cade o chamado Sistema Brasileiro de Defesa da Concorrência (SBDC).

A SDE tem, de acordo com a Lei 8.884, a competência de instruir atos de concentração econômica e processos que apuram infrações contra a ordem econômica. A instrução desses expedientes envolve uma análise de natureza jurídica e uma de natureza econômica. Com respeito à análise econômica, esta tem o objetivo de fornecer a compreensão do mercado envolvido em cada caso, em todos os seus aspectos. A análise antitruste exige uma definição de mercado relevante, o entendimento de como é o funcionamento desse mercado e o impacto que resulta do ato de concentração ou da conduta sob apreciação da SDE. O resultado dessa análise é imprescindível à emissão de um parecer jurídico satisfatório pela SDE e ao julgamento posterior pelo Cade.

## MOTIVAÇÃO

Durante algum tempo não se teve uma moldura analítica única no SBDC. Só em 1999 a Seae adotou um *Guia de Análise de Atos de Concentração Horizontal*, que, em 2001, evoluiu para um guia de análise conjunto SDE-Seae. Nota-se, contudo, uma acentuada evolução técnica e processual das instruções desde 1994. A adoção de guias de análise (a Seae também elaborou um guia para práticas de preços predatórios em 2002) certamente foi fundamental para a harmonização e a maior celeridade das análises nas duas secretarias.

Ainda assim, há uma parcela de casos — tanto atos de concentração quanto processos de apuração de condutas — que exige uma análise econômica mais profunda e completa, para que os órgãos de instrução possam emitir pareceres mais precisos e o órgão julgador possa tomar decisões mais eficazes e tempestivas. A realização da análise econômica com o grau de rigor aqui mencionado requer maior *expertise* e maior facilidade de levantamento de dados econômicos e de mercado. Nesses casos, também pode ser necessária a utilização de modelos conceituais sofisticados, que pressupõem o tratamento por profissionais qualificados para tal tarefa. O SBDC, como um todo, tem compreendido que cada tipo de caso pode exigir análises diferenciadas e com grau distinto de complexidade.

Essa sofisticação já vinha sendo percebida nos países de mais longa tradição na defesa da concorrência e, junto com o surgimento de novos modelos teóricos de análise de casos antitruste, observou-se um rápido desenvolvimento de métodos quantitativos aplicados à averiguação dos casos.<sup>3</sup> Essa sofisticação de análise

3. A esse respeito, ver, por exemplo, as resenhas de Baker e Rubinfeld (1999) sobre o uso de métodos quantitativos nos Estados Unidos e da LECG (1999) sobre as experiências britânica e europeia.

vem sendo, de certa forma, aplicada no Brasil — ver, por exemplo, os casos da Companhia Brasileira de Bebidas (Ambev) e da Nestlé/Garoto —, com técnicas de análise que combinam modelos econôméticos e teoria dos jogos com aplicação de algoritmos de simulação de mercado. Mas esse nível de conhecimento ainda não é absorvido de forma plena e simétrica pelos diversos atores do SBDC.

## OBJETIVOS

Métodos quantitativos são técnicas para a aplicação de modelos teóricos a dados empíricos. Assim, a extensão e a complexidade de técnicas quantitativas requerem também a compreensão dos modelos teóricos. Dessa forma, a SDE lançou o projeto Estudos em Métodos Quantitativos na Defesa da Concorrência e na Regulação Econômica para traduzir de forma didática a fronteira atual desse conhecimento, com ênfase nas aplicações à análise antitruste.

Embora o estado da arte dos métodos fosse o principal objetivo, isso não implica necessariamente que modelos sofisticados sejam sempre os mais recomendáveis. Considerando as características microeconômicas dos casos, o analista pode ser conclusivo, muitas vezes, com técnicas simples e de execução rápida, desde que aplicadas no contexto adequado. Por outro lado, em muitos casos as estimativas dos parâmetros econômicos, dada a disponibilidade de dados, devem ser interpretadas com cautela. Dessa forma, longe de restringir ou fixar procedimentos, os estudos apresentados nesta publicação traduzem a literatura econômica quantitativa antitruste na forma de proposições de metodologia e procedimentos estimativos.

O principal objetivo era o de desenvolver estudos de tal forma que os integrantes do SBDC pudessem ser orientados quanto:

- a) aos pressupostos teóricos dos métodos quantitativos;
- b) às limitações de predição decorrentes desses pressupostos;
- c) à aplicabilidade dos métodos ante a disponibilidade de dados; e
- d) à integração desses métodos em procedimentos concatenados junto com outras evidências que instruem os casos.

Com base nas proposições deste estudo, a SDE está promovendo no âmbito do SBDC um esforço para selecionar e definir aquelas que possam vir a compor guias oficiais, a exemplo do atualmente adotado para a análise de atos de concentração horizontal.

Para melhor difundir esses estudos no âmbito do SBDC, um minicurso foi oferecido aos seus integrantes. Todavia, vale ressaltar que, para pôr em prática

os futuros guias que eles podem vir a influenciar, o SBDC precisará dar continuidade aos esforços de treinamento e consolidar um quadro permanente de especialistas.

Naturalmente esta publicação não esgota os temas envolvendo métodos quantitativos aplicados à defesa da concorrência. Testes estatísticos de preços e tendências de preços (usando *cross-sections*, fazendo análise hedônica, testando correlação e cointegração, causalidade e velocidade de ajustamento etc.) na delimitação de mercado relevante já têm sido adotados em alguns casos tanto pelas partes como pela própria SDE e, assim, existe material que pode dar origem a um estudo sistematizador.

Devido às suas especificidades, estudos de demanda e guias de análise ajustados para supermercados e indústria farmacêutica são outra frente a ser levada adiante.<sup>4</sup> Por fim, a análise quantitativa de mercados caracterizados por processos de leilões ou licitações requer um arcabouço específico tendo em vista a recente ênfase da SDE na apuração de cartéis em licitações.

## HISTÓRICO

A Coordenação de Estudos de Mercado e Regulação do Instituto de Pesquisa Econômica Aplicada (Ipea) foi selecionada para compartilhar a coordenação técnica do projeto junto à Coordenação-Geral de Análise Econômica da SDE. A parceria com a Anpec foi decisiva para aglutinar os melhores centros de pesquisa para a execução do projeto.

Dois editais de chamada de propostas foram realizados pela Anpec. O primeiro em junho de 2004 e o segundo em agosto do mesmo ano. Uma comissão de avaliação coordenada pelo Ipea analisou as 26 propostas recebidas<sup>5</sup> e três tipos de estudos foram contratados:

*a)* manuais que apresentam uma descrição detalhada e didática de questões relacionadas às técnicas quantitativas na análise antitruste e que propõem procedimentos metodológicos e estimativos em temas gerais, tais como funções de demanda e oferta, mensuração de eficiências e modelos de simulação;

---

4. Vale notar, contudo, que, em outra frente de pesquisa, a Associação Nacional dos Centros de Pós-Graduação em Economia (Anpec) também mobilizou uma equipe de economistas e farmacêuticos para desenvolver uma metodologia de delimitação de mercado relevante para a indústria farmacêutica, a qual já está em fase final de assimilação pela SDE e pela Seae.

5. Os pareceristas que colaboraram na seleção das propostas foram Ajax Moreira, Alexandre Marinho, Armando Castelar Pinheiro, Carlos Henrique Corseuil Leite, Carlos Octávio Ocké-Reis, Elyson Caiado Rocha Lima, Gabriel Godofredo Fiúza de Bragança, José Gustavo Feres, Luiz Carlos Garcia de Magalhães, Paulo Tafner e Thompson Andrade.

*b)* análise regulatória que compreende estudos de demanda de certos setores específicos — no caso, saúde suplementar e bancos, em que predominam assimetrias de informação; e

*c)* análises de atos de concentração e de conduta, que discutem os modelos teóricos e seus métodos quantitativos, além de determinar possíveis seqüências de decisões na averiguação dos casos. Os temas foram as concentrações verticais, poder de compra em concentrações horizontais, conluio tácito, efeitos coordenados, preços predatórios e restrições verticais. Esses assuntos foram tratados também separadamente para áreas de saúde suplementar e de bancos, em que a regulação prudencial, por vezes, conflita com a defesa da concorrência.

Nos estudos de análises de atos de concentração e conduta, três consultores sistematizadores, Thompson Andrade, Cleveland Prates Teixeira e Ruy Santacruz, ex-conselheiros do Cade, colaboraram no processo de revisão dos guias de análise junto à Coordenação Técnica.

As etapas do projeto que culminou nesta publicação foram:

*a)* O lançamento dos editais e a seleção das propostas, realizados entre julho e setembro de 2004.

*b)* As propostas e os seus relatórios parciais foram disponibilizados no portal do Ipea e discutidos em um *workshop* interno na sede do Rio de Janeiro do Ipea entre 15 e 17 de dezembro de 2005, ao qual compareceram técnicos do SBDC, pesquisadores do Ipea e de outras instituições de pesquisa. Com base nas questões debatidas no evento, a coordenação técnica preparou e enviou aos autores uma avaliação com os comentários e sugestões para a versão do relatório final.

*c)* As primeiras versões dos relatórios finais foram analisadas pela coordenação técnica durante o mês de março e as versões revisadas, também disponibilizadas no portal do Ipea, foram apresentadas e discutidas no seminário de 25 a 28 de abril de 2005, em Brasília, aberto ao público em geral.

*d)* Com base nos resultados do evento de abril e de comentários recebidos de seus participantes, versões definitivas foram entregues até o final de agosto de 2005.

*e)* Após comentários e aprovação das versões finais dos relatórios no âmbito do SBDC, iniciou-se em dezembro de 2005 a editoração da publicação.

*f)* Entre janeiro e abril de 2006 realizaram-se os minicursos de treinamento dos relatórios e guias para os técnicos do SBDC.

## A ESTRUTURA DO LIVRO<sup>6</sup>

Esta publicação contém dois tomos com um total de nove capítulos, distribuídos em quatro seções. O Tomo 1 apresenta manuais e estudos de demanda e o Tomo 2 se dedica às análises complementares de atos de concentração e às análises de condutas.

A primeira parte do Tomo 1 divide-se em dois capítulos, que analisam os modelos econométricos estruturais de demanda e oferta (Capítulo 1) e margens de erro nas previsões e mensuração de eficiências em fusões (Capítulo 2).

Além de extremamente útil para a delimitação de mercados relevantes, a estimativa de modelos estruturais de demanda é fundamental para a simulação do novo equilíbrio de mercado nos casos em que a análise de atos de concentração efetuada pela autoridade antitruste chega a um ponto em que é necessário avaliar quantitativamente o real grau de rivalidade entre as firmas do mercado relevante e as consequências do ato de concentração para a relação de forças entre elas, tendo em vista a substituibilidade dos bens produzidos por elas. A estimativa de demanda também é fundamental, caso se queira estimar parâmetros de conduta e, assim, testar condutas concertadas em mercados específicos.

Ao longo de sua curta história, o SBDC foi percebendo como estimativas apresentadas por pareceristas — a partir de diferentes hipóteses a respeito de ordenação da escolha do consumidor, bens alternativos e situação de compra — poderiam trazer resultados bem diferentes. Reconheceu-se a importância da capacitação da autoridade antitruste para avaliar criticamente as vantagens e desvantagens de cada estimativa, bem como a confiabilidade das previsões resultantes.

O Capítulo 1, de Huse e Salvo, começa resenhando as principais especificações funcionais utilizadas em estudos de demanda aplicados à organização industrial. Com relação ao tratamento dado ao problema da dimensionalidade (o número de elasticidades cruzadas cresce quadraticamente com o número de bens do mercado), os modelos foram agrupados em modelos contínuos (em geral, restritos por hipóteses de orçamento multiestágio) e modelos de escolha discreta. O primeiro grupo lida com a maldição da dimensionalidade assumindo alguns pressupostos de agregação dos bens em categorias e de separabilidade das preferências, o que permite seqüenciar a decisão de consumo em conjuntos de escolha cada vez mais desagregados, até

6. Esta descrição está baseada nas apresentações dos capítulos, escritas por Eduardo P. S. Fiúza nos Capítulos 1 e 2; Ronaldo Seroa da Motta nos Capítulos 3 e 4; Ruy Santacruz e Thompson Andrade nos Capítulos 5, 6 e 7; João Manuel P. de Mello no Capítulo 8; e Maria Teresa Duclos no Capítulo 9.

chegar à substituição entre marcas, produtos ou ofertantes de interesse. O segundo grupo resolve a dimensionalidade modelando a demanda como função de um conjunto pequeno de características, de modo que as elasticidades passam a ser, elas também, funções desse mesmo conjunto.

Os autores discutem, em seguida, os desafios econométricos clássicos que surgem na estimativa da demanda, principalmente: *a)* a simultaneidade, que causa viés e inconsistência e deve ser corrigida por variáveis instrumentais adequadas; e *b)* a identificação dos parâmetros estruturais de oferta (custos e conduta). Para o segundo desafio, são examinadas três abordagens: *a)* estimativa simultânea (em dois estágios) de demanda, custos e conduta, estimando esta última como um parâmetro livre ou em função da estrutura de mercado (por exemplo, concentração); *b)* seleção entre modelos comportamentais; e *c)* assumir, em vez de estimar, o modelo de conduta.

Por fim, atentos aos critérios de tempestividade, parcimônia, operacionalidade e simplicidade de interpretação, tão caros ao SBDC, os autores cobrem brevemente os modelos PCAIDS e ALM, de simples e rápida operacionalização.

Como discutido no Capítulo 1, a identificação de custos e condutas no equilíbrio de mercado pode ser tratada pela estimativa simultânea de oferta e demanda. Em complemento à excelente resenha feita naquele texto era importante estudar o lado da oferta em casos de atos de concentração, pois, na ausência dessa estimativa, não se poderiam simular variantes de equilíbrio pós-fusão para diferentes cenários de variação dos custos ao longo da curva de custos original ou em resposta a deslocamentos da mesma propiciados pelas eficiências (específicas e não-pecuniárias) do ato.

Pioneer e Pinheiro, no Capítulo 2, dedicaram-se a examinar como as eficiências alegadas pelas requerentes deveriam ser incorporadas na simulação do novo equilíbrio ou, invertendo a pergunta, de quanto deveriam ser as eficiências para que os preços e as quantidades de equilíbrio pré-fusão fossem mantidos. Além disso, investigaram qual o grau de incerteza — traduzido por margem de erro estatístico — incorrido nesses exercícios. Tentou-se, ainda, avaliar a probabilidade de entrada propiciada pelo aumento da concentração, mas isso exige modelos computacionalmente complexos e que ainda demandam bastante pesquisa para se firmarem como alternativas viáveis ao analista antitruste.

Para a estimativa da oferta do mercado, os autores apresentam duas alternativas: a estimativa direta da função custo e a estimativa da função de produção. A escolha entre uma e outra depende do conjunto de dados disponível. Os

dados requeridos para a estimação e a simulação em cada caso são explicitamente listados pelos autores, o que ajuda o analista a escolher o método e/ou modelo mais adequado. Um procedimento simples de *bootstrap* é sugerido como metodologia de cálculo das margens de erro dos novos equilíbrios. Rotinas de MatLab para a simulação são incluídas no Apêndice. Os autores também discutem a quais eficiências e parâmetros o analista antitruste deve atentar, para cada medida de bem-estar cuja perda a autoridade antitruste queira minimizar.

Ainda no Tomo 1, a Parte 2 apresenta estudos de demanda por serviços bancários (Capítulo 3) e de saúde (Capítulo 4), que são objeto de regulação prudencial e para os quais, no Tomo 2, são propostos guias de análise específicos.

Com a estabilidade monetária, ficou mais fácil observar os elevados *spreads* bancários no país. Mais ainda, com um regime monetário de aperto fiscal com elevadas taxas de juros reais, sucessivos e elevados indicadores de lucratividade dos bancos operando no Brasil acabam reforçando a impressão de que os bancos atuam em mercados imperfeitos.

A literatura econômica, por outro lado, mostrou até agora evidências de que o setor é caracterizado por uma estrutura de mercado imperfeita, rejeitando a hipótese de que os bancos brasileiros operam sob a forma de cartel. Nakane, Alencar e Kanczuk, no Capítulo 3, aumentam o grau de conhecimento a respeito dessa polêmica com uma abordagem econométrica que incorpora recentes avanços na modelagem de mercados imperfeitos. Para tal, desenvolvem modelos de demanda *logit* multinomial sob cenários alternativos de concorrência entre bancos — um de concorrência de Bertrand e outro de cartel — para três produtos bancários: *a*) depósitos a prazo; *b*) soma de depósitos à vista e poupança; e *c*) empréstimos.

Quatro modelos de demanda foram estimados, combinando-se diferentes conjuntos de instrumentos e a inclusão (ou não) de interações entre as características do banco e a renda por habitante do município. Os resultados mostram que as elasticidades-preço próprias para a demanda de depósitos a prazo são bastante elevadas, enquanto as elasticidades da demanda por depósitos à vista e de poupança com relação a variações na taxa de juros de depósitos a prazo são bem mais modestas. As demandas por ambos os tipos de depósitos foram encontradas inelásticas com relação a variações nos preços dos serviços bancários.

No lado da oferta de serviços bancários, o teste de poder de mercado apresentado no estudo confronta as margens preço-custo empiricamente observadas com aquelas preditas pelos modelos de Bertrand e de cartel. Os resultados apresentados mostram que o modelo de Bertrand descreve bem o comportamento

dos preços de serviços bancários, mas superestima o poder de mercado observado na fixação de preços para depósitos a prazo e para empréstimos. Assim, sob bases empíricas e teóricas mais bem fundamentadas que em estudos anteriores, os resultados do estudo de Nakane, Alencar e Kanczuk tendem a confirmar que o setor bancário não atua na forma de cartel.

Cerca de 38 milhões de indivíduos (ou 21% da população residente) são beneficiários de algum tipo de plano ou seguro de saúde no Brasil. O Capítulo 4 analisa diferentes aspectos da demanda de saúde suplementar no Brasil. Inicialmente, no Subcapítulo 4A, Andrade discute a regulamentação dos planos e seguros de saúde no Brasil quanto às regras de determinação de preços e de benefícios definidas na Lei 9.656 de 1998.

Em seguida, três estudos econométricos tratam com originalidade os diversos aspectos da demanda do setor, que lida com duas falhas de mercado derivadas da assimetria de informação entre seguradora e segurado: risco moral e seleção adversa. Quanto maior a cobertura do plano ou seguro de saúde, maior é o incentivo para os indivíduos escolherem mais utilização do cuidado — um problema de risco moral. A seleção adversa, *ex ante* ao estabelecimento do contrato, consiste na seleção, por parte da seguradora, de clientes com risco maior que o risco médio da sociedade. Em ambos os casos a provisão dos serviços é ineficiente.

Andrade e Maia, no Subcapítulo 4B, medem uma *proxy* do risco moral através de medidas de sobreutilização dos serviços de saúde suplementar em relação aos serviços públicos de saúde, aplicando modelos de contagem (binomial negativo) com dados da PNAD 1988. Embora as autoras indiquem a existência de sobreutilização, esta pode refletir tanto a presença de problemas nos incentivos dos contratos de saúde suplementar quanto um problema de acesso ao setor público de saúde, ou mesmo as duas situações simultaneamente.

Ainda com base na PNAD 1998, Andrade e Maia estimam, no Subcapítulo 4C, a presença de seleção adversa dentro da população que possui plano de saúde privado e entre esta e a população coberta apenas pelo Sistema Único de Saúde (SUS). A seleção adversa é parametrizada pela correlação positiva entre variáveis de risco e pela probabilidade de terem plano de saúde, num modelo de contagem estimado por Máxima Verossimilhança com Informação Completa (FIML). Os principais resultados sugerem que atributos de risco observáveis, como sexo e idade, são importantes na decisão de adquirir plano de saúde. As características de risco não-observáveis não foram significativas.

No Subcapítulo 4D, Andrade, Paraiso e Brito, estimam diretamente a demanda por serviços de saúde considerando a população beneficiária do plano oferecido pela Sabesperv. O comportamento dos indivíduos é modelado através de quatro equações para identificar separadamente gastos ambulatoriais e de internação. Conclui-se que os atributos de risco, idade, sexo e estado de saúde passada são bastante relevantes na determinação dos gastos com serviços de saúde.

O Tomo 2 é todo dedicado a análises de atos de concentração e conduta. A Parte 3, com os Capítulos 5, 6 e 7, cobre as concentrações verticais, o poder de compra em concentrações horizontais, o conluio tácito, os efeitos coordenados, os preços predatórios e as restrições verticais. Esses temas são retomados na quarta e última parte do livro, porém agora com foco voltado especificamente para os setores bancário (Capítulo 8) e de saúde suplementar (Capítulo 9).

Atos de concentração horizontal podem reforçar o poder de mercado tornando as empresas capazes de elevar seus preços de forma não-transitória. Esse é, de modo geral, o maior temor das autoridades antitruste, explicitado na literatura e nos guias de análise existentes em vários países.

Porém, esses atos também podem reduzir a eficiência econômica ao criarem ou reforçarem o poder de monopsônio, ou seja, o poder de compra de insumos por parte das empresas envolvidas. Nesse sentido, Pinheiro e Pioner constroem, no Subcapítulo 5A, um arcabouço teórico fundamental para a avaliação do efeito de fusões e aquisições sobre o poder de compra das firmas. A partir das contribuições mais recentes da teoria econômica para a discussão, os autores apresentam uma resenha dos aspectos envolvidos no exercício do poder de monopsônio e discutem o objetivo da intervenção antitruste quanto a esses aspectos. Finalmente, apresentam, com base na teoria econômica, um conjunto de procedimentos a serem utilizados, apontando para a necessidade da aplicação de métodos empíricos de simulação e avaliação dos seus efeitos anticompetitivos.

Alguns atos de concentração resultam também na associação de empresas que estão em diferentes posições da mesma cadeia produtiva. Essa concentração integra fases referentes aos elos a montante e a jusante do processo produtivo dessa cadeia produtiva. Do ponto de vista competitivo, essa situação poderia propiciar efeitos danosos ao mercado, pois empresas verticalmente integradas podem fechar o mercado — dificultando ou impedindo o acesso de concorrentes a insumos essenciais e concentrando o mercado do produto — ou facilitar condutas colusivas. Reações contrárias a essa conclusão, como as dos adeptos da Escola de Chicago, chegaram a dominar o pensamento econômico durante

algum tempo, argumentando que algumas integrações verticais podem ser pró-competitivas e que elas não provocariam necessariamente o fechamento dos mercados de insumos ou do produto.

No Subcapítulo 5B, Pinheiro e Pioner examinam a literatura mais recente sobre o tema, a qual contém estudos que revigoraram a idéia de que atos de concentração vertical podem gerar preocupações quanto aos seus efeitos competitivos. São revistas análises com base em diversos modelos teóricos, que utilizam diferentes hipóteses para as características competitivas do mercado do insumo, a concentração do mercado a montante, a substituibilidade dos produtos concorrentes e as condutas das empresas participantes dos mesmos. No que tange ao efeito de facilitação de colusão decorrente de uma concentração vertical, os autores mostram que são importantes na análise concorrencial apenas o caso especial em que o mercado a montante seja muito concentrado ou o caso em que haja um histórico de tal comportamento por parte das empresas relevantes.

O Subcapítulo 5C oferece um rol de procedimentos a serem utilizados pelos analistas de atos de concentração que envolvam integração vertical. Fica claro que, embora haja alguma semelhança com procedimentos de análise empregados para o exame de atos de concentração horizontal, no caso de integração vertical a análise necessita ser mais abrangente, pois esta implica estudar mercados que estão necessariamente em elos diferentes da cadeia produtiva. Assim, a preocupação com o nível de concentração não ficará restrita a um mercado, mas sim a mais de um, para verificar se o fechamento do mercado pode ter alguma racionalidade no que toca aos incentivos e à viabilidade de uma conduta de mercado como essa, e quais são os custos e as eficiências gerados pela integração vertical.

O Capítulo 6, de Matsumura e Mello, concentra-se em três tipos de conduta: colusão, coordenação e predação. As características e as formas da cooperação entre firmas em um oligopólio são discutidas no âmbito dos denominados modelos “colusivos”, nos quais as firmas procuram estabelecer algum procedimento de coordenação conjunta das decisões, através de acordo explícito ou conluio tácito, que supõe a não-comunicação entre as firmas e não é considerado uma infração à ordem econômica.

A dificuldade de distinguir entre essas duas formas de coordenação é um problema com que se deparam os órgãos antitruste de todo o mundo. Nesse sentido, o Subcapítulo 6A representa uma enorme contribuição à defesa da concorrência ao sistematizar uma metodologia para a identificação de conluio tácito nos mercados. Caso a análise descritiva do caso indique a presença no

mercado investigado das condições necessárias ao conluio tácito, os autores propõem uma segunda etapa na qual seria realizada uma análise quantitativa com a finalidade de rejeitar ou não essa hipótese, comparando-se o comportamento das firmas que participam do mercado relevante com modelos econômicos de competição.

As abordagens empíricas propostas relacionam os parâmetros estimados das equações de comportamento das firmas com os valores dos parâmetros identificados na teoria econômica. A vantagem principal dessa abordagem é que as equações de equilíbrio a serem estimadas são derivadas da teoria econômica. Com isso, a natureza da inferência, embora potencialmente complexa, é clara, permitindo-se ainda que um conjunto de hipóteses alternativas seja formulado e testado.

Atos de concentração podem aumentar a probabilidade de exercício coordenado ou unilateral de poder de mercado. O Subcapítulo 6B apresenta o que a literatura econômica disponibiliza para a caracterização e estimativa dos efeitos da coordenação entre firmas sobre as variáveis de mercado. Os autores descrevem os procedimentos de coleta e transformação de dados em estatísticas de interesse na caracterização da coordenação de mercado. Em seguida, analisam os métodos quantitativos empregados para estimar os efeitos de atos de concentração sobre a capacidade de coordenação das empresas, mais especificamente: *a)* estudos de eventos; e *b)* estudos relacionados a fusões e seus impactos sobre a assimetria na indústria.

No Subcapítulo 6C, os autores analisam a prática de preços predatórios, que tem como finalidade dominar o mercado pela eliminação de concorrentes. Essa é uma das condutas anticompetitivas sobre a qual há maior controvérsia na literatura quanto a possibilidade e efeitos da predação sobre a eficiência econômica. As autoridades antitruste também encontram enorme dificuldade para superar problemas tanto de ordem conceitual quanto de ordem prática nas investigações desse tipo de conduta comercial. Por essa razão a análise de Matsumura e Mello traz luz à discussão, ao apresentar uma resenha da literatura econômica e da prática antitruste sobre o tema, além de uma proposta de investigação que incorpora os mais modernos elementos teóricos.

De um modo geral, as investigações de preços predatórios são realizadas em dois estágios. Em uma primeira etapa investiga-se se estão presentes no mercado as condições estruturais para a conduta (que inclui a existência de poder de mercado por parte da firma investigada), além da possibilidade de recuperação das perdas incorridas na etapa de afastamento ou redução substancial da

concorrência. Em um segundo estágio é analisado o diferencial entre preços e custos e a existência de eventual racionalidade na fixação de preços abaixo de custos.

Os autores mantêm essa mesma estrutura da análise, mas adicionam ao primeiro estágio um enquadramento teórico da estratégia predatória, fundamental para uma análise efetivamente econômica da conduta. Com isso, reduz-se a possibilidade de que práticas competitivas que conduzam a preços baixos sejam confundidas com práticas predatórias que visam à dominação de mercado.

A identificação do esquema de preços predatórios proposta é um instrumento adicional na investigação da conduta, tendo por objetivo demonstrar que as estratégias do caso em análise são economicamente racionais. Com essa análise investiga-se se a estratégia de preços predatórios é teoricamente plausível, sendo para isso necessário identificar e classificar a conduta investigada dentre os possíveis tipos de esquemas predatórios e suas respectivas evidências de sustentação. Dentre esses esquemas, dois modelos são identificados: *a)* os modelos predatórios de restrição financeira, em que a estratégia predatória é viabilizada pela existência de informação assimétrica no mercado de capitais; e *b)* os modelos de reputação, nos quais o predador tem mais conhecimento das condições de mercado do que os rivais efetivos e potenciais e, com isso, reduz seus preços com o objetivo de fazer com que os rivais acreditem que as condições de mercado são desfavoráveis.

O Capítulo 7, analisa as restrições verticais, outro tipo de conduta que permeia de controvérsia a literatura econômica. Na busca por competitividade, as empresas cada vez mais procuram intervir direta ou indiretamente nos mercados relacionados verticalmente — a montante ou a jusante — ao longo de suas cadeias produtivas. As intervenções diretas ocorrem por meio de fusões e aquisições, enquanto as indiretas se dão através de restrições impostas por produtores/ofertantes de bens ou serviços em determinado mercado (“de origem”) sobre outras empresas em mercados relacionados verticalmente (“alvo”). Essas restrições verticais, dependendo das condições, podem elevar ou reduzir a eficiência econômica.

A discussão não é trivial, constituindo-se provavelmente em uma das mais complexas da literatura econômica e da prática antitruste. Mello e Matsumura buscam, dessa forma, identificar as condições gerais sob as quais as principais restrições verticais observadas nos mercados reduzem a eficiência econômica, sendo, por essa razão, infrações à ordem econômica. Os autores resenham a literatura teórica de restrições verticais e as jurisprudências norte-americana e

da Comunidade Européia. São destacados como elementos de análise na proposta de um guia brasileiro: a observabilidade e duração dos contratos; a unilateralidade das restrições; o poder de mercado ou controle de recurso essencial; a capacidade de criar compromisso crível; e as características dos mercados adjacentes. Outra observação importante dos autores é que, muitas vezes, a origem de um problema de concorrência pode estar em outra etapa da cadeia produtiva, a montante ou a jusante. Assim, ante uma representação, o analista antitruste deve examinar toda a cadeia produtiva e não apenas o segmento para o qual a queixa aponta.

Apesar das dificuldades teóricas, os autores identificam na literatura testes empíricos capazes de auxiliar a autoridade antitruste na identificação dos efeitos negativos sobre a eficiência decorrentes de restrições verticais — ou seja, na identificação de restrições verticais anticompetitivas. A primeira forma de análise quantitativa de uma restrição vertical continua sendo uma apresentação de estatísticas descritivas, comparando-se o comportamento de variáveis em mercados com e sem restrições verticais. Porém, como as regularidades encontradas podem ter outras explicações, é necessário realizar também testes econométricos. Em termos metodológicos, uma resenha de estudos empíricos revela que vêm sendo usadas basicamente sete técnicas: descritiva; dados em corte (*cross-section*); séries de tempo; dados em painel; experimentos naturais; estudos de evento; e análises estruturais. A resenha mostra que há limitações nos estudos empíricos analisados, em termos de técnicas utilizadas, cobertura da indústria e possibilidades de interpretação dos resultados obtidos, mas que, ainda assim, eles podem auxiliar bastante a atividade antitruste, trazendo a teoria econômica para as decisões administrativas.

As dificuldades para sanar a assimetria de informação e a possibilidade de risco sistêmico impõem uma certa peculiaridade na provisão dos serviços financeiros e de saúde. Dessa forma, os Capítulos 8 e 9 fecham o Tomo 2 e a publicação com estudos que identificam possíveis diferenças analíticas de casos de concentração e conduta nesses setores.

Usando o estado da arte do conhecimento teórico e empírico a respeito da especificidade dos problemas bancários, Mello propõe, no Capítulo 8, adaptações específicas para o setor dos procedimentos já usualmente adotados na análise de fusões e conduta.

Em um mundo com informação e contratibilidade completas, não haveria razão para a existência de intermediação financeira. Em tal mundo, poupadores e tomadores jamais se encontrariam. Suas transações seriam mediadas pela troca anônima de papéis, ou seja, contratos de compra e venda de capital. Informação

completa garantiria que tudo o que é relevante a respeito do tomador estivesse no contrato, de modo que a taxa de juros contratual especificaria perfeitamente os riscos envolvidos na transferência de recursos dos poupadouros para os investidores. Com completude contratual, seria possível especificar todas as possíveis contingências no contrato e dizer quanto se pagaria em cada uma delas. Em outras palavras, nesse mundo não haveria nem seleção adversa nem risco moral.

No mundo real, entretanto, a intermediação financeira existe e é difícil subestimar sua importância. No Brasil, por exemplo, ela já representou quase 20% do Produto Interno Bruto (PIB). Bancos existem precisamente para aliviar os problemas de incompletude contratual e de informação incompleta. Ou seja, a *raison d'être* da indústria é resolver falhas de mercado. Isso torna a indústria bancária especial em geral, e no que se refere à defesa da concorrência em particular. O Capítulo 8 trata dessa especificidade. A autoridade de defesa da concorrência deveria tratar de forma especial fusões no setor bancário? Se sim, deveria ser mais ou menos tolerante? A análise de conduta deveria ser diferente? Um mesmo *mark-up*, com uma mesma estrutura de mercado, no setor bancário e no setor de bebidas não-alcoólicas implicaria a mesma probabilidade de conluio?

Um exemplo é ilustrativo. O setor bancário está sujeito à retirada em massa de seu principal insumo: depósitos à vista. Dá-se a esse fenômeno o nome de corrida bancária. A origem é uma falha de coordenação por parte dos depositantes, além do monitoramento imperfeito do intermediário (banco). Nesse contexto, a regulação prudencial em defesa da firma falida como justificativa de fusões é especial no setor bancário. Enquanto a falência de uma fábrica de cerveja normalmente implica realocação de fatores de produção, sem maiores consequências, a falência de um banco pode ter implicações sistêmicas, tanto para a indústria bancária como para a economia em geral. Além desse exemplo, há inúmeras outras razões para a adaptação da defesa da concorrência ao setor bancário. Invariavelmente elas têm origem em problemas de seleção adversa (informação incompleta) e risco moral (incompletude contratual), conforme se discute no início daquele capítulo.

A publicação se encerra no Capítulo 9 com um estudo realizado por Duclos sobre um dos temas mais controversos na análise antitruste, que diz respeito aos serviços de saúde. O setor de serviços de atenção à saúde é vital e, nas últimas décadas, devido aos avanços tecnológicos, tem melhorado substancialmente a forma pela qual esses serviços são prestados, assim como as perspectivas de recuperação dos pacientes. No entanto, aumentos substanciais nos custos desses serviços pressionam o sistema de seguro saúde em vários países.

A concorrência pode trazer benefícios importantes ao mercado de saúde. Entretanto, um ambiente competitivo pode não eliminar as incertezas inerentes aos serviços de saúde — a assimetria de informação entre consumidores, provedores e compradores (seguradoras e empregadores). Mais ainda, os benefícios dos consumidores vão depender do acesso à informação, assim como da compatibilidade e do alinhamento entre os incentivos percebidos pelos diversos agentes econômicos que atuam no setor. A regulação tem papel fundamental e pode afetar a competição nesse mercado, restringindo as escolhas do consumidor, aumentando os custos do seguro e inibindo os incentivos a inovações.

O capítulo inicia abordando a transformação ocorrida na estrutura dessa indústria nos Estados Unidos após a introdução do *managed care* nos anos 1980 e os desafios enfrentados pelas autoridades americanas antitruste. Com base na experiência americana, são analisados os principais aspectos envolvidos nos atos de concentração horizontal, no exercício do poder de monopólio e nas condutas verticais. As limitações analíticas vão além da compreensão teórica do setor e abrangem também os estudos empíricos. A autora faz uma extensa revisão da literatura empírica, que, embora incipiente e ainda tecnicamente defasada em relação a outras áreas da economia antitruste, pode oferecer procedimentos estimativos bastante elucidativos para a grande maioria de casos. Para tal, revisa os métodos utilizados pelas cortes americanas para a análise desses atos e condutas, além de apresentar as críticas e alternativas de pesquisadores e especialistas.

## AGRADECIMENTOS

Conforme demonstra a sucinta descrição que fizemos, o projeto que culminou nesta publicação foi bastante ambicioso, mas ainda considerando o prazo de execução. Para tanto, sua consecução contou com a colaboração de muitos.

Temos certeza de que os estudos aqui publicados atingiram os objetivos propostos com uma excelente qualidade acadêmica, graças à excelência dos seus autores e às contribuições dos especialistas e praticantes do SBDC durante o processo de disseminação dos estudos. Esse processo, muitas vezes exaustivo, envolveu a passagem de todos os estudos pela avaliação de especialistas e integrantes do SBDC, além dos dois eventos mencionados. A Coordenação Técnica é muito grata a todos que nele colaboraram, os técnicos do Ipea, da SDE, da Seae e do Cade, incluindo seus conselheiros, que junto com outros colegas da academia apostaram na qualidade do projeto e encontraram nas suas agendas atribuídas um espaço para formularem suas valiosas contribuições.

Todo o processo foi bastante rico e gerou um esforço contínuo por parte dos autores para adicionar, reformular e reescrever passagens dos estudos de forma a garantir-lhes a aderência desejada ao SBDC. Assim, além da competência, esta publicação foi beneficiada pelo alto profissionalismo dos autores que, sem exceção, empenharam-se nas revisões quase sempre solicitadas em prazos exígues. Para a Coordenação Técnica foi um grande prazer desfrutar do convívio com esses profissionais. Mais uma vez, agradecemos sua colaboração.

Para realizar este projeto no prazo proposto, contamos com o apoio das equipes administrativas da SDE e Anpec, lideradas, respectivamente, por Francisco Rogério Lima da Silva e Ângela Moura. Destacamos a eficiência de Marcos Hecksher e da equipe do Serviço Editorial do Ipea na coordenação da editoração deste livro. Somos gratos pela dedicação de todos esses profissionais.

Ao presidente do Ipea, Glauco Arbix, e ao diretor de Estudos Macroeconômicos do Ipea, Paulo Levy, agradecemos pelo entusiasmo e motivação que sempre dedicaram ao projeto. Nossos agradecimentos se estendem da mesma forma ao apoio irrestrito de Mauro Borges Lemos e Fabiana Fontes Rocha, que se sucederam na Secretaria-Executiva da Anpec.

A percepção aguçada do secretário Daniel Goldberg sobre as oportunidades de aprimoramento do SBDC permitiu que este projeto fosse concebido, e sua generosa influência foi do formato ao conteúdo dos estudos, incluindo revisões escritas a autores. Agradecemos pelas contribuições e em especial pela inabalável confiança que nos depositou na Coordenação Técnica.

O canal entre a literatura científica e a prática da análise antitruste já foi aberto. Ainda falta muito a realizar, mas acreditamos que a contribuição desta publicação será crucial nesse processo contínuo de aperfeiçoamento do SBDC.

## BIBLIOGRAFIA

- BAKER, J., RUBINFELD, D. Empirical methods in antitrust litigation: review and critique. *American Law and Economics Review*, v. 1, n. 1-2, p. 386-435, 1999.
- BID, OCDE. *Lei e política de concorrência no Brasil: uma revisão pelos pares*, 2005.
- FIUZA, E. P. S. *Três ensaios sobre diferenciação de produto*. Rio de Janeiro: FGV/EPGE, 2001 (Tese de Doutorado).
- LECG Ltd. *Quantitative techniques in competition analysis*. Londres: Office of Fair Trading, 1999 (OFT Research Paper 17).
- SALGADO, L. H., SEROA DA MOTTA, R. (eds.). *Marcos regulatórios no Brasil: o que foi feito e o que falta fazer*. Rio de Janeiro: Ipea, 2005.



# **Modelos Econométricos Estruturais**

parte 1



A estimativa de modelos estruturais de demanda, além de extremamente útil para a delimitação de mercados relevantes, é fundamental para a simulação do novo equilíbrio de mercado nos casos em que a análise de atos de concentração efetuada pela autoridade antitruste chega a um ponto em que é necessário avaliar quantitativamente o real grau de rivalidade entre as firmas do mercado relevante e as consequências do ato de concentração para a relação de forças entre essas firmas, tendo em vista a substituibilidade dos bens produzidos por elas. A estimativa de demanda também é fundamental, caso se queira estimar parâmetros de conduta e, assim, testar condutas certificadas em mercados específicos.

Ao longo de sua curta história, o Sistema Brasileiro de Defesa da Concorrência (SBDC) foi percebendo como estimativas apresentadas por pareceristas, a partir de diferentes hipóteses a respeito de ordenação da escolha do consumidor, bens alternativos e situação de compra, poderiam trazer resultados bem diferentes. Reconheceu-se a importância da capacitação da autoridade antitruste para avaliar criticamente as vantagens e desvantagens de cada estimativa, bem como a confiabilidade das previsões resultantes.

O Capítulo 1, de Huse e Salvo, após apresentar brevemente o problema da simulação dos efeitos da fusão, segue resenhando as principais especificações funcionais utilizadas em estudos de demanda aplicados à organização industrial. Com relação ao tratamento dado ao problema da dimensionalidade (o número de elasticidades cruzadas cresce quadraticamente com o número de bens do mercado), os modelos foram agrupados em modelos contínuos (em geral, restritos por hipóteses de orçamento multiestágio) e modelos de escolha discreta. O primeiro grupo lida com a maldição da dimensionalidade assumindo alguns pressupostos de agregação dos bens em categorias e de

1  
0  
Capítulo

separabilidade das preferências, o que permite seqüenciar a decisão de consumo em conjuntos de escolha cada vez mais desagregados, até chegar à substituição entre marcas, produtos ou ofertantes de interesse. O segundo grupo resolve a dimensionalidade modelando a demanda como função de um conjunto pequeno de características, de modo que as elasticidades passam a ser, elas também, funções desse mesmo conjunto.

Os autores discutem, em seguida, os desafios econôméticos clássicos que surgem na estimação da demanda, principalmente: *a*) a simultaneidade, que causa viés e inconsistência e deve ser corrigida por variáveis instrumentais adequadas; e *b*) a identificação dos parâmetros estruturais de oferta (custos e conduta). Para o segundo desafio, são examinadas três abordagens: *a*) estimativa simultânea (em dois estágios) de demanda, custos e conduta, estimando esta última como um parâmetro livre ou em função da estrutura de mercado (por exemplo, concentração); *b*) seleção entre modelos comportamentais; e *c*) assumir, em vez de estimar, o modelo de conduta.

Por fim, atentos aos critérios de tempestividade, parcimônia, operacionalidade e simplicidade de interpretação, tão caros ao SBDC, os autores cobrem brevemente os modelos PCAIDS e ALM, de simples e rápida operacionalização.

## ESTIMAÇÃO E IDENTIFICAÇÃO DE DEMANDA E DE OFERTA\*

Cristian Huse  
Da London School of Economics  
Alberto Salvo  
Da London School of Economics

### 1.1 INTRODUÇÃO

A estimativa de modelos estruturais de indústrias, objeto de análise deste estudo, tem papel central na análise antitruste. Além do interesse tradicional por esses modelos, pelo fato de fornecerem os determinantes da demanda por um bem, eles são, hoje em dia, parte fundamental da análise de atos de concentração: sua relevância vem do fato de permitirem investigar em que medida os efeitos anticompetitivos de, por exemplo, uma fusão, reduzem o bem-estar dos consumidores, mesmo que ocorram ganhos de eficiência produtiva.

A quantidade e a complexidade das informações referentes às interações entre firmas em um dado mercado, bem como entre estas e os consumidores, tornam necessário o uso de ferramentas quantitativas que permitam um melhor entendimento dessa realidade. Modelos de demanda são, portanto, uma ferramenta fundamental na definição de mercados relevantes, além de parte crucial na obtenção de equilíbrios pós-fusão, efeitos de integração vertical, reações a mudanças de preços etc. Adicionalmente, tais modelos fornecem como importante subproduto a matriz de elasticidades-preço de demanda dos produtos em um determinado mercado. O conceito de elasticidade é fundamental na análise antitruste: enquanto a elasticidade-preço própria da demanda de um bem fornece a variação da demanda por esse bem dada uma pequena variação de seus preços, refletindo a sensibilidade dos consumidores à alteração de preços desse bem, a elasticidade-preço

---

\* Este trabalho, fruto de nossas atividades de pesquisa e ensino nas nossas respectivas instituições, foi elaborado como parte do convênio Ipea/Anpec/SDE. Agradecemos a Eduardo Fiúza, Alessandro Oliveira e Ronaldo Seroa da Motta por seus comentários e sugestões.

cruzada (da demanda de um bem com relação ao preço de outro bem) fornece a variação da demanda do primeiro bem dada uma pequena variação do preço do segundo bem, refletindo o grau de substituibilidade entre esses dois bens e, em última análise, a dimensão do mercado relevante — definido pelo lado da demanda — em uma análise antitruste.

O objetivo central do presente trabalho é oferecer uma visão geral dos modelos de demanda mais relevantes atualmente, tanto em estudos teóricos, como em aplicados, com o intuito de fornecer subsídios ao Sistema Brasileiro de Defesa da Concorrência (SBDC) na condução de seu trabalho. Para isso, adotamos uma sistematização informal desses modelos, iniciando pela discussão de como discernir entre produtos homogêneos e diferenciados. Os desenvolvimentos mais recentes na modelagem de produtos homogêneos datam da década de 1980, e sua descrição é, portanto, breve, inclusive pelo fato de discussões sobre modelos como o Sistema de Demanda Ideal [Almost Ideal Demand System (AIDS)] e o modelo de despesa linear [Linear Expenditure System (LES)] já se encontrarem em livros-texto como o de Deaton e Muellbauer (1980*b*).

Tratamos em seguida da modelagem de produtos diferenciados — modelos que têm recebido crescente atenção, tanto na literatura acadêmica, como no dia-a-dia de autoridades antitruste, pela parcela significativa da economia que atualmente representam, bem como em função da necessidade de se modelar o efeito reputacional de marcas, publicidade, e de se levar em conta a heterogeneidade dos consumidores. Na última década, duas classes de modelos têm prevalecido na literatura. A primeira classe — modelos de escolha contínua — desenvolvida em Hausman, Leonard e Zona (1994) e Hausman (1997*a*), adota o uso de um modelo multiestágio para estimar a demanda de bens diferenciados. As idéias fundamentais presentes nessa classe de modelos são as de agregação e separabilidade, que justificam o uso de diferentes estágios de demanda. O primeiro estágio é a demanda pela categoria onde se insere o bem que é o objeto de estudo (automóveis de passeio, cereais matinais). O estágio intermediário considera a substituibilidade entre diferentes segmentos de mercado (automóveis pequenos ou grandes, cereais naturais ou infantis), enquanto o estágio final compreende as diversas marcas dentro de um determinado segmento de mercado. Cada estágio do sistema de demanda é estimado com o uso de um modelo suficientemente flexível para gerar padrões realistas de substituição entre os produtos. A segmentação da decisão do consumidor por um determinado produto permite ainda uma redução do número de parâmetros que se necessita estimar (na proporção do inverso do número de estágios considerados), de forma que, com um número suficientemente pequeno de produtos, ou um número suficientemente grande de estágios, o problema de estimar a demanda

é factível com o uso de modelos como o AIDS de Deaton e Muellbauer (1980a). Infelizmente, no entanto, a implementação de modelos contínuos deixa de se tornar factível em alguns casos, pois o número de parâmetros que se necessita estimar aumenta na razão do quadrado do número de produtos em um dado segmento. Por exemplo, no caso de se ter um mercado com 300 modelos de automóveis [tal como em Berry, Levinsohn e Pakes (1995)], tornar-se-ia necessário estimar 90 mil elasticidades-preço — mesmo ao se impor restrições oriundas da teoria econômica, o número de parâmetros é ainda muito grande para ser considerado factível de se estimar.

Uma solução clássica para o problema da dimensionalidade é a adoção de modelos de escolha discreta [McFadden (1974, 1978a e 1984), Cardell (1989), Berry (1994), Berry, Levinsohn e Pakes (1995) e Nevo (2001)]. Nessa classe de modelos o problema de dimensionalidade é resolvido projetando-se os produtos em um espaço de características, com a consequência de reduzir um problema de dimensão igual ao quadrado do número de produtos para um problema de dimensão igual à dimensão do espaço de características. Alguns desses modelos de escolha discreta são ainda extremamente restritivos, como será aprofundado ao longo do estudo, e devem ser empregados com cuidado, por exemplo, ao se simular fusões. Mais recentemente, uma nova geração de modelos de escolha discreta vem tendo um maior sucesso na literatura, gerando padrões de substituibilidade flexíveis e condizentes com o que se espera na teoria e na prática. No entanto, a aplicabilidade de tais modelos ainda esbarra no fato de a sua implementação ser não-trivial.

Um componente fundamental de qualquer modelo de demanda que se deseje estimar é a sua identificação. A importância da identificação da demanda é crucial para que se obtenham resultados numéricos que façam sentido e sejam consistentes do ponto de vista teórico: em uma especificação típica de demanda, preços são endógenos, ou seja, os preços observados decorrem da interação entre produtores e consumidores, sendo essencial distinguir as alterações de preços e de quantidades que resultam do deslocamento da curva de oferta das alterações de preços e de quantidades que resultam do deslocamento da curva de demanda. Como consequência de tal fato, métodos “tradicionais” de estimativa resultam em coeficientes-preço menos negativos (ou “viesados”) do que na realidade eles são, e a identificação de demanda visa exatamente à obtenção de estimativas livres desse viés (ou “consistentes”, como são denominados pelos economistas). Por essa razão, discutimos diversas classes de instrumentos de demanda, abordando com cuidado as premissas identificadoras subjacentes a elas, bem como casos em que tais instrumentos foram aplicados.

Uma vez estudada a demanda, discutimos a obtenção do equilíbrio de mercado e como proceder à recuperação de parâmetros estruturais de oferta (custos e conduta) a partir de dados observados. Abordamos as diferentes estratégias que têm sido utilizadas na literatura como, por exemplo, *a)* a estimação simultânea (ou em estágios) de demanda, de custos e de conduta, considerando este último como um parâmetro livre, ou em função da estrutura de mercado (por exemplo, concentração); *b)* o método de “seleção entre modelos comportamentais”; e *c)* assumir em vez de estimar o modelo de conduta (hipótese de concorrência).

A seguir, tratamos da experiência internacional do antitruste. Em particular, fazemos uma breve visita aos modelos calibrados de simulação de efeitos unilaterais, como o Proporcionality-Calibrated AIDS (PCAIDS) (adotado pelo SDC da Nova Zelândia) e o ALM (adotado pelo SDC dos Estados Unidos) — objetos de estudo mais detalhado do Capítulo 2 deste livro — que surgem como resposta a uma característica institucional de SDCs, que é o pouco tempo concedido à análise dos efeitos de uma fusão. Como o pouco tempo disponível dificulta em muito a implementação de métodos mais demandantes em termos de técnica e de dados, os modelos discutidos foram desenvolvidos com o intuito de serem parcimoniosos e de cálculo rápido, ainda que sob o risco de serem considerados demasiadamente simples.

Tendo coberto em detalhe a parte conceitual dos métodos empíricos em organização industrial, reservamos uma seção para enfatizar as considerações práticas na estimação estrutural. Tal seção visa fazer a conexão entre os modelos teóricos e o dia-a-dia do SBDC, em que os dados disponíveis e os prazos nem sempre atuam a favor do analista. Iniciamos tal discussão detalhando as necessidades de dados para os principais modelos anteriormente descritos, e descrevendo as várias dimensões em que esses dados podem ser agregados (ou desagregados), enfatizando suas implicações. A seguir, fazemos uma comparação “horizontal” entre os modelos de demanda mais simples, considerando especificações funcionais alternativas (curvas de demanda lineares, logarítmicas etc.). Tecemos comentários também sobre a escolha da variável dependente (preço ou quantidade) e de que forma as estimativas de uma dada especificação estão relacionadas com a “regressão reversa”. Entre outras questões, discutimos a importância de testes de especificação (por exemplo, testes de restrições “superidentificáveis”, testes de endogeneidade), a escolha de regressores, a estrutura do erro (como tratar a correlação serial, por exemplo) e a especificação de modelos dinâmicos. Fazemos também uma comparação entre modelos de demanda mais sofisticados.

Durante todo o trabalho, faremos referência a artigos científicos em maior ou menor detalhe por acreditarmos que a interação entre teoria e prática é fundamental na construção de um SDC saudável e atualizado, mas sem deixar de lado

os critérios de *precisão de previsão, tempestividade, parcimônia, operacionalidade, e simplicidade de interpretação* — equilíbrio nem sempre fácil de ser obtido. Como exemplo de artigos, citamos Genesove e Mullin (1998) e Parker e Röller (1997), assim como as considerações metodológicas de Bresnahan (1982 e 1989) e Reiss e Wolak (2002), ou o estudo de hipóteses alternativas de concorrência, como em Gasmi, Laffont e Vuong (1990 e 1992).

Por fim, tecemos algumas recomendações de política no âmbito do SBDC — nossas principais recomendações são no sentido do seu fortalecimento institucional. Entendemos que um SBDC não pode prescindir de quadros bem formados e em processo de constante aperfeiçoamento: o material coberto neste trabalho mostra a importância do conhecimento de organização industrial e econometria, uma vez que se tenha optado pela utilização de modelos estruturais. Do ponto de vista da adoção de modelos descritos no trabalho, entendemos que, num primeiro estágio, o SBDC veja como mais conveniente adotar métodos parcimoniosos como o PCAIDS e o ALM. No entanto, sob os critérios de *precisão de previsão, tempestividade, parcimônia, operacionalidade e simplicidade de interpretação*, entendemos que, num horizonte de longo prazo, métodos mais adequados seriam, por exemplo, o modelo *logit* aninhado [Nested Logit Model (NL)] e o sistema de estimação em estágios, desde que adotados com as precauções detalhadas no texto.

Antes de proceder à análise de demanda propriamente dita, ilustramos com um exemplo estilizado os principais passos (e componentes) de um estudo de previsão de efeitos unilaterais de uma fusão, com o intuito de dar um sabor dos ingredientes a serem estudados em seguida.

### **1.1.1 Ilustração: Como as elasticidades estimadas podem ser utilizadas para prever os efeitos “unilaterais” de uma fusão**

Com o intuito de motivar a importância da estimativa de demanda, ilustramos aqui como as elasticidades podem ser utilizadas para prever os efeitos de um ato de concentração sobre preços, decorrentes do exercício unilateral de poder de mercado.<sup>1</sup> Consideraremos uma indústria de bens diferenciados onde as firmas concorrem à Bertrand, cada firma tomando os preços dos bens oferecidos por firmas rivais como dados ao estabelecer seus preços. Delegaremos a discussão da adequação dessa premissa sobre conduta, assim como possíveis testes estatísticos, a seções posteriores do trabalho.

---

1. A abordagem aqui ilustrada foi inicialmente proposta por Baker e Bresnahan (1985), sendo posteriormente estendida em Berry e Pakes (1993), Hausman, Leonard e Zona (1994) e Nevo (2000). Exemplos de indústrias estudadas incluem automobilística [Berry e Pakes (1993), Ivaldi e Verboven (2004)]; cereais [Nevo (2000)]; cerveja [Baker e Bresnahan (1985), Hausman, Leonard e Zona (1994), Pinkse e Slade (2004)]; refrigerantes [Dubé (2004)]; transporte aéreo [Peters (2001)]; e telecomunicações a longa distância [Werden e Froeb (1994)].

Começamos considerando uma indústria de bens diferenciados onde diferentes firmas detêm inicialmente apenas um produto (ou seja, uma marca) cada uma. Ao estabelecer o preço, cada firma há de equilibrar dois efeitos distintos sobre seu lucro decorrentes de um acréscimo (infinitesimal) no preço de seu produto. Por um lado, um aumento de preço resulta em um aumento de lucro a ser realizado de consumidores que continuarão comprando seu produto, apesar do aumento de preço. Esses consumidores são os chamados “consumidores inframarginais”, por desfrutarem de um excedente suficientemente alto no consumo desse produto a ponto de continuarem comprando-o ainda que a um preço mais alto. Por outro lado, o aumento de preço fará com que alguns consumidores — os chamados “consumidores marginais” — deixem de consumir o produto da firma em questão, seja porque irão optar agora por consumir um outro produto (marca) ou porque deixarão de comprar produtos nesse mercado (ou seja, optarão agora pelo produto externo). Claramente, enquanto o primeiro efeito sobre o lucro é positivo, o segundo efeito é negativo: o preço escolhido pela firma em equilíbrio é aquele que balanceia os dois efeitos.

Esses dois efeitos podem ser vistos imediatamente a partir do problema de otimização da firma e a correspondente condição de primeira ordem. Assumamos que existam  $n$  firmas, cada uma com um único produto, indexados por  $i = 1, \dots, n$ . O lucro da firma  $i$  (proprietária do produto  $i$ ) é dado por  $(p_i - c_i) D_i(p)$ , onde  $p_i$  é o preço,  $c_i$  é o custo marginal (assumindo que seja constante, isto é, que não varie em função de quantidade no intervalo relevante), e  $q_i := D_i(p)$  é a demanda pelo produto quando os preços de todos os produtos no mercado são dados pelo vetor de preços  $p := (p_1, p_2, \dots, p_n)$ . Diante da premissa de concorrência em preços (Bertrand),<sup>2</sup> a condição de primeira ordem (CPO) da firma  $i$  é

$$D_i(p) + (p_i - c_i) \frac{\partial D_i(p)}{\partial p_i} = 0 \quad (1)$$

Ao passo que o segundo termo captura a perda de vendas para os “consumidores marginais” (o termo é negativo para funções de demanda decrescentes em preço, e onde há uma margem  $p_i - c_i$  positiva), o primeiro termo reflete o efeito do

---

2. Portanto  $\partial p_j / \partial p_i = 0$  para  $j \neq i$ .

aumento de preço nas vendas para os “consumidores inframarginais” remanescentes. A CPO (1) pode ser reescrita na forma de *markup* preço-custo no lado esquerdo:

$$\frac{p_i - c_i}{p_i} = \frac{-D_i(p)}{\frac{\partial D_i(p)}{\partial p_i} p_i} = -\frac{1}{\eta_{ii}(p)} \quad (2)$$

Referimo-nos à recíproca da expressão do lado direito como a elasticidade-preço própria de demanda: esta é claramente uma função de preços, denotada por  $\eta_{ii}(p)$ . Corresponde, em valores absolutos, à queda percentual na demanda pelo

produto  $i$  (equivalentemente  $\partial \ln D_i(p) = \frac{\partial D_i(p)}{D_i(p)}$ ) quando o preço do produto

$i$  é aumentado em 1% (isto é  $\partial \ln p_i = \frac{\partial p_i}{p_i} = 1\%$ ). A rigor, trata-se da razão entre

$\partial \ln D_i(p)$  e  $\partial \ln p_i$  para um aumento infinitesimal de preço  $\partial \ln p_i \rightarrow 0$ , ou seja, na margem.

A CPO é satisfeita no equilíbrio: o equilíbrio é dado por um sistema de equações composto de uma CPO para cada produto. A equação (2) pode então ser usada para estimar os *markups* preço-custo a partir das elasticidades-preço estimadas, assumindo que o modelo Bertrand seja um modelo suficientemente apropriado de comportamento na indústria. Alternativamente, as elasticidades-preço podem ser utilizadas junto com uma medida direta de *markups* (obtida das firmas) de forma a testar a premissa de conduta Bertrand, como trataremos adiante.

Vê-se, claramente, da CPO (2) que no caso de firmas proprietárias de apenas um produto no mercado, um produto com alto *markup* em equilíbrio estará associado a uma baixa elasticidade-preço própria em equilíbrio. A intuição desse resultado é clara a partir de (1): se, no equilíbrio, um produto usufrui de alta margem  $p_i - c_i$  é porque a proporção de consumidores marginais necessária a compensar o efeito “inframarginal” positivo de um aumento de preços é baixa. Por exemplo, considere um produto com preço de 20 e custo de 10 cujo consumo no mercado seja de 1.000 unidades em equilíbrio. A razão pela qual a firma precisa o produto em 20 é porque o coeficiente angular da curva de demanda nesse ponto

da curva é de  $\frac{\partial D_i(p)}{\partial p_i} = -100$ , tal que  $1.000 + (20 - 10).(-100) = 0$ . Ou seja, o

preço não é maior do que 20 porque um aumento (marginal) unitário acarretaria a perda de vendas de 100 unidades, resultando em um aumento de preços não lucrativo. Considere agora outro produto com preço de 30 e o mesmo custo marginal de 10 com consumo da mesma quantidade de 1.000 unidades em equilíbrio. A razão pela qual cobra-se um preço mais elevado por esse outro produto reside no fato de a curva de demanda ter coeficiente angular de apenas  $-50$ , já que  $1.000 + (30 - 10) \cdot (-50) = 0$ . Ou seja, quanto maior a margem preço-custo, menor é a proporção de vendas marginais necessária para tornar o acréscimo de preço não-lucrativo.

Fica claro a partir dessas considerações que (em condições normais) são os consumidores marginais, e não os consumidores inframarginais “leais”, que a firma considera ao fixar seus preços. Assim, o que restringe o aumento de preços são as alternativas dos consumidores ao se depararem com esse aumento: o quanto atraente é a opção de deixar de comprar o produto (ou reduzir sua compra) cujo preço está sendo acrescido, seja para deixar de consumir no mercado (ou seja, optar por consumir o produto externo) ou optar pelo consumo de produtos rivais, ou substitutos. Dessa forma, a existência de produtos substitutos restringe o preço de um produto: quanto maior a proporção de consumidores marginais prestes a trocar sua decisão de consumo por um produto rival, maior é a substituibilidade, ou a “proximidade”, entre o produto cujo preço está sendo acrescido e os produtos-destino dos consumidores marginais. Enquanto a elasticidade-preço própria captura a proporção de consumidores marginais, a elasticidade-preço cruzada reflete a proximidade entre dois produtos. Define-se a elasticidade-preço cruzada de demanda pelo produto  $i$  com relação ao preço do produto  $j$ , denotada  $\eta_{ij}(p)$ , como sendo a alteração percentual na demanda pelo produto  $i$  que resulta do aumento

de 1% no preço do produto  $j$ , ou seja  $\eta_{ij}(p) := \frac{\partial \ln D_i(p)}{\partial \ln p_j} = \frac{\partial D_i(p)}{\partial p_j} \frac{p_j}{D_i(p)}$ .

*Caeteribus paribus*, quanto maior for a proporção de consumidores marginais do produto  $j$  que trocar pelo produto  $i$  quando o produto  $j$  sofrer um acréscimo de preço, maior será o valor de  $\eta_{ij}$ .

Uma fusão entre firmas proprietárias de produtos suficientemente próximos poderá, assim, levar a um aumento não-desprezível de preços, na medida em que uma parcela importante da restrição competitiva antes imposta pelo(s) produto(s) substituto(s) é eliminada. Considere a fusão entre duas empresas 1 e 2, proprietárias de produtos substitutos 1 e 2, respectivamente. A firma proprietária do produto 1 que, antes da fusão, não aumentava o preço além de um certo patamar, devido à existência de um suficiente número de consumidores marginais prestes a trocar o

produto 1 pelo produto 2, após a fusão já não perderá as vendas advindas desses consumidores caso eleve o preço do produto 1. A resposta ótima à fusão — na ausência de alterações de custo, como veremos adiante — será então aumentar o preço do produto 1 (e muito provavelmente aumentar também o preço do produto 2). A atenuação (ou internalização) das restrições competitivas advinda do ato de concentração, com o consequente aumento de preços, é o chamado “efeito unilateral” sobre preços.

Esse efeito pode novamente ser evidenciado pela condição de primeira ordem da firma fusionada. Lembrando nosso ponto de partida, onde cada firma detém inicialmente um produto (uma marca), consideremos a fusão entre firmas 1 e 2. A firma fusionada irá agora escolher os preços dos produtos 1 e 2 de forma a maximizar seu lucro pós-fusão, equivalente a  $(p_1 - c_1) D_1(p) + (p_2 - c_2) D_2(p)$ . A CPO com relação a  $p_1$  (inverter os algarismos 1 e 2 para obter a CPO em relação a  $p_2$ ) é, agora:

$$D_1(p) + (p_1 - \tilde{c}_1) \frac{\partial D_1(p)}{\partial p_1} + (p_2 - \tilde{c}_2) \frac{\partial D_2(p)}{\partial p_1} = 0 \quad (3)$$

onde  $\tilde{c}_i$  corresponde ao respectivo custo marginal pós-fusão (como veremos, a fusão poderá ou não alterar os custos marginais dos produtos envolvidos, a parte marginal das chamadas “eficiências” da fusão). Uma rápida comparação entre a CPO da entidade fusionada (3) e a CPO da firma quando independente (1) aponta para a presença de um terceiro termo, referente ao efeito cruzado: alguns consumidores marginais do produto 1, deparados com um aumento no preço do produto 1, trocarão sua decisão de consumo pelo produto 2, e essa troca que ocorre em favor do produto 2 já não conta como uma perda para a firma fusionada quando esta estabelece o preço do produto 1. Em outras palavras, ao ser internalizada pela fusão, essa troca não mais restringe a precificação do produto 1. Como fizemos, a CPO pode ser reescrita em termos de *markups* e elasticidades. Multiplicando os dois lados de (3) por  $p_1$  e dividindo pela receita total da indústria,

$\sum_{i=1}^n p_i D_i(p)$ , obtemos:

$$\begin{aligned} & \frac{p_1 D_1(p)}{\sum_{i=1}^n p_i D_i(p)} + \frac{p_1 D_1(p)}{\sum_{i=1}^n p_i D_i(p)} \frac{p_1 - \tilde{c}_1}{p_1} \frac{p_1}{D_1(p)} \frac{\partial D_1(p)}{\partial p_1} + \\ & + \frac{p_2 D_2(p)}{\sum_{i=1}^n p_i D_i(p)} \frac{p_2 - \tilde{c}_2}{p_2} \frac{p_1}{D_2(p)} \frac{\partial D_2(p)}{\partial p_1} = 0 \end{aligned}$$

ou, denotando a fatia (participação) da receita total da indústria correspondente ao produto  $i$  (*share* de valor) como  $s_i(p) = \frac{p_i D_i(p)}{\sum_{j=1}^n p_j D_j(p)}$ , e lembrando a definição das elasticidades-preço própria e cruzada, temos:<sup>3</sup>

$$s_1(p) + s_1(p) \frac{p_1 - \tilde{c}_1}{p_1} \eta_{11}(p) + s_2(p) \frac{p_2 - \tilde{c}_2}{p_2} \eta_{21}(p) = 0 \quad (4)$$

As CPOs das  $n - 2$  firmas remanescentes, cada uma detendo apenas um produto, ainda são dadas pela equação (2), a qual podemos reescrever como [simplesmente multiplique os dois lados por  $s_i(p)$ ]:

$$s_i(p) + s_i(p) \frac{p_i - c_i}{p_i} \eta_{ii}(p) = 0 \quad (5)$$

Assim como na situação pré-fusão, os preços em equilíbrio dos  $n$  produtos satisfarão simultaneamente o sistema composto por  $n$  equações dado por (4) para o produto 1, uma equação similar a (4) para o produto 2, conforme mencionado, e  $n - 2$  equações (5) para os  $n - 2$  produtos remanescentes. Esse sistema pode ser expresso em formato matricial:

$$\begin{pmatrix} s_1(p) \\ s_2(p) \\ s_3(p) \\ \vdots \\ s_n(p) \end{pmatrix} + \begin{pmatrix} \eta_{11}(p) & \eta_{21}(p) & 0 & \cdots & 0 \\ \eta_{12}(p) & \eta_{22}(p) & 0 & \cdots & 0 \\ 0 & 0 & \eta_{33}(p) & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \eta_{nn}(p) \end{pmatrix} \begin{pmatrix} s_1(p) \frac{p_1 - \tilde{c}_1}{p_1} \\ s_2(p) \frac{p_2 - \tilde{c}_2}{p_2} \\ s_3(p) \frac{p_3 - c_3}{p_3} \\ \vdots \\ s_n(p) \frac{p_n - c_n}{p_n} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (6)$$

3. A CPO (4) pode ser rearranjada tal que o *markup* do produto 1 no equilíbrio pós-fusão é dado por

$$\frac{p_1 - \tilde{c}_1}{p_1} = -\frac{1}{\eta_{11}(p)} - \frac{s_2(p)}{s_1(p)} \frac{p_2 - \tilde{c}_2}{p_2} \frac{\eta_{21}(p)}{\eta_{11}(p)}$$

Fica claro a partir desta equação que o *markup* do produto 1 será tanto maior quanto: *a*) maior for a elasticidade-preço cruzada do produto 2 em relação ao (preço do) produto 1,  $\eta_{21}$ ; *b*) maior for o *markup* do produto cruzado 2,  $(p_2 - \tilde{c}_2)/p_2$ ; e *c*) maior for a participação da marca 2 em relação à participação da marca 1,  $s_2(p)/s_1(p)$ .

ou:

$$s(p) + E(p)' \omega = 0 \quad (7)$$

onde  $s(p)$  é um vetor de dimensão  $n \times 1$  de *shares* de valor dos produtos,  $E(p)$  é uma matriz de dimensão  $n \times n$  de elasticidades onde os elementos fora da diagonal equivalem a 0 a não ser pelas elasticidades cruzadas referentes aos produtos 1 e 2 (ou seja,  $\eta_{21}$  e  $\eta_{12}$ ), e  $\omega$  é um vetor de dimensão  $n \times 1$  referente aos *markups* de preço-custo multiplicados pelos *shares* de valor. Assumindo que o analista tenha em mãos estimativas para as elasticidades e os *shares* de valor após a fusão, a solução desse sistema passa pela inversão da (transposta) da matriz de elasticidades  $E(p)'$ , de forma a obter o vetor de *markup*  $\times$  *share* de valor  $\omega$ , conforme segue:

$$\omega = -\left( E(p)' \right)^{-1} s(p) \quad (8)$$

Os *markups* pós-fusão podem ser calculados a partir da divisão elemento-por-elemento do vetor de *markup*  $\times$  *share* de valor  $\omega$  por  $s(p)$ , ou seja:

$$\frac{p_i - \tilde{c}_i}{p_i} = \frac{\omega_i}{s_i(p)} \quad (9)$$

Procedemos a uma ilustração numérica dos efeitos de uma fusão sobre preços a partir de elasticidades estimadas e *shares* de valor observadas pré-fusão. Antes, porém, convém tecer um comentário a respeito do sistema (7), que simplesmente empilha as condições de primeira ordem de todos os produtos no mercado. Quaisquer outras estruturas de mercado, referentes à propriedade dos  $n$  produtos por  $m \leq n$  firmas, podem facilmente ser acomodadas no sistema (7), simplesmente alterando a matriz de elasticidades. Por exemplo, caso os produtos 1, 2 e 4 pertençam à mesma firma, as linhas 1, 2 e 4 da matriz referentes às CPOs dos produtos 1, 2 e 4 consistirão em 3 elementos (elasticidades) não-nulos. Relevante para estabelecer o preço do produto 1, por exemplo, serão a elasticidade-preço própria do produto 1 ( $\eta_{11}$ ) e as elasticidades-preço cruzadas dos produtos 2 e 4 com relação ao preço do produto 1 ( $\eta_{21}$  e  $\eta_{41}$ ).

*Ilustração numérica: quantificando efeitos unilaterais na análise de um ato de concentração em uma indústria de bens diferenciados, sem ou com “eficiências” previstas.*

Como podemos então determinar os efeitos estimados de uma fusão proposta sobre preços? Tipicamente, esse exercício consiste em duas etapas, a serem elaboradas em seções posteriores do presente trabalho:

a) Estima-se o sistema de demanda a partir do equilíbrio pré-fusão, de forma a obter estimativas consistentes das elasticidades-preço próprias e cruzadas.

b) Assume-se um modelo comportamental (que poderá, em princípio, variar com a fusão) e calcula-se, a partir do sistema de demanda estimado na primeira etapa, as elasticidades-preço e *shares* de valor para cada produto para o equilíbrio após a fusão. Quanto ao modelo comportamental, manteremos aqui a premissa de conduta Bertrand como sendo suficientemente apropriada, utilizando assim as CPOs recém-deduzidas; discutiremos essa premissa, e também seu teste, posteriormente. Quanto à estimativa das elasticidades e *shares* pós-fusão, uma possível aproximação é utilizar as mesmas elasticidades e *shares* da situação pré-fusão para a situação pós-fusão.<sup>4</sup> Evidentemente trata-se de uma aproximação de primeira ordem (pois, ao alterar preços, a fusão impactará elasticidades e *shares*), que poderá ser mais adequada na medida em que a variação de preços seja pequena. Efetua-se, então, o cálculo dos *markups* preço-custo utilizando a expressão (8) para o vetor de *markup × share* de valor  $\omega$ , seguido de (9).

Assumimos agora que o sistema de demanda, na primeira etapa, já tenha sido estimado. Mantemos nosso exemplo de uma indústria com  $n$  bens diferenciados, inicialmente com  $n$  firmas, cada uma proprietária de um produto. Propõe-se a fusão das firmas 1 e 2. Suponhamos que a seguinte situação exista para os produtos 1 e 2:

Dados referentes à situação pré-fusão	Produto 1	Produto 2
Share de valor	20%	10%
Elasticidade-preço própria de demanda	-2,0	-3,0
Elasticidade-preço cruzada de demanda com relação ao preço do outro produto	0,20	0,20

4. Observe que especificamente na função de demanda log-linear (isoelástica) as elasticidades são invariantes a preços. Uma alternativa a essa aproximação consiste em utilizar as funções de demanda estimadas e as expressões para as elasticidades resultantes para resolver (numericamente) o novo sistema de CPO. Contudo, vale apontar que a mudança no valor das elasticidades (e dos *shares*) em função da fusão irá depender sensivelmente da forma funcional adotada para estimar demanda (ver adiante). Essa forma funcional adotada, qualquer que seja, irá restringir a maneira como as elasticidades variam na medida em que os preços se alteram, não necessariamente restringindo-a “menos” do que a função de demanda isoelástica as restringe (a ponto de serem constantes em preços). No entanto, na estimação de uma função de demanda AIDS (ver adiante) para molhos para massas, Capps, Church e Love (2003) obtém diferenças não-triviais nos efeitos unilaterais estimados com a aproximação em comparação com os efeitos estimados mediante a alternativa numérica “exata”. Remetemos o leitor ao Capítulo 2 deste livro, que trata de simulação.

Os *markups* pré-fusão para os produtos 1 e 2, usando (2) — já que antes da fusão ora proposta esses produtos pertenciam a firmas donas de apenas um produto — correspondem simplesmente ao valor recíproco das respectivas elasticidades-preço próprias (multiplicados por -1):

$$\frac{p_1^{pre} - c_1}{p_1^{pre}} = -\frac{1}{-2.0} = 50\% \quad \text{e} \quad \frac{p_2^{pre} - c_2}{p_2^{pre}} = -\frac{1}{-3.0} = 33,3\% \quad (10)$$

Aproximando os *shares* e as elasticidades pós-fusão por seus respectivos valores pré-fusão, as primeiras duas equações do sistema de condições de primeira ordem (6), referentes à situação após a fusão, são:<sup>5</sup>

$$0.2 + 0.2 \frac{p_1 - \tilde{c}_1}{p_1} (-2.0) + 0.1 \frac{p_2 - \tilde{c}_2}{p_2} (0.20) = 0$$

$$0.1 + 0.2 \frac{p_1 - \tilde{c}_1}{p_1} (0.20) + 0.1 \frac{p_2 - \tilde{c}_2}{p_2} (-3.0) = 0$$

Os *markups* pós-fusão que satisfazem esse par de equações simultâneas são dados por:

$$\frac{p_1^{pos} - \tilde{c}_1}{p_1^{pos}} = 52,0\% \quad \text{e} \quad \frac{p_2^{pos} - \tilde{c}_2}{p_2^{pos}} = 40,3\% \quad (11)$$

Assim, estima-se que o *markup* do produto 1 aumente de 50% para 52% mediante a fusão, ao passo que o *markup* do produto 2 eleve-se de 33% para 40%. Com relação ao efeito da fusão sobre preços, assumindo inicialmente que a fusão não alterará custos marginais, ou seja, que *eficiências não são previstas* ( $\tilde{c}_i = c_i$ ), temos a partir dos *markups* para o produto 1 que<sup>6</sup>  $c_1 = (1 - 0.5)p_1^{pre}$  e  $\tilde{c}_1 = c_1 = (1 - 0.520)p_1^{pos}$ , e, portanto,  $p_1^{pos}/p_1^{pre} = (1 - 0.5)/(1 - 0.520) = 1.042$ ,

5. Observe que em virtude da simplificação (ver nota 4) apenas essas duas equações são necessárias para deduzir os *markups* pós-fusão dos produtos da empresa fusionada.

6. De (10) podemos escrever  $1 - c_1/p_1^{pre} = 0.50$  e, portanto,  $c_1 = (1 - 0.50)p_1^{pre}$ . De (11) podemos escrever  $1 - \tilde{c}_1/p_1^{pos} = 0.52$  e, portanto,  $\tilde{c}_1 = (1 - 0.52)p_1^{pos}$  onde, como não há eficiências,  $\tilde{c}_1 = c_1$ .

correspondendo a um aumento de 4,2% no preço do produto 1. De maneira semelhante,  $p_2^{pos} / p_2^{pre} = (1 - 0.333) / (1 - 0.403) = 1.116$ , o que se traduz em um aumento de preço de 11,6% para o produto 2.

Suponhamos que se preveja uma redução de 4% no custo marginal em decorrência da fusão. Portanto,  $\tilde{c}_1 = 0.96c_1$  e assim  $p_1^{pos} / p_1^{pre} = (0.96)(1 - 0.5) / (1 - 0.520) = 1.000$ ; de forma similar,  $p_2^{pos} / p_2^{pre} = (0.96)(1 - 0.333) / (1 - 0.403) = 1.071$ . Enquanto agora não se prevê alteração no preço do produto 1, quantifica-se um aumento de 7,1% no preço do produto 2.

Por fim, vale registrar a importância da magnitude das elasticidades-preço próprias em relação à magnitude das elasticidades cruzadas, como apontam Hausman e Leonard (1997). Repetindo-se o cálculo para elasticidades próprias de -4,0 e -6,0 para os produtos 1 e 2, mantendo as mesmas elasticidades cruzadas (e os mesmos *shares*), os aumentos previstos de preços são consideravelmente menores. O *markup* estimado para o produto 1 sobe agora de 25% pré-fusão para 25,5% pós-fusão, ao passo que o *markup* estimado para o produto 2 sobe de 16,7% pré-fusão para 18,3% pós-fusão. Para o primeiro caso em que eficiências não são previstas, os aumentos de preços para os produtos 1 e 2 são agora de apenas 0,6% e 2,1%, respectivamente, a serem comparados com 4,2% e 11,6% no exemplo acima, onde as elasticidades próprias são mais baixas. Hausman e Leonard (1997) apontam para sua experiência em estimar (para mercados nos Estados Unidos) elasticidades próprias altas para marcas de cerveja e papel higiênico acima de -4, a serem comparadas com elasticidades próprias de outros produtos na faixa de -2 a -3, como cereais matinais.

## 1.2 DEMANDA

### 1.2.1 Especificação geral e considerações iniciais

Iniciamos esta seção especificando o lado de demanda em termos gerais. Na sequência, consideraremos sucintamente o caso mais simples de demanda por um bem homogêneo, partindo então para uma apresentação de modelos empíricos de demanda por produtos diferenciados, tanto modelos contínuos como modelos discretos. Na Subseção 2.5 tratamos da identificação empírica da relação de demanda.

A função de demanda ordinária, ou incondicional, é tipicamente especificada em trabalhos empíricos por:

$$q_i = D(p, Y, \alpha_i, \varepsilon_i) \quad (12)$$

onde  $q_i$  é a demanda pelo bem  $i$ ,  $p$  é um vetor de preços de todos os bens no mercado (bens internos),  $Y$  é um vetor de variáveis exógenas que deslocam a curva

de demanda (ver a seguir),  $\alpha_i$  são parâmetros a serem estimados e  $\varepsilon_i$  é o erro econômico, possivelmente entrando em (12) de forma não-linear. Exemplos típicos de deslocadores da curva de demanda  $Y$  são renda, tamanho da população e efeitos sazonais. No caso de produtos diferenciados, como veremos, o vetor de preços de todos os bens internos  $p$  pode ser escrito como  $(p_i, p_{-i})$ , onde se destaca o preço do bem  $i$  (o bem próprio) dos preços dos bens substitutos ou complementares a  $i$ ,  $\{-i\} := \{j \mid j \neq i\}$  (os bens cruzados). Para cada dado valor (ou combinação de valores)  $Y$  tem-se, para o bem  $i$ , uma curva de demanda no espaço  $p_i$  por  $q_i$ . Alterações no valor de  $Y$ , assim como alterações no preço de algum bem cruzado  $p_{-i}$ , deslocam essa curva de demanda do bem  $i$ .<sup>7</sup>

### 1.2.2 Produtos homogêneos versus diferenciados

Um mercado de produtos homogêneos é aquele onde o consumidor não percebe diferenciação, em qualquer dimensão, pelos produtos ofertados. Por diferenciação entende-se comumente as características técnicas do produto como, por exemplo, sabor do alimento, cor do vestido ou potência do carro. Outra dimensão importante em que pode haver diferenciação é a dimensão geográfica. Assim, uma dimensão em que claramente pode haver diferenciação no mercado de minério de ferro ou de petróleo cru, digamos, é a localização geográfica. Dado o mesmo preço por um barril de petróleo a ser entregue em São Paulo e um barril de petróleo a ser entregue em Londres, o consumidor terá preferência por algum desses dois bens *diferenciados*.

A especificação de um mercado de bens homogêneos poderá ser justificável quando a diferenciação entre produtos em todas as suas dimensões *a)* não for importante perante o consumidor (ou mesmo for inexistente!); ou *b)* puder ser controlada na modelagem (por exemplo, situações onde faz sentido definir um mercado em termos locacionais desagregados, especificando-se uma função de demanda para cada localidade, não ocorrendo *arbitragem* entre mercados locais para a faixa de preços observada).<sup>8</sup> A função de demanda ordinária pelo bem homogêneo é então um caso particular da especificação (12), a saber:

$$q = D(p, Y, \alpha, \varepsilon) \quad (13)$$

onde suprime-se o subscrito  $i$  (e  $p$  é agora um escalar).

7. Para a dedução da curva de demanda ordinária a partir do problema do consumidor, bem como a agregação das demandas individuais para demanda agregada, remetemos o leitor a um livro-texto de microeconomia. No entanto, ao longo do presente trabalho, faremos algumas considerações pertinentes e demonstraremos, a título de ilustração, a dedução de uma função de demanda importante (AIDS) a partir do problema primitivo do consumidor.

8. Salvo (2004), por exemplo, estima uma função de demanda por cimento para cada unidade federativa do Brasil.

*Um exemplo de especificação de demanda por produto homogêneo.* Estaremos analisando posteriormente vários aspectos da estimação estrutural da indústria de açúcar dos Estados Unidos na virada do século XX, por Genesove e Mullin (1998). Convém, portanto, ilustrar aqui a especificação de demanda por produtos homogêneos fazendo referência a este trabalho (relegamos a discussão sobre identificação de demanda à Subseção 2.5). Os autores estimam (13) utilizando quatro formas funcionais alternativas<sup>9</sup> — como veremos na Seção 5, o intuito é verificar a robustez dos resultados à escolha de forma funcional — a saber:

Especificação	Fórmula
Quadrática	$\ln q = \ln \alpha_1 + 2\ln(\alpha_2 - p) + \varepsilon$
Linear	$q = \alpha_1 (\alpha_2 - p) + \varepsilon$
Log-linear	$\ln q = \ln(-\alpha_1) + \alpha_3 \ln p + \varepsilon$
Exponencial	$\ln q = \ln \alpha_1 + \frac{\alpha_3}{\alpha_2} p + \varepsilon$

Quanto aos deslocadores de demanda  $Y$ , Genesove e Mullin (1998) especificam uma variável *dummy ALTA\_ESTACAO* que toma o valor 1 quando a observação corresponde a um período (trimestral) de verão. Os autores argumentam que no verão, em virtude da complementaridade de demanda entre açúcar e fruta, para a confecção de doces, a curva de demanda por açúcar se desloca exogenousamente para fora. Tipicamente em trabalhos dessa natureza,  $Y$  entra na curva de demanda como um termo aditivo (por exemplo,  $+ Y_{it} + \ln Y$ ) e também possivelmente interagindo com preços de alguma forma (por exemplo,  $+ Y.P$  ou  $+ \ln P.\ln Y$ ). Os autores, entretanto, optam por estimar a equação de demanda duas vezes (para cada forma funcional), uma vez para observações referentes à alta estação ( $ALTA\_ESTACAO = 1$ ), e outra vez para observações referentes à baixa estação ( $ALTA\_ESTACAO = 0$ ). ■

Já a demanda por produtos diferenciados é o caso mais geral da especificação (12), pondendo reescrevê-la como:

$$q_i = D(p_i, p_{-i}, Y, \alpha_i, \varepsilon_i) \quad (14)$$

Será discutido posteriormente como definir as fronteiras de um mercado de produtos diferenciados. Por agora, apesar de admitir que tal tarefa não é necessa-

---

9. Todas essas quatro formas funcionais são casos específicos da forma mais geral de curva de demanda  $q = \alpha_1(\alpha_2 - p)^{\alpha_3}$ .

riamente simples,<sup>10</sup> assumimos que o pesquisador tenha delineado de modo satisfatório o mercado sob análise: digamos que há  $J$  bens diferenciados no mercado.

O esforço de pesquisa com o intuito de especificar  $D(\cdot)$  ao mesmo tempo flexível e consistente com a teoria econômica vem desde meados do século passado. Contribuições importantes incluem o LES [Stone (1954) e Pollak e Wales (1992)]; o modelo de Rotterdam [Theil (1965) e Barten (1966)]; o modelo Translog [Christensen, Jorgenson e Lau (1975)]; e AIDS [Deaton e Muellbauer (1980a)].

Pela equação (14) vê-se claramente que um desafio presente na estimativa de demanda por produtos diferenciados é o elevado número de parâmetros a serem estimados.<sup>11</sup> Com  $J$  bens ( $J \geq 100$ , em estudos clássicos tanto no caso de automóveis, como no caso de cereais), o número de elasticidades-preço (próprias e cruzadas) a ser estimado é da ordem de  $J^2$ , de modo que mesmo um sistema de demanda extremamente restritivo consistiria em um grande número de parâmetros a serem estimados<sup>12</sup> (por contraste, há apenas uma elasticidade-preço a ser estimada em um mercado de produtos homogêneos). Como veremos nos modelos a seguir, maneiras de lidar com esse número excessivo de parâmetros incluem (além da opção de agregar produtos): *a*) decompor a decisão de compra do consumidor em estágios, de forma que alguns bens são melhores substitutos de determinada marca do que outros; *b*) utilizar restrições advindas da teoria econômica para restringir padrões de substituição, como restrições de homogeneidade; e *c*) modelar a decisão de consumo de forma que o consumidor escolha produtos indiretamente através das características que esses produtos oferecem, em vez de escolher diretamente os produtos. Como o número de características pode ser substancialmente inferior ao número de produtos (por exemplo na demanda por veículos), a dimensão do problema é reduzida.

### **1.2.3 Produtos diferenciados 1: modelos contínuos**

#### **1.2.3.1 Sistema de escolha em estágios múltiplos**

Uma forma de lidar com o problema de ter de estimar um número excessivo de parâmetros é decompor a decisão de consumo em múltiplos estágios (ou níveis),

10. Por exemplo, é fácil concebermos situações onde não há uma clara “quebra na cadeia de produtos substitutos”, ou há considerável heterogeneidade nas preferências dos consumidores. Ou mesmo situações onde os produtos ofertados apresentam múltiplas características (em dimensões tanto *horizontais* como *verticais*), ou apresentam grandes variações em preço e qualidade.

11. Discutiremos a identificação dessa quantidade elevada de parâmetros na Subseção 1.2.5.

12. Por exemplo, suponha um modelo da forma  $D(p) = B \cdot p$ , onde  $B$  é uma matriz simétrica de coeficientes  $\beta$ . No caso de haver  $J$  produtos, a matriz  $B$  teria um número de parâmetros da ordem  $J^2$ . Obviamente, pode-se impor simetria à Slutsky e restrições *adding-up*, mas o problema de o número de parâmetros crescer na ordem do quadrado do número de produtos persiste, sendo amplificado com o uso de formas funcionais mais flexíveis.

segundo a metodologia de orçamentos em estágios múltiplos de Gorman (1995). Tal estratégia tem sido utilizada amplamente por Hausman e seus co-autores.<sup>13</sup> O sistema de demanda a ser estimado é composto por distintos níveis de agregação. Consideremos, como ilustração, um sistema de três níveis. O nível “superior” corresponde à decisão de consumo entre o produto do mercado (ou indústria) de interesse (por exemplo, o mercado de cerveja) e o produto externo. No nível “intermediário”, dado o montante a ser gasto no mercado de interesse advindo do estágio superior, o consumidor aloca essa despesa entre os diferentes tipos de “segmentos” oferecidos nesse mercado; no exemplo da cerveja, dado o montante a ser gasto nesse produto, o consumidor aloca essa despesa entre os segmentos *premium*, *light* e popular. Já no nível “inferior”, condicionando na despesa a ser alocada em determinado segmento segundo o estágio intermediário, o consumidor aloca essa despesa entre as diversas “marcas” oferecidas naquele segmento; por exemplo, dada a despesa alocada para o segmento de cerveja popular, o consumidor escolhe entre as diversas marcas de cerveja popular. Assim, especifica-se uma equação para o nível superior, uma equação de nível intermediário para cada segmento, e uma equação de nível inferior para cada marca em cada segmento. De certa forma, as marcas que figuram no mesmo segmento fazem parte do mesmo “ninho” e estariam mais “próximas” entre si em termos do padrão de substituibilidade. Isso se reflete na especificação das equações como veremos a seguir.<sup>14</sup>

Como especificar a decisão de consumo em termos do número de estágios, bem como a divisão e a agregação entre segmentos e marcas,<sup>15</sup> é em, última instância, uma questão empírica, em nada diferente do antigo debate na literatura em torno da definição de um mercado (e a tal “quebra na cadeia de substitutos”). Em algumas situações, a divisão parecerá natural, baseada, por exemplo, em entrevistas junto a executivos de *marketing* da indústria. Claramente, a especificação a ser utilizada irá depender também do objetivo do estudo, assim como da estrutura dos dados com a qual o pesquisador procederá à estimativa do sistema de demanda. Quando em dúvida entre especificações alternativas, é recomendável efetuar a estimativa dessas especificações, para investigar a robustez dos resultados à especificação. Dos diferentes níveis de alocação, o nível intermediário (ou níveis intermediários,

13. Para um exemplo de aplicação a uma indústria brasileira (cerveja), ver Cysne *et alii* (2001). Hausman, Leonard e Zona (1994), que estimam a demanda por cerveja; Hausman (1997a), que estimou a demanda por cereal matinal; Hausman e Leonard (1997 e 2002), que estimam a demanda por papel higiênico.

14. Esse caráter seqüencial na modelagem da decisão do consumidor é adotado também no modelo NL, como veremos na Subseção 2.4.

15. Por exemplo, devemos contentar-nos com cerveja Antarctica sendo uma marca, ou deveríamos especificar cerveja Antarctica em lata, diferente de cerveja Antarctica em garrafa *long neck*? Ver a Seção 1.5.

se houver mais de um) pode parecer um tanto *ad hoc* e, portanto, em algumas situações pode ser conveniente suprimi-lo,<sup>16</sup> especificando apenas dois estágios [ver, por exemplo, Hausman e Leonard (2002)]. No que segue, apresentaremos um sistema de apenas dois estágios, comentando que a especificação de um nível intermediário é comumente similar à especificação do nível superior. Tipicamente especifica-se uma equação log-linear para o nível superior, utilizando-se para o nível inferior a função de demanda AIDS, proposta por Deaton e Muellbauer (1980a), em virtude de suas propriedades flexíveis. Tendo estimado os diferentes estágios do sistema de demanda, os parâmetros estimados são então utilizados para calcular as elasticidades-preço próprias e cruzadas das diferentes marcas. Comumente esse sistema é estimado a partir de dados em painel, onde dados referentes às diversas marcas são coletados ao longo do tempo (mercados temporais) em diferentes regiões (mercados geográficos).<sup>17</sup>

*Estágio inferior.* A começar pelo estágio inferior, a equação de demanda AIDS é especificada no formato de *share* de valor. [No Apêndice, deduzimos esta expressão, seguindo Deaton e Muellbauer (1980a)]. Assim, o *share* de valor correspondente à marca  $i$  na região (ou “cidade”)  $n$  e no período  $t$ , denotado por  $s_{int}$ , é dado por:

$$s_{int} = \alpha_{in} + \beta_i \log\left(\frac{Y_{nt}}{P_{nt}}\right) + \sum_j \gamma_{ij} \log p_{jnt} + Z_{nt} \theta_{in} + \varepsilon_{int} \quad (15)$$

onde  $s_{int} := \frac{p_{int} q_{int}}{Y_{nt}}$ ,  $p_{int}$  é o preço da marca  $i$  na região  $n$  e no período  $t$ ,  $q_{int}$  é o

consumo da marca  $i$  na região  $n$  e no período  $t$ , e  $Y_{nt} := \sum_i p_{int} q_{int}$  é a receita total da indústria (mercado) na região  $n$  e no período  $t$ . (Caso seja especificado um estágio intermediário,  $Y_{nt}$  será a receita total do segmento do qual a marca  $i$  faz parte.)  $P_{nt}$  é um índice de preços da indústria (ver o Apêndice), normalmente

16. Por exemplo, quando o número de marcas é suficientemente pequeno, ou algumas marcas de menor interesse podem ser agregadas.

17. Uma outra possibilidade é observar dados referentes a vendas através de diferentes canais, como por exemplo, observar o consumo de cerveja em bares separado de vendas através de supermercados.

aproximado (para tornar a equação linear) por um índice de preços de Stone (1954) *médio*,<sup>18</sup> dado por:

$$\log P_{nt} := \sum_i w_{in} \log p_{int} \quad (16)$$

onde o peso  $w_{in}$  é o *share* de valor médio da marca  $i$  na região  $n$ , cuja média é computada ao longo de todos os períodos  $t$ :

$$w_{in} := \frac{\sum_t p_{int} q_{int}}{\sum_i \sum_t p_{int} q_{int}} = \frac{Y_{in}}{Y_n}$$

Ainda com relação a (15), os interceptos  $\alpha_{in}$  denotam efeitos fixos específicos a marca e região, de forma a capturar diferenças demográficas e de preferências invariantes no tempo (para refletir, por exemplo, o fato de que algumas marcas simplesmente são mais populares em algumas regiões do que em outras).  $Z_{nt}$  são variáveis que capturam alterações demográficas e de preferências, como efeitos sazonais, condições climáticas e tendências temporais (específicas a cada região). Por fim,  $\varepsilon_{int}$  é o erro econométrico.

Convém tecer alguns comentários sobre a função AIDS. A restrição de homogeneidade, advinda da teoria econômica (ou seja, se preços e renda são multiplicados por uma constante, os *shares* devem permanecer inalterados) requer que  $\sum_j \gamma_{ij} = 1$ . O uso dessa restrição no procedimento de estimação deve aumentar sua eficiência. Já o uso de restrições de simetria ( $\gamma_{ij} = \gamma_{ji}$ ), apesar de ser um tanto comum, não é indicado na estimativa com dados de demanda agregada, podendo gerar inconsistência nas estimativas. Ao passo que um sistema de demanda especificado para um indivíduo deve satisfazer, segundo a teoria econômica, a propriedade

18. Vale observar que essa linearização da equação de demanda AIDS (*Linear Approximate AIDS*) não é isenta de crítica, tais como a inconsistência das estimativas (devido ao problema de "erro de variável") [ver, por exemplo, Pashardes (1993), Alston, Foster e Green (1994) e Buse e Chan (2000)]. Observe que (16) é uma versão *média* do índice de preços de Stone  $\sum_i s_{int} \log p_{int}$  [ver (46) no Apêndice] onde, no lugar dos pesos  $s_{int}$  que variam no tempo, utilizam-se pesos  $w_{in}$  invariantes no tempo. Apesar de requerer mais trabalho, o próprio sistema não-linear AIDS, onde o índice de preços é dado pela expressão translog (44) no Apêndice, pode em princípio ser estimado. No entanto, como apontam Capps, Church e Love (2003), a resolução do sistema não-linear AIDS — ou seja, a resolução exata em vez da resolução aproximada — tem de lidar com o problema de falta de convergência, decorrente de sua não-linearidade. De fato, em uma aplicação ao mercado de molhos para massas, os autores não conseguem estimar o sistema não-linear devido a esse motivo [ver também Deaton e Muellbauer (1980a, p. 316) e Green e Alston (1990)]. Outras alternativas incluem estimar versões linearizadas com índices de preços alternativos ao Stone, como o Laspeyres ou o Tornqvist.

de simetria Slutsky (ver o Apêndice), as condições necessárias para assegurar que essa propriedade continue valendo em nível de demanda agregada são demasiadamente restritivas [Deaton e Muellbauer (1980b, cap. 6)].

Um tema recorrente na Seção 5, que versa sobre considerações práticas na estimativa estrutural, trata do *trade-off* entre eficiência e consistência. Esse *trade-off* está potencialmente presente na inclusão por parte do analista de restrições de homogeneidade e de simetria na estimativa. Capps, Church e Love (2003), que estimam sistemas de demanda multiestágio com funções AIDS (e Rotterdam) para o mercado de molhos para massas, analisam tais restrições usando testes de verossimilhança relativa. Vale apontar que na estimativa do sistema com funções AIDS tanto as restrições de simetria como as de homogeneidade são rejeitadas. No entanto, a inclusão das restrições reduz substancialmente a variância das elasticidades estimadas pelos autores (e, aliás, reduz o número de elasticidades cruzadas *negativas* obtidas na estimativa).

Finalmente, vale lembrar que proponentes da especificação de demanda AIDS ressaltam sua flexibilidade quando comparada à de outros sistemas, como a especificação *logit* ou LES.<sup>19</sup> Como vimos, o que é importante na especificação de um sistema de demanda flexível é que este limite o número de parâmetros a serem estimados mas, ao mesmo tempo, não impõe padrões de substituição entre marcas, e sim permita aos dados informarem esses padrões de substituição.

*Estágio superior.* O consumo total da indústria na região  $n$  e no período  $t$ , definido como  $Q_{nt} := \sum_i q_{int}$  é, então, especificado na equação de estágio superior em termos do índice de preços da indústria  $P_{nt}$  e da renda (real) disponível aos consumidores na região  $n$  e no período  $t$ , denotado  $X_{nt}$ . Tal equação pode ser especificada no formato log-linear, conforme segue:

$$\log Q_{nt} = \alpha_n + \beta \log X_{nt} + \gamma \log P_{nt} + Z_{nt} \theta + \varepsilon_{nt} \quad (17)$$

---

19. Como veremos ao longo deste trabalho, na especificação *logit* (assim como em especificações onde a elasticidade de substituição é constante, conhecidas como *Constant Elasticity of Substitution*, ou CES), as elasticidades-preço cruzadas são, por definição, iguais umas às outras, a ponto de apenas uma elasticidade-preço cruzada ter de ser estimada. Na especificação de gastos lineares (LES), as elasticidades-preço cruzadas são restrinvidas a serem proporcionais às participações (*shares*) de mercado. Nessa mesma linha, o sistema PCAIDS aplica uma premissa semelhante de proporcionalidade ao sistema AIDS.

onde  $\alpha_n$  é um efeito fixo específico à região que novamente captura diferenças demográficas e de preferências invariantes no tempo, as variáveis  $Z_n$  são definidas como na equação inferior (15), e  $\varepsilon_{nt}$  corresponde ao erro econométrico.<sup>20</sup>

*Elasticidades-preço de demanda.* As elasticidades-preço próprias (para  $i = j$ ) ou cruzadas (para  $i \neq j$ ) podem ser calculadas a partir das estimativas dos coeficientes  $\gamma$ ,  $\beta_i$  e  $\gamma_{ij}$  das equações superior e inferior, conforme segue:<sup>21</sup>

$$\eta_{ij} = \frac{1}{s_i} (\gamma_{ij} - \beta_i w_j) + \left(1 + \frac{\beta_i}{s_i}\right) (1 + \gamma) w_j - 1[i = j] \quad (18)$$

(A função indicadora  $1[i = j]$  retorna 1 se  $i = j$ , retornando 0 no caso contrário.) Relegamos a dedução e breve discussão desta expressão ao Apêndice. Dois pontos merecem ser comentados. Primeiro, os erros-padrão das elasticidades podem ser calculados a partir dos erros-padrão associados aos coeficientes estimados (usando, por exemplo, o método delta: ver o Apêndice<sup>22</sup>). Segundo, as estatísticas  $t$  normalmente apresentadas ao reportar a significância dos resultados devem levar em conta que a hipótese nula tem de ser adaptada, conforme a equação (18) deixa claro. Em vez de testar a hipótese nula de a estimativa ser igual a zero, ou seja, que  $\eta_{ij} = 0$ , hipóteses nulas mais apropriadas são  $\eta_{ij} = w_j - 1$  para a significância das elasticidades próprias e  $\eta_{ij} = w_j$  para a significância das elasticidades cruzadas.<sup>23</sup> Isso (infelizmente) poucas vezes é feito, não surpreendendo, portanto, estatísticas  $t$  apresentadas de magnitude bastante alta para as elasticidades próprias (onde o último termo de (18) é, por definição, igual a -1).

---

20. Quando o sistema especificado incluir um estágio intermediário, uma equação semelhante à do nível superior (17) será incluída para cada segmento, entre os níveis superior e inferior. A equação referente ao (log do) consumo do segmento / especificará esse consumo em termos (do log) de: a) a receita total (real) da indústria  $Y_n$  (no lugar de renda disponível  $X_n$ ); e b) os índices de preços (Stone, por exemplo) para cada segmento (no lugar do índice de preços da indústria  $P_n$ ).

21. Podemos usar as médias amostrais  $s_i$  e  $w_j$  das variáveis  $s_{int}$  e  $w_{int}$ , respectivamente.

22. Capps, Church e Love (2003) argumentam que o uso do método delta para o cálculo dos erros-padrão dos preços pós-fusão a partir das estimativas das elasticidades não é indicado, pelo fato de a solução dos preços pós-fusão envolver a transformação não-linear de elasticidades no sistema de condições de primeira ordem (lembre a ilustração na introdução). Eles propõem a utilização do procedimento de *bootstrap* de Freedman e Peters (1984) que leva em consideração a autocorrelação dos resíduos. A correção da autocorrelação, argumentam Capps, Church e Love (2003), é importante já que a correlação serial tende a ser comum ao utilizar-se dados de *scanner* (ver a discussão na Seção 5).

23. Para ver isso, lembre-se que  $1[i = j] = 1$  nos casos de elasticidade própria e 0 nos de elasticidade cruzada, e que pela hipótese nula (resultados insignificantes)  $\gamma = \beta_i = \gamma_{ij} = 0$  em ambos os casos.

### 1.2.3.2 A métrica de distância

A métrica de distância, proposta por Pinkse, Slade e Brett (2002), permite ao analista testar a relevância da diferenciação ao longo de diferentes dimensões no espaço de características. Assim, diferentes hipóteses a respeito das dimensões em que os produtos competem entre si, gerando diferentes medidas de distância, podem ser avaliadas. O termo distância é adotado no sentido da proximidade (inversa) entre produtos, ou seja, sua substituibilidade. A métrica de distância, baseada na função de utilidade indireta quadrática-normalizada [ver Berndt, Fuss e Waverman (1977) e McFadden (1978b)] não é um modelo de escolha discreta. Ela assume que o consumidor possui uma preferência sistemática por diversidade, assim consumindo mais do que uma marca, e em quantidades variáveis.<sup>24</sup> Da função de utilidade indireta quadrática decorre a linearidade em preços e renda da função de demanda agregada da marca  $i$ :

$$q_i = a_i + \sum_{j=1}^n b_{ij} p_j - \gamma_i y \quad (19)$$

onde  $B = [b_{ij}]$  é uma matriz simétrica negativa semidefinida de dimensão  $n \times n$ , e onde preços  $p_j$  e renda agregada  $y$  estão normalizados pelo preço do bem exterior  $p_0$ . Para lidar com o problema de dimensionalidade — o elevado número de parâmetros em (19) — especifica-se  $a_i$  e  $b_{ii}$  como funções das características da marca  $i$ ,  $a_i = a(x_i)$  e  $b_{ii} = b(x_i)$ . Pinkse e Slade (2004) e Slade (2004), estudando a indústria de cerveja, tomam como características o teor alcoólico, o tipo de cerveja (por exemplo, Pilsen) e a identidade da cervejaria. Os elementos fora da diagonal em  $B$  são especificados como função de um vetor de medidas da distância entre marcas de acordo com alguma métrica escolhida, ou seja,  $b_{ij} = g(d_{ij})$ . Voltando ao exemplo de cervejas, Pinkse e Slade escolhem como medidas de distância a proximidade no teor alcoólico, variáveis *dummies* que capturam o tipo de cerveja e *dummies* que denotam a cervejaria. Como veremos adiante em modelos de escolha discreta — por exemplo, o modelo NL —, através da transformação da equação de demanda (19) para uma situação onde consumidores demandam as características de produtos e não os produtos em si, a dimensionalidade do problema é reduzida. No entanto, diferentemente dos modelos discretos mais restritos como o *logit*, a

---

24. Nesse modelo, o consumidor sempre adquire o bem exterior. Os preços dos produtos diferenciados e as rendas dos consumidores são normalizados (divididos) pelo preço do bem exterior. A função de utilidade está no formato polar de Gorman, simplificando a agregação de demanda pela marca no mercado. Essa agregação independe da distribuição de heterogeneidade individual não-observada das preferências dos consumidores ou da distribuição de renda. O custo dessa simplicidade, no entanto, reside na imposição de que todos os consumidores detêm a mesma utilidade marginal de renda constante [ver resumo em Slade (2004)].

especificação dos elementos fora da diagonal  $b_{ij}$  como sendo função de medidas de distância  $d_{ij}$  permite que marcas com “menor distância” entre si possam ser substitutas mais próximas.

Podemos definir, ainda,  $\xi_i$  como sendo um componente que captura variáveis não-observadas de marca e de mercado na demanda pela marca  $i$ . Denotando  $Z$  como a matriz de variáveis observadas de marca e de mercado, com fileira típica  $z_i = (x'_i, y)$ , (19) pode ser reescrita usando notação matricial

$$q = Z\beta + Bp + \xi \quad (20)$$

onde  $\beta$  é um vetor de parâmetros a ser estimado e o erro  $\xi$  pode ser heterocedástico e correlacionado espacialmente, apesar da premissa mantida de que  $E(\xi_i | X) = 0$  (premissa identificadora também adotada nos modelos de escolha discreta, expostos a seguir).

Pode-se deduzir a partir de (20) que a elasticidade-preço cruzada  $\eta_{ij}$  depende das características de ambas as marcas,  $i$  e  $j$ . Apesar desse aspecto do modelo parecer tão-somente sensato, tal propriedade flexível não é satisfeita no modelo *logit*, como veremos na Subseção 2.4.<sup>25</sup>

#### **1.2.4 Produtos diferenciados 2: modelos de escolha discreta — demanda no espaço de características**

Uma alternativa aos modelos contínuos consiste na classe de modelos de escolha discreta [McFadden (1974, 1978a e 1984); Cardell (1989); Berry (1994); Berry, Levinsohn e Pakes (1995); e Nevo (2001)]. Modelos de escolha discreta resolvem o problema da dimensionalidade tratando produtos como cestas de características<sup>26</sup> [Gorman (1955) e Lancaster (1966)]. No caso de automóveis, por exemplo, cada modelo será definido como a combinação de características tais como seu tamanho, peso, potência, consumo de combustível, e itens opcionais (ar condicionado, direção hidráulica). Assim, as preferências dos indivíduos são definidas no espaço de características, fazendo da dimensão desse espaço a dimensão relevante para a análise empírica. Cada consumidor vai então escolher a cesta de características que maximiza sua utilidade, sendo que as preferências dos consumidores diferem com relação às características. A demanda agregada, ou de mercado, é obtida como

---

25. Adicionalmente, as medidas de distância podem ser escolhidas tal que o padrão de substituição refletirá agrupamentos de marcas, como no modelo NL.

26. Mais formalmente, o problema de dimensionalidade é resolvido projetando-se os produtos em um espaço de características.

soma das demandas individuais, e depende da distribuição das preferências dos consumidores.

Em termos mais formais, a utilidade que o consumidor  $i$  percebe ao consumir o produto  $j$  é denotada por

$$U_{ij} = U(x_j, p_j, v_i; \theta), \quad j = 0, 1, \dots, J$$

onde:

$x_j$  é um vetor de características de produtos, e tais características são as mesmas para todos os consumidores;

$p_j$  representa o preço do produto  $j$  (um caso especial de característica, tratado separadamente em função de seu papel central, e de sua endogeneidade, como exploraremos adiante);

$v_i$  representa fatores gerando preferências, variando de indivíduo para indivíduo;

$J$  é o número de produtos disponíveis no mercado em questão; e

$\theta$  representa o vetor de parâmetros a ser estimado, medindo o impacto das preferências na utilidade dos agentes.

O bem 0 é o chamado bem exterior (*outside good*), um bem não competindo com os produtos da indústria (apesar de competir pelo dinheiro do consumidor), e que corresponde à escolha do consumidor de não adquirir nenhum dos produtos ofertados no mercado em estudo — por exemplo, em se estudando o mercado de transporte aéreo, o bem exterior poderia ser definido como “não comprar passagem aérea”, equivalendo a fazer de ônibus, trem, automóvel etc. o trajeto que poderia ser feito de avião, ou “não realizar a viagem”, gastando o dinheiro em outros bens, ou simplesmente não gastando o dinheiro. A presença do bem exterior é necessária para que seja possível o estudo da demanda agregada — a não-existência do bem exterior no caso do transporte aéreo faria com que consumidores não deixassem de consumir passagens aéreas caso houvesse um aumento generalizado no preço das mesmas. Ou seja, caso não haja um bem exterior, ter-se-ia um estudo de demanda condicional à compra de um dos bens interiores, o que é indesejável. Portanto, nossa análise a seguir será feita supondo-se a existência do bem exterior que, usualmente, dará ao consumidor utilidade média zero, uma normalização comumente usada, sem perda de generalidade.

O subconjunto de preferências levando à escolha do produto  $j$  é dado por

$$A_j(\theta) = \{v_i : U_{ij} > U_{ik}, \text{ para todo } i \text{ e } k \neq j\} \quad (21)$$

sendo corriqueira a hipótese de não haver igualdade (tal caso é ignorado por ter probabilidade nula).

Nesse ponto vale abordar brevemente um conceito a ser muito utilizado adiante. A equação (21) pode ser interpretada como a escolha do indivíduo pela alternativa disponível que lhe dá a máxima utilidade. Convém ressaltar que tais utilidades são variáveis aleatórias, sendo descritas como a soma de um componente determinístico e um componente estocástico, destinado a capturar uma série de incertezas com que o econometrista se depara. Previsivelmente, a esse arcabouço se dá o nome de maximização da utilidade aleatória — Random Utility Maximization (RUM) — [McFadden (1974)], e é importante frisar desde já que modelos da família RUM supõem que indivíduos têm capacidade de discriminação perfeita, mas o econometrista não, deparando-se com informação incompleta, de forma que tal incerteza deve ser levada em consideração, originando-se o termo estocástico na utilidade individual. Em outras palavras, tanto a utilidade individual como a regra de decisão quanto à escolha do produto são determinísticas, e a incerteza resulta do fato de o econometrista não ter perfeita observação da realidade. Por exemplo, Manski (1977) destaca quatro fontes de incerteza:<sup>27</sup>

- a) atributos não-observados, pois o vetor de características afetando a escolha do indivíduo é conhecido apenas em parte pelo econometrista;<sup>28</sup>
- b) variações não-observadas nas preferências individuais, uma consequência da heterogeneidade individual, que introduz variância nas preferências;
- c) erros de medida, pelo fato de as quantidades de cada característica observada não serem perfeitamente conhecidas; e
- d) má especificação funcional, pelo fato de a função utilidade individual não ser conhecida, mas aproximada, o que envolve a adoção de hipóteses sobre a sua forma (linearidade, aditividade).

---

27. Tais componentes serão inseridos paulatinamente na discussão, mas são listados desde já para uma discussão em um nível mais geral.

28. Adicionalmente, segundo Manski (1977), a escolha individual pode também ser afetada por fatores sobre os quais os indivíduos não estão conscientes (ou seja, o componente subconsciente na escolha individual).

Uma vez determinado o subconjunto de preferências  $A_j(\theta)$ , para todo  $j = 1, \dots, J$ , intuitivamente tem-se que “contar” as probabilidades de escolha para que se possa obter as *market shares* dos produtos de um dado mercado. Definindo  $f(v)$  como a distribuição de preferências na população de interesse, as probabilidades de escolha, que coincidem com as *market shares* geradas pelo modelo (supondo-se a correta especificação de  $f(.)$ ), são dadas por:

$$s_j(x, p; \theta) = \int_{v \in A_j(\theta)} f(v) d(v)$$

onde  $(x, p)$  são vetores de características observadas e preços, respectivamente, dos  $J$  produtos considerados.<sup>29</sup>

A demanda agregada pelo bem  $j$  é então dada por  $M \times s_j(x, p; \theta)$ , onde  $M$  é o tamanho do mercado, tomado, por exemplo, como o número de unidades familiares (*households*) em Berry, Levinsohn e Pakes (1995) e Nevo (2000 e 2001), que estudam, respectivamente, os mercados americanos de automóveis e cereais; alternativamente,  $M$  pode ser estimado [ver Reiss e Wolak (2002) para discussão e problemas envolvidos].

Para se estimar os parâmetros do modelo são necessárias hipóteses sobre a distribuição das variáveis não-observadas. O passo seguinte consiste em estimar o modelo usando dados agregados de mercado [Berry, Levinsohn e Pakes (1995)], ou dados individuais de compra (*scanner data*) [Nevo (2000 e 2001)]. De maneira geral, os dados usualmente disponíveis são  $\{(s_j^o, p_j, x_j)\}_{j=1}^J$ , ou seja, *market shares observadas* (daí o sobrescrito  $o$ ), preços e características dos produtos, possivelmente para vários mercados e/ou vários períodos (*market level data*). Mais recentemente, bases de dados mais ricas, com características detalhadas dos consumidores (microdados), também têm se tornado menos raras, mas elas são ainda a exceção, e não a regra, de forma que daremos ênfase a métodos empregando dados em nível de mercado (ver a Seção 1.5 para detalhes).

De modo intuitivo, o pesquisador, usando os modelos descritos a seguir, busca obter um vetor de parâmetros  $\theta$  que minimize uma distância  $\|s^o - s(x, p; \theta)\|$  entre vetores de *market shares* observados e *market shares* gerados pelo modelo estrutural por ele adotado. O que vai diferenciar os modelos que se seguem é a “riqueza” dos *market shares* gerados por cada modelo ou, alternativamente, os padrões das elasticidades geradas por cada modelo, como ilustraremos adiante.

---

29. Tais conceitos serão abordados em maior detalhe adiante.

Dada essa estrutura geral, inúmeros modelos de escolha discreta foram desenvolvidos nas últimas décadas. Faremos a seguir uma revisão seletiva de tal literatura, dando ênfase a modelos que tenham sido usados com algum sucesso em estudos empíricos.<sup>30</sup>

#### 1.2.4.1 Modelo *logit*

O modelo *logit* pode ser escrito como:

$$u_{ij} = \delta_j + \varepsilon_{ij}$$

$$i = 1, \dots, I_t; j = 1, \dots, J; t = 1, \dots, T$$

onde, por exemplo,  $\delta_j = x'_j \beta + \alpha(y - p_j \cdot 1[j > 0])$  ou  $\delta_j = x'_j \beta - \alpha p_j$ , ignorando-se o efeito da renda  $y$ ,<sup>31</sup> e as preferências dos indivíduos diferem apenas por um termo aditivo.<sup>32</sup> Ou seja, a utilidade que o produto  $j$  proporciona ao indivíduo  $i$  é descrita como a soma da utilidade média do bem  $j$  com um termo idiosincrático  $\varepsilon_{ij}$  — tal índice de utilidade é composto por características observadas (pelo econometrista e pelos consumidores)  $x_j$  (além de  $y$  e  $p_j$ ).

O modelo *logit* pode ser motivado tanto a partir de uma perspectiva predominantemente econométrica, ou seja, um modelo de escolha discreta, quanto de uma perspectiva mais econômica, como um modelo simétrico de consumidor representativo. Apesar de suas limitações do ponto de vista aplicado, esse modelo

30. Ver Anderson, de Palma e Thisse (1992) para uma excelente revisão bibliográfica dos métodos disponíveis até o início da década de 1990, bem como métodos alternativos (e não tão difundidos).

31. No que se refere à especificação da renda na modelagem da demanda, primeiro, note que, no caso mais simples, cada consumidor tem a mesma utilidade marginal da renda  $\alpha$  para cada opção, pois

$$\max_j u_{ij}(x, \varepsilon_i) + \alpha(y - p_j \cdot 1[j > 0]) = \alpha y + \max_j u_{ij}(x, \varepsilon_i) - p_j \cdot 1[j > 0]$$

resultando em

$$s_j = \frac{\exp(\delta_j + \alpha y)}{\sum_{k=0}^J \exp(\delta_k + \alpha y)} = \frac{\exp(\delta_j)}{\sum_{k=0}^J \exp(\delta_k)}$$

Assim, nesse caso o nível da renda não influí nas escolhas do consumidor, o que não é necessariamente realista. Formas clássicas de evitar esse inconveniente são a especificação de  $\alpha$  indexado por  $j$  por alguma razão, bem como a especificação não-linear do termo de renda, tal como em Berry, Levinsohn e Pakes (1995), que adotam  $\ln(y - p)$  e, usando dados do censo americano sobre renda familiar, eliminam a distribuição (empírica) da renda ao se calcular a integral para obter a demanda agregada para cada produto.

32. Como vimos antes, o termo  $1[A]$  consiste na função indicadora, tomando valor 1 caso o evento  $A$  ocorra, e 0, caso contrário.

será descrito em algum detalhe pelo fato de ele e suas variantes terem recebido considerável atenção na aplicação ao antitruste de simulação de fusões [Werden e Froeb (1994)].

Motivação econômica.<sup>33</sup> Uma função utilidade muito utilizada em modelos teóricos pioneiros de diferenciação de produto [Dixit e Stiglitz (1977) e Spence (1976)] é a função utilidade com elasticidade de substituição constante (CES),

$$U(q_1, \dots, q_J) = \left( \sum_{j=1}^J q_j^\rho \right)^{\frac{1}{\rho}}$$

onde  $\rho$  é um parâmetro medindo o grau de substituibilidade entre os produtos ( $\rho < 1$  assegura concavidade, e  $\rho > 0$  garante a existência de quantidades nulas, ou seja, garante que o consumidor não compre necessariamente quantidades estritamente positivas de todos os bens).

A demanda do consumidor representativo obtida a partir dessa função utilidade é dada por:

$$q_k = \frac{p_k^{-1/(1-\rho)}}{\sum_{i=1}^J p_i^{-\rho/(1-\rho)}} \times Y, \quad k = 1, \dots, J$$

onde  $Y$  é a renda do consumidor representativo. Convém notar que o problema da dimensionalidade é resolvido impondo-se simetria entre os diferentes produtos, fazendo com que o problema de estimativa se reduza à estimativa de um único parâmetro ( $\rho$ ), independentemente do número de produtos no mercado. Tal imposição é, no entanto, bastante restritiva, por implicar que

$$\frac{\partial q_i}{\partial p_j} \frac{p_j}{q_i} = \frac{\partial q_k}{\partial p_j} \frac{p_j}{q_k}, \text{ para todo } i, j, k$$

ou seja, as elasticidades-preço cruzadas são necessariamente iguais, independentemente da proximidade entre os bens em um dado espaço de atributos (ou seja, de

---

33. O modelo *logit* é estudado em profundidade em Anderson, de Palma e Thisse (1992), cuja leitura é recomendada.

quão parecidos são os produtos). Em problemas aplicados, a violação de tal implicação fatalmente leva a conclusões equivocadas no que se refere ao objeto de estudo.

Uma alternativa à especificação CES é dada por:

$$U(q_1, \dots, q_J) = \sum_{i=1}^J a_i q_i - b \sum_{i=1}^J q_i \ln(q_i) \quad (22)$$

A fórmula (22) tem duas partes. A primeira sugere que o consumidor representativo consome o produto com maior  $a_j$ . Já a segunda é um termo de entropia,<sup>34</sup> e expressa o gosto pela variedade por parte do consumidor — se  $b \rightarrow 0$ , o consumidor não atribui valor algum à variedade, enquanto, se  $b \rightarrow \infty$ , tem-se consumo de todos os bens.<sup>35</sup> Pode-se mostrar [Anderson, de Palma e Thisse (1992, cap. 3)] que tal função utilidade resulta em demandas (*market shares*) da forma *logit*, gerando padrões de substituição mais ricos do que no caso CES. No entanto, como discutido adiante, as elasticidades no modelo *logit* são funções apenas dos *market shares*, e não das características dos produtos considerados.

De modo intuitivo, a simetria de todos os produtos presentes no segundo termo leva a um resultado análogo ao da simetria no caso CES. De forma resumida, modelos como o CES e o *logit* resolvem o problema da dimensionalidade impondo restrições de simetria, que implicitamente sugerem que a competição entre os produtos não depende de sua similaridade (ou seja, despreza-se a informação de quão próximos dois produtos se situam no espaço de características).

*Motivação econométrica.*<sup>36</sup> O modelo *logit* pode também ser motivado como um caso particular de modelos de variáveis dependentes limitadas — Limited Dependent Variable models (LDV) — em que a variável dependente é discreta (por exemplo, tomando o valor 1 caso ocorra a compra do bem  $j$ , e valor 0 caso contrário). Modelos de regressão lineares são, em geral, inapropriados para tratar de tais situações [ver Wooldridge (2002) e Lee (1996) para tratamentos de livro-texto do assunto], visto que tais modelos são intrinsecamente não-lineares.

34. A entropia mede o grau de desordem de um sistema. Para uma discussão detalhada, ver Anderson, de Palma e Thisse (1992).

35. Em trabalhos empíricos, usualmente adota-se a normalização  $b = 1$  para identificação.

36. Uma referência clássica sobre modelos discretos (e *logit*, em particular) é McFadden (1984). Atualmente, livros-texto de econometria em nível de pós-graduação, como Wooldridge (2002), oferecem tratamentos plenamente satisfatórios do tema. Tais leituras são, também, recomendadas. Na discussão que se segue, supõe-se que o leitor esteja familiarizado com o Método da Máxima Verossimilhança (GMU) e o Método Generalizado dos Momentos (GMM).

Modelos LDV podem ser escritos de forma geral como:

$$\begin{aligned}y_i &= E(y_i | x_i) + \varepsilon_i \\&= g(x'_i \beta) + \varepsilon_i, \quad i = 1, \dots, N\end{aligned}$$

onde  $g(\cdot)$  é uma função não-linear conhecida. No caso da variável dependente tomar apenas dois valores (0 ou 1, para fixar idéias), tem-se  $g(x'_i \beta) = E(y_i | x_i) \equiv P(y_i = 1 | x_i)$ . É fácil notar que usar um modelo linear da forma  $y_i = x'_i \beta + \varepsilon_i$  pode, em geral, não ser adequado, pois por  $x'_i \beta$  ser interpretado como uma probabilidade, deve assumir apenas valores compreendidos entre 0 e 1, fato possível apenas com  $x_i$  sendo limitado e  $\beta$  satisfazendo hipóteses extremamente restritivas.

Além disso, pode-se mostrar que o erro não é normalmente distribuído e heterocedástico: pelo fato de  $y_i$  tomar apenas dois valores, o termo de erro, para um dado valor de  $y_i$  pode também tomar apenas dois valores:

$$\begin{aligned}\varepsilon_i &= 1 - g(x'_i \beta), \text{ se } y_i = 1 \\&= -g(x'_i \beta), \text{ se } y_i = 0\end{aligned}$$

sendo claramente não-normal. Convém notar que  $y_i$  é, na realidade, uma variável aleatória seguindo a distribuição de Bernoulli, com função densidade:

$$\begin{aligned}f(y_i | x_i) &= P(y_i = 1 | x_i)^{y_i} \cdot P(y_i = 0 | x_i)^{1-y_i} \\&= g(x'_i \beta)^{y_i} \cdot [1 - g(x'_i \beta)]^{1-y_i}\end{aligned}$$

como consequência, a sua variância é dada por:

$$\text{Var}(y_i | x_i) = P(y_i = 1 | x_i) \cdot (1 - P(y_i = 1 | x_i)) = \text{Var}(\varepsilon_i | x_i)$$

Supondo-se então  $g(x'_i \beta)$  como sendo a probabilidade de  $y_i = 1$ , tem-se que  $\text{Var}(\varepsilon_i | x_i) = g(x'_i \beta) \cdot (1 - g(x'_i \beta))$ , obtendo-se a heterocedasticidade do erro.

Dados os problemas com o modelo linear, uma alternativa intuitiva para modelar a escolha entre duas variáveis discretas é escolher especificações para  $g(\cdot)$

que tomem valores no intervalo  $(0,1)$  — por exemplo, especificar  $g$  como sendo uma função distribuição  $G$ .<sup>37</sup> Casos clássicos incluem os modelos *probit* e *logit*, com  $G(z) = \Phi(z) = \int_{-\infty}^z \phi(z) dz$ , onde  $\phi$  é a função densidade da distribuição Normal, e  $G(z) = \exp(z)/(1 + \exp(z))$ , a distribuição logística, respectivamente.

Dada uma função  $F(.)$ , pode-se estimar o vetor de parâmetros  $\beta$  usando-se o princípio da Máxima Verossimilhança [Maximum Likelihood (ML)]. O estimador ML para  $\beta$  é obtido como solução do seguinte problema de maximização:

$$\max_{\beta} \sum_{i=1}^N \ln g(y_i, X, \beta)$$

Como  $y_i$  segue a distribuição de Bernoulli, tem-se:

$$\begin{aligned} \ln g(y_i, X, \beta) &= y_i \cdot \ln P(y_i = 1 | x_i) + (1 - y_i) \cdot \ln P(y_i = 0 | x_i) \\ &= y_i \cdot \ln G(x'_i \beta) + (1 - y_i) \cdot \ln [1 - G(x'_i \beta)] \end{aligned}$$

de modo que a função (log) verossimilhança correspondente é dada por:

$$\ell(\beta; y, X) = \sum_{i=1}^N y_i \cdot \ln G(x'_i \beta) + (1 - y_i) \cdot \ln [1 - G(x'_i \beta)]$$

No caso de um modelo *probit*, tem-se:

$$\ell(\beta) = \sum_{i=1}^N y_i \cdot \ln (\Phi(-x'_i \beta)) + (1 - y_i) \cdot \ln (1 - \Phi(-x'_i \beta))$$

enquanto no caso de um modelo *logit*,

$$\ell(\beta) = \sum_{i=1}^N y_i \cdot \ln \frac{\exp(-x'_i \beta)}{1 + \exp(-x'_i \beta)} + (1 - y_i) \cdot \ln \frac{1}{1 + \exp(-x'_i \beta)}$$

---

37. Convém aqui notar a equivalência entre se impor estrutura sobre a média condicional (especificada de acordo com a função logística, por exemplo) e sobre o erro do modelo (distribuição de valores extremos do tipo I).

A obtenção de estimadores *logit* ou *probit* consiste em obter (usando-se, por exemplo, o método de Newton) a solução das condições de primeira ordem — *first-order conditions* (FOC) —, podendo ser implementada sem maiores dificuldades. Atualmente, pacotes econôméticos como o Stata têm essas funções implementadas. Ambas as funções critério são globalmente côncavas nos parâmetros, de forma que a obtenção de um máximo global é assegurada — ponto esse que é em geral atingido muito rapidamente. Vale notar que, enquanto no modelo *logit* as probabilidades envolvem expressões fechadas (semelhantes às frações presentes na função log-verossimilhança), no caso do modelo *probit* as probabilidades não têm expressão fechada, aparecendo na forma de integrais. Em caso de um modelo de escolha discreta com múltiplas respostas, a diferença entre *probit* e *logit* se aprofunda, uma vez que, para o primeiro, necessita-se resolver integrais multidimensionais, o que pode ser feito por métodos de aproximação (tal como a regra de Simpson, aprendida em cursos de Cálculo) ou, com melhores propriedades, métodos simulados [ver McFadden (1989) e Pakes e Pollard (1989)]. Em todo caso, a demanda computacional do modelo *probit* se torna muito maior à medida que se aumenta o número de alternativas envolvidas (tem-se uma integral de dimensão  $J - 1$  em um problema com  $J$  alternativas), o que não ocorre com o *logit*. Decorre daí a prevalência deste último em estudos empíricos.

Para obter a covariância assintótica, recorde que  $G_\beta := \partial G / \partial \beta = \phi(z)$  no caso *probit*, e  $G_\beta = \exp(z) / [1 + \exp(z)]^2$  no caso *logit*. Pode-se mostrar que, sendo  $G(\cdot)$  duas vezes continuamente diferenciável, a função *score* condicional associada à observação  $i$  é dada por:

$$Score_i(\beta) := \frac{G_\beta(x_i \beta) \cdot x'_i \cdot [y_i - G(x_i \beta)]}{G(x_i \beta) [1 - G(x_i \beta)]}$$

enquanto o valor esperado do Hessiano condicionado em  $x_i$  é dado por:

$$-E[H_i(\beta) | x_i] = \frac{[G_\beta(x_i \beta)]^2 \cdot x'_i x_i}{G(x_i \beta) [1 - G(x_i \beta)]}$$

sendo uma matriz positiva semidefinida de dimensão  $K \times K$  para cada  $i$ . Pela teoria-padrão de verossimilhança, a covariância assintótica de  $\hat{\beta}$  é estimada por:

$$\widehat{\text{Var}}(\hat{\beta}) = \left( \sum_{i=1}^N \frac{\left[ G_{\beta}(x_i \hat{\beta}) \right]^2 \cdot x_i' x_i}{G(x_i \hat{\beta}) \cdot [1 - G(x_i \hat{\beta})]} \right)^{-1}$$

Convém ainda ressaltar que não é possível identificar tanto os componentes de  $\beta$  quanto  $\sigma$  simultaneamente, de forma que é usual se adotar a normalização  $\sigma = 1$ . Além disso, para que se possam comparar estimativas de modelos *logit* e *probit* (com a normalização acima), deve-se multiplicar o vetor  $\hat{\beta}$  do primeiro por  $\sqrt{3}/\pi$ , pelo fato de sua variância ser  $\pi^2/3$ . Em virtude de as funções distribuição logística e Normal não diferirem substancialmente (diferem marginalmente nas caudas), as estimativas obtidas por ambos os modelos tendem a ser próximas — exceto para dados muito concentrados nos extremos.

Uma forma de comparar os parâmetros e as implicações dos modelos é considerar os efeitos marginais (*marginal effects*), isto é, as derivadas das probabilidades com relação a um dado regressor (no caso de regressores contínuos — para regressores discretos, basta tomar a diferença). No caso de regressores contínuos, as fórmulas são dadas por:

$$\begin{aligned} \frac{\partial P(y_i = 1 | X)}{\partial x_{ik}} &= \phi(x_i' \beta) \cdot \beta_k, \text{ para } \textit{probit} \\ &= \frac{\exp(x_i' \beta)}{\left[ 1 + \exp(x_i' \beta) \right]^2} \cdot \beta_k, \text{ para } \textit{logit} \end{aligned}$$

Normalmente, os efeitos marginais são computados nos valores médios amostrais dos regressores<sup>38</sup> (como no pacote Stata, por exemplo), e o método delta pode ser usado para obter fórmulas analíticas para os erros-padrão das estimativas. Alternativamente, podem-se comparar as elasticidades geradas pelos modelos, opção geralmente preferida em função de sua interpretabilidade econômica.

---

38. Uma alternativa consiste em calcular os efeitos marginais ponto a ponto e tirar a média.

#### 1.2.4.2 Modelos de resposta múltipla: o modelo *logit* multinomial

Como mencionado anteriormente, há aplicações em que o número de alternativas (produtos) entre as quais o indivíduo pode escolher é claramente maior do que dois, o que motiva o uso de modelos discretos de resposta múltipla, uma generalização imediata dos modelos dicotômicos citados. O *logit* multinomial [Multinomial Logit (MNL)] é o mais utilizado entre os modelos de escolha discreta, principalmente em decorrência de sua simplicidade, além do fato de também poder ser considerado dentro do arcabouço RUM.

Considere um espaço de escolhas com  $J$  alternativas mutuamente excludentes e contemplando todas as possibilidades de escolha.<sup>39</sup> A alternativa  $j = 1, \dots, J$  gera utilidade  $y_{ij}^* = x'_{ij} \beta + \varepsilon_{ij}$ . De acordo com a hipótese RUM, um indivíduo  $i$  vai escolher a  $k$ -ésima alternativa tal que  $k = \arg \max_{1 \leq j \leq J} y_{ij}^*$ , isto é, a alternativa que lhe fornece a maior utilidade. Note que:

$$\begin{aligned} P(\text{indivíduo } i \text{ escolhe produto } j) &= P(y_i^* = y_{ik}^*) \\ &= P(y_{ik}^* > y_{ij}^*, \quad j \neq k) \\ &= P(x'_{ik} + \varepsilon_{ik} > x'_{ij} + \varepsilon_{ij}, \quad j \neq k) \end{aligned}$$

Convém notar que a probabilidade de igualdade entre as alternativas é nula para distribuições contínuas. Além disso, vale ressaltar que um problema com  $J$  alternativas e utilidade em níveis pode ser reescrito como um problema com  $J - 1$  alternativas e utilidades tomadas como diferenças entre as alternativas — tal fato decorre do caráter ordinal das preferências.

Conforme já mencionado, os modelos de escolha discreta mais difundidos são o *probit* e o *logit* multinomial [Multinomial Probit (MNP)] e MNL, respectivamente. Em ambos os casos, pode-se escrever a utilidade indireta condicional do indivíduo como:

$$u_{ij} = \delta_j(y, p) + \varepsilon_{ij} \quad (23)$$

$$i = 1, \dots, I; j = 1, \dots, J$$

39. Convém ressaltar que nem sempre é uma tarefa trivial definir o espaço de alternativas e as alternativas mutuamente excludentes que contemplam todas as possibilidades de escolha, como será abordado adiante.

onde, usualmente,  $\delta_j = x'_j \beta + \alpha(y - p_j \cdot 1[j > 0])$  ou  $\delta_j = x'_j \beta - \alpha p_j$ , ignorando-se o efeito-renda, e o “tipo do consumidor” é dado por  $\varepsilon_i = (\varepsilon_{i0}, \varepsilon_{i1}, \dots, \varepsilon_{ij})'$ .

No caso MNP, supõe-se que  $\varepsilon_i = (\varepsilon_{i0}, \varepsilon_{i1}, \dots, \varepsilon_{ij})'$  tem distribuição Normal, enquanto o MNL supõe que  $\varepsilon_i = (\varepsilon_{i0}, \varepsilon_{i1}, \dots, \varepsilon_{ij})'$  é i.i.d. em  $i$  e  $j$  com distribuição de valor extremo do tipo I (também chamada Gumbel), resultando numa função distribuição sendo dada por  $F(\varepsilon) = \exp(-\exp(-\varepsilon))$  e densidade

$$f(\varepsilon) = \frac{1}{\phi_1} \exp\left[-\exp\left(-\frac{(\varepsilon - \phi_2)}{\phi_1}\right)\right] - \frac{(\varepsilon - \phi_2)}{\phi_1}. \text{ No caso de uma distribuição de valores extremos “padrão”,}^{40} \text{ tem-se } \phi_1 = 1 \text{ e } \phi_2 = 0, \text{ resultando em } f(\varepsilon) = \exp[-\exp(-\varepsilon) - \varepsilon].^{41}$$

O MNL goza da propriedade de suas demandas agregadas (equivalentemente, suas probabilidades e *market shares*) poderem ser escritas em forma fechada (análítica) [ver, por exemplo, Anderson, de Palma e Thisse (1992)]:

$$q_j(p, y, x) = M \cdot \int_{-\infty}^{+\infty} q_{ij}(p, y, x, \varepsilon) \cdot f_\varepsilon(\varepsilon) d\varepsilon = M \cdot \frac{\exp(\delta_j)}{\sum_{k=0}^J \exp(\delta_k)}$$

onde  $M$  é o tamanho do mercado potencial, que relaciona as quantidades agregadas demandadas como  $q_j(p, y, x) = M \cdot s_j(p, y, x)$ , resultando em *market shares* da forma

$$s_j := P(y = j) = \frac{\exp(\delta_j)}{\sum_{k=0}^J \exp(\delta_k)}$$

ou seja, o MNL expressa a probabilidade de uma alternativa (no caso, a alternativa  $j$ ) ser escolhida como a razão entre o expoente do componente determinístico de sua utilidade e a soma dos expoentes dos componentes determinísticos das utilidades de todas as alternativas (tanto as escolhidas como as não-escolhidas). É imediato

40. Tal distribuição é também conhecida como exponencial dupla, de Gumbel e de Gnedenko, entre outras denominações.

41. Tal como mencionado na nota 37, há uma relação de equivalência entre se impor uma forma funcional logística para a esperança condicional  $E(y|x)$  de um modelo de regressão e a imposição da distribuição de valores extremos do tipo I para o erro (impondo-se uma normalização conveniente). Tal equivalência é demonstrada nos Lemas 1 e 2 de McFadden (1974). Para resultados relacionados, ver Anderson, de Palma e Thisse (1992, subseção 2.6.1).

notar que isso resulta em probabilidades estimadas (ou *market shares*) assumindo valores entre 0 e 1.

Convém agora mencionar algumas das hipóteses implícitas nessa formulação. A primeira hipótese a se notar na especificação MNL em (23) é a linearidade dos parâmetros, que torna o modelo simples e computacionalmente atrativo. A segunda é a chamada propriedade IIA, de acordo com a qual, para um dado indivíduo, a razão entre as probabilidades de escolha de quaisquer duas alternativas não é afetada pelo componente determinístico de qualquer uma das alternativas que não as duas alternativas envolvidas

$$\frac{s_j}{s_k} = \frac{P(y=j)}{P(y=k)} = \frac{\frac{\exp(\delta_j)}{\sum_{i=0}^J \exp(\delta_i)}}{\frac{\exp(\delta_k)}{\sum_{i=0}^J \exp(\delta_i)}} = \frac{\exp(\delta_j)}{\exp(\delta_k)} \quad (24)$$

A relação (24) acima mostra que as alternativas “irrelevantes” nesse contexto ( $j \neq k$ ) não influenciam as alternativas sobre as quais se tem interesse.<sup>42</sup> Tal propriedade decorre da independência (entre indivíduos) do termo estocástico da utilidade. Ou, de outro modo, que atributos não-observados das alternativas são independentes, o que não necessariamente é uma hipótese realista (por exemplo, poderíamos esperar que um consumidor com elevada preferência por Mercedes tenha elevada preferência por BMW). Finalmente, a relação entre *shares* e demanda agregada remete o leitor à discussão dos modelos de consumidor representativo (a ser retomada ao se tratar os modelos *logit* com coeficientes aleatórios).

#### 1.2.4.3 MNL com características não-observáveis

Uma das possíveis fontes de erro ao se estimar modelos baseados em características é a existência de características não-observadas, não-mensuradas, ou simplesmente omitidas (pelo econometrista). Tipicamente, pode-se pensar em características observadas pelos consumidores (e produtores), mas não pelo econometrista, por exemplo, *status*, qualidade, durabilidade, atendimento, serviço etc. Tal fato foi levado em conta em Berry (1994) que propõe a especificação

$$u_{ijt} = \delta_{jt} + \varepsilon_{ijt}$$

$$i = 1, \dots, N; \quad j = 1, \dots, J; \quad t = 1, \dots, T$$

---

42. Tal propriedade será estudada em maior detalhe adiante.

onde, agora,  $\delta_{jt} = x'_{jt} \beta - \alpha p_{jt} + \xi_{jt}$ . Ou seja, a utilidade que o produto  $j$  proporciona ao indivíduo  $i$  no período (e/ou mercado geográfico)  $t$  é descrita como a soma da qualidade média do bem  $j$  com um termo idiosincrático  $\varepsilon_{ijt}$  — tal índice de qualidade é composto por características observadas (pelo econometrista e pelos consumidores)  $x_{jt}$  e não-observadas (pelo econometrista, mas observadas pelos consumidores<sup>43)</sup>)  $\xi_{jt}$ . É importante salientar que tal modelo difere do *logit* clássico pela presença do termo adicional  $\xi_{jt}$  representando as características não-observadas dos produtos; no entanto, a hipótese de erro  $\varepsilon_{ijt}$  aditivo e i.i.d. (independente e identicamente distribuído) com distribuição de valor extremo do tipo I é mantida.

Em vez de tentar estimar  $\{\xi_j\}$ , suponha que tais elementos sejam variáveis aleatórias geradas a partir de uma distribuição de probabilidade, e use as propriedades de tal distribuição para estimar  $\beta$ . Tipicamente,  $\xi_j$  são supostos ter média condicional zero, ou seja,  $E[\xi_j | x_j] = 0$ <sup>44</sup> e variância finita. Convém notar que a independência na média não vale para preço pois, pelo fato de os consumidores conhecerem  $\xi$ , é provável que as firmas também o conheçam, e é de seu interesse que ele seja considerado ao se fixar o preço de um produto — ou seja, tem-se simultaneidade entre  $\xi$  e  $p$ .

A estratégia de estimação consiste em, supondo-se que o número de consumidores,  $N$ , seja grande, e definido-se  $\theta := (\alpha, \beta)'$ , considerar a relação entre *market shares* observados  $s_j^{obs}$  e *market shares* gerados pelo modelo,  $s_j(\xi, \dots; \theta)$  através do sistema

$$s_j^{obs} = s_j(\xi, \dots; \theta_0)$$

O primeiro fato a se notar aqui é que se tem um sistema de  $J$  equações ( $s_1, \dots, s_J$ ) e  $J$  incógnitas ( $\xi_1, \dots, \xi_J$ ), de forma que, para cada  $\theta$ , existe apenas um  $\xi$  para o qual vale a igualdade. Intuitivamente, para obter  $\xi$  basta apenas que se *inverta* o sistema  $s_j^{obs} = s_j(\xi, \dots; \theta_0)$ , escrevendo  $\xi$  como função dos demais elementos.<sup>45</sup>

43. Características não-observadas pelo econometrista, apesar de observadas pelos consumidores e firmas incluem, entre outros, variáveis difíceis de quantificar, tais como reputação (de uma firma ou modelo), confiabilidade, prestígio, serviço de assistência técnica etc. Como tais fatores são observados pelos participantes, eles serão correlacionados com os preços de equilíbrio, fazendo com que a estimativa dos coeficientes associados a preço, na ausência de tratamento adequado (ver a Subseção 1.2.5), seja viesada em direção a 0.

44. A hipótese de média condicional 0 é feita sem perda de generalidade, em virtude da presença de um termo de intercepto, gerando uma condição de ortogonalidade a ser utilizada no processo de estimação.

45. Tal resultado técnico encontra-se fora do escopo do trabalho. Para fixar idéias, a função  $s(\cdot)$  deve ser suficientemente “suave”. Ou seja, deve ser diferenciável em quase toda parte com relação a  $\delta$ , e as derivadas devem satisfazer  $\frac{\partial s_j}{\partial \delta_j} > 0, \frac{\partial s_j}{\partial \delta_k} < 0, k \neq j$ . Para sua derivação, bem como discussão dos detalhes técnicos, ver Berry (1994) e Berry, Levinsohn e Pakes (1995).

Uma vez obtido  $\xi = \xi(\theta)$ , basta agora aplicar técnicas-padrão de estimação (GMM, por exemplo), ou seja, impor restrições de covariância nula entre  $\xi(\theta_0)$  e outras variáveis observáveis, obtendo-se o valor de  $\theta$  que torna o análogo amostral tão próximo de 0 quanto possível.

O modelo *logit* é um caso particular da inversão muito conveniente para fixar idéias, sendo um dos poucos obtidos com forma fechada para *market shares*. Dada a especificação *logit* para o erro  $\varepsilon$ , tem-se:

$$s_j = \frac{\exp(\delta_j)}{\sum_{k=0}^J \exp(\delta_k)} = \frac{\exp(\delta_j)}{\exp(\delta_0) + \sum_{k=1}^J \exp(\delta_k)} = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^J \exp(\delta_k)}$$

Como  $s_0 = 1 / \left( 1 + \sum_{k=1}^J \exp(\delta_k) \right)$ , tem-se  $s_j = s_0 \cdot \exp(\delta_j)$ , mas  $\delta_j$  agora inclui a característica não-observável (erro)  $\xi_j$ , ou seja,

$$\ln(s_j) - \ln(s_0) = \delta_j(\theta) := x'_j \beta - \alpha p_j + \xi_j \quad (25)$$

Assim, o modelo *logit* com características não-observadas se reduz a simplesmente estimar uma regressão (linear) de variáveis instrumentais da diferença em logaritmos dos *market shares* do bem  $j$  e do bem exterior, tendo características observadas e preço (a ser instrumentado) como regressores. Ou seja, o modelo a se estimar é dado por:

$$\ln(s_j) - \ln(s_0) = x'_j \beta - \alpha p_j + \xi_j$$

A inversão de Berry, nesse caso particular, resulta em:

$$\xi_j = \ln(s_j) - \ln(s_0) - x'_j \beta + \alpha p_j$$

onde, na verdade,  $\xi_j = \xi_j(s, x, p; \theta)$ .

Convém notar que o *logit* é uma exceção, visto que  $\xi_j$  não precisa ser obtido numericamente como a solução de uma integral multidimensional. O passo seguinte, dada a muito provável simultaneidade entre  $\xi$  e  $p$ , é obter condições de momento com o uso de instrumentos para  $p$ . A hipótese tradicional, feita em Berry, Levinsohn e Pakes, e remetendo à literatura clássica de oferta e demanda é  $E[\xi_j | x, w] = 0$ , onde  $x$  consiste em características observadas (exceto preço) de todos os produtos (e não apenas o produto  $j$ ), e  $w$  são deslocadores de oferta, tais como variáveis de custo (variáveis exógenas excluídas da equação de demanda; ver a Subseção 2.5).

#### 1.2.4.4 Uma análise crítica do modelo *logit*

O principal problema do modelo *logit* é o fato de os efeitos substituição entre os produtos serem resultado apenas, e necessariamente, das participações de mercado (*market shares*) dos produtos — alternativamente, das quantidades consumidas pelo consumidor agregado — mas não das *características dos produtos*. Isso ocorre também com a derivada própria do *market share* com relação ao preço, cujo valor depende apenas do *market share* e não das características do produto envolvido. Uma consequência indesejável a ser explorada adiante é que dois produtos com o mesmo *market share* têm de ter o mesmo *markup* em um equilíbrio Nash-Bertrand com firmas produzindo um único produto.

Somada ao problema acima, tem-se a IIA, de acordo com a qual, dado um aumento no preço do produto  $l$ ,  $p_l$  o consumidor representativo vai manter constante a razão entre as quantidades  $q_j/q_k$ , para todo  $j$  e para todo  $k$  diferente de  $l$ . Ou seja, dado um aumento do preço de um bem, espera-se que consumidores consumam maiores quantidades dos produtos mais parecidos com o produto  $l$ , mas isso não ocorre. Usando-se os resultados da seção anterior,

$$\frac{s_j(\theta)}{s_k(\theta)} = \frac{P(y=j)}{P(y=k)} = \frac{\frac{\exp(\delta_j)}{\sum_{h=0}^J \exp(\delta_h)}}{\frac{\exp(\delta_k)}{\sum_{h=0}^J \exp(\delta_h)}} = \frac{\exp(\delta_j)}{\exp(\delta_k)}$$

As consequências da IIA no que concerne à estimativa da demanda são indesejáveis em situações normais. Tal fato foi primeiramente ilustrado por McFadden como sendo o *red bus-blue bus problem* (problema do ônibus vermelho-ônibus azul).

*Exemplo.* Suponha que as probabilidades iniciais de escolha do meio de transporte de um indivíduo sejam dadas por

$$P(\text{automóvel}) = P(A) = 70\%$$

$$P(\text{blue bus}) = P(BB) = 20\%$$

$$P(\text{trem}) = P(T) = 10\%$$

Pela propriedade IIA, tem-se  $P(A)/P(BB) = 70/20 = 3.5$ ,  $P(T)/P(BB) = 10/20 = 0.5$ , e  $P(A)/P(T) = 70/10 = 7$ . Suponha agora a introdução de um ônibus vermelho (RB - red bus), que se distingue do ônibus azul unicamente pela sua cor. Para computar as probabilidades finais, devem-se respeitar a propriedade IIA e o fato de a soma das probabilidades dos meios de transporte ser igual a 1. Daí resultam:

$$P(A) = 3,5P(BB)$$

$$P(T) = 0,5P(BB)$$

$$P(RB) = P(BB)$$

bem como  $P(A) + P(T) + P(BB) + P(RB) = 1$ . Fazendo-se as substituições, chega-se a:

$$P(A) = 7/12 = 58,33\%$$

$$P(T) = 1/12 = 8,33\%$$

$$P(RB) = P(BB) = 1/6 = 16,67\%$$

A estrutura imposta pelo modelo *logit* claramente leva a uma situação totalmente contra o que se esperava, pois calculava-se que o BB capturasse *market share* apenas do RB (de quem é substituto perfeito), o que não ocorreu. Na realidade, a introdução do BB gera uma previsão de acordo com a qual o indivíduo tenderá a usar ônibus com maior freqüência (33,3% contra os 20% iniciais) sem que tenha

havido a introdução de um novo meio de transporte, com características novas. Assim, recomenda-se cuidado ao se utilizar modelos MNL (ver a Seção 1.5). ■

Mais geralmente, o preço a pagar pela tratabilidade do modelo *logit* se expressa de duas formas no que se refere a elasticidades-preço: com elasticidades próprias e cruzadas. Com relação às elasticidades cruzadas — quando o preço de um produto aumenta, a adoção do modelo *logit* implica que os consumidores substituirão esse produto pelos concorrentes de forma proporcional às *market shares*, não levando em conta as características dos produtos. Quanto às elasticidades próprias, seu valor é consequência direta da forma funcional adotada. A seguir, derivamos as elasticidades do modelo *logit* para duas das especificações mais usadas, no que se refere a preços.

*Exemplo.* Relembrando que no modelo *logit* tem-se:

$$s_j(\theta) = \frac{\exp(\delta_j)}{\sum_{k=0}^J \exp(\delta_k)}, \quad j = 1, 1, \dots, J$$

com  $\delta_0 = 0$ , a elasticidade-preço do *share* do bem  $j$  com relação ao preço do bem  $k$  é definida como:

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k} \frac{p_k}{s_j}$$

Calcularemos a derivada acima supondo-se primeiramente a utilidade média linear em preços, passando então à especificação logarítmica. Para a especificação linear em preços,  $\delta_j = x'_j \beta - \alpha p_j + \xi_j$ , tem-se:

$$\frac{\partial s_j}{\partial p_k} = \frac{\partial}{\partial p_k} \left[ \frac{\exp(x'_j \beta - \alpha p_j + \xi_j)}{\sum_{i=0}^J \exp(x'_i \beta - \alpha p_i + \xi_i)} \right]$$

daí resulta,

$$\begin{aligned}\frac{\partial s_j}{\partial p_k} &= -\alpha s_j (1 - s_j), \text{ se } j = k \\ &= \alpha s_j s_k, \text{ caso contrário.}\end{aligned}$$

No caso da especificação logarítmica,  $\delta_j = x'_j \beta - \alpha \ln(p_j) + \xi_j$ , tem-se:

$$\frac{\partial s_j}{\partial p_k} = \frac{\partial}{\partial p_k} \left[ \frac{\exp(x'_j \beta - \alpha \ln(p_j) + \xi_j)}{\sum_{i=0}^J \exp(x'_i \beta - \alpha \ln(p_i) + \xi_i)} \right]$$

daí resulta,

$$\begin{aligned}\frac{\partial s_j}{\partial p_k} &= -\alpha \frac{s_j}{p_j} (1 - s_j), \text{ se } j = k \\ &= \alpha \frac{s_j s_k}{p_k}, \text{ caso contrário.}\end{aligned}$$

As elasticidades são, então, dadas por:

a) Caso linear

$$\begin{aligned}\eta_{jk} &= -\alpha p_j (1 - s_j), \text{ se } j = k \\ &= \alpha p_k s_k, \text{ caso contrário,}\end{aligned}$$

onde um valor alto para  $p_k$  fatalmente leva a elasticidades altas, *caeteris paribus*.

Conclui-se, então, que as elasticidades são determinadas pelo coeficiente  $\alpha$ , pelos preços, e pelos *market shares*. Quanto menor o preço, menor a elasticidade (em valor absoluto), fato que, associado aos modelos usuais de especificação, resulta em um *markup* maior para produtos com preços menores (o caso de automóveis é um bom contra-exemplo). Isso é razoável apenas caso o custo marginal de um

produto mais barato seja menor — não apenas em valor absoluto, mas também como percentagem do preço — que o de um produto mais caro, o que não é necessariamente verdadeiro, em geral. Além disso, note-se que  $\eta_{jk} = \eta_{lk}$ , para todo  $j$  e  $l$ . Ou seja, a elasticidade-preço cruzada entre tanto  $j$  e  $k$ , como  $l$  e  $k$ , é a mesma: depende apenas de variáveis relativas ao produto  $k$ , independentemente de quão próximos os produtos  $j$ ,  $l$  e  $k$  estejam no espaço de características.

### b) Caso logarítmico

$$\begin{aligned}\eta_{jk} &= -\alpha(1 - s_j), \text{ se } j = k \\ &= \alpha s_k, \text{ caso contrário,}\end{aligned}$$

onde as elasticidades são inteiramente determinadas por  $\alpha$  e pelos *market shares*. No caso de a especificação ser logarítmica, a elasticidade resultante será aproximadamente constante.

Conclui-se então que, além de não dependerem das características dos produtos,<sup>46</sup> as elasticidades são consequência direta da forma funcional utilizada. Ou seja, um econometrista, desejando obter altos valores para elasticidades cruzadas, preferiria estimar um modelo linear nos preços, em vez de um modelo logarítmico, pelo fato de as elasticidades do primeiro sabidamente serem maiores do que as do segundo (supondo-se preços maiores que a unidade). ■

No contexto de automóveis, Berry, Levinsohn e Pakes (1995) discutem o exemplo de um Yugo (modelo importado da Iugoslávia) e um Mercedes, carros com *market shares* similares, mas elasticidades-preço cruzadas claramente distintas com relação a um BMW. Apesar de um aumento de preço do BMW tender a gerar um impacto muito maior sobre um Mercedes do que sobre um Yugo (modelo barato importado da então Iugoslávia), consumidores *logit* teriam padrões de substituição iguais. Ou seja, dois produtos claramente em nichos de mercado diferentes terão as mesmas elasticidades cruzadas com relação ao BMW. Alternativamente, se o preço de um Yugo aumenta, o impacto sobre a demanda por um BMW é o mesmo que o impacto sobre a demanda por um Fiat Uno. Vale ressaltar que, no caso de uma análise antitruste, em que elasticidades cruzadas geram informação sobre o mercado relevante, a adoção desavisada desse tipo de modelo parece temerária, como apontado em McFadden (1981); Berry, Levinsohn e Pakes (1995); Nevo (2000); entre outros.

---

46. Note-se que não há termos em  $x$  nas fórmulas das elasticidades.

Concluímos então que, apesar de sua simplicidade e tratabilidade, por resolver o problema da dimensionalidade impondo restrições de simetria, o modelo *logit* paga um preço extremamente alto no que se refere à matriz de elasticidades. Apesar de tal especificação poder se adequar a alguns mercados de bens diferenciados, é heróico esperar que isso ocorra no caso geral.

#### 1.2.4.5 Modelo *logit* aninhado (NL)

Dados os inconvenientes do uso de modelos *logit*, uma forma alternativa sugerida por Ben-Akiva (1973) e racionalizada como um modelo discreto consistente com RUM por McFadden (1978a) é o NL em que a hipótese de erros i.i.d. seguindo uma distribuição de valor extremo é substituída por uma estrutura de componentes de variância (*variance components structure*). Intuitivamente, a decisão de comprar (ou não) um produto é racionalizada através de uma árvore de decisão (tal como na Subseção 1.2.3.1): tipicamente, a primeira ramificação dessa árvore representa a decisão entre adquirir um bem no mercado em estudo, automóveis, por exemplo, ou adquirir um bem exterior. Tendo decidido pela compra de um automóvel, o passo seguinte consiste em, digamos, escolher o tamanho do automóvel e, a partir daí, a sua marca e, assim, sucessivamente. O consumidor se depara com choques comuns a todos os produtos de um determinado ramo, tendendo a substituir um produto por produtos de um mesmo ramo, o que gera maior correlação entre produtos mais próximos no espaço de características (ou seja, que estejam num mesmo ramo).

*Exemplo.* Consideramos, a seguir, o exemplo de Berry (1994), que insere o NL no arcabouço de um modelo linear de variáveis instrumentais. Para isso, suponha a existência de  $G + 1$  conjuntos mutuamente exclusivos,  $g = 0, 1, \dots, G$ , e denote o número de produtos no grupo  $g$  como  $J_g (\sum_g J_g = J)$ . O bem exterior ( $j = 0$ ) é suposto ser o único elemento de  $g = 0$ . Para o produto  $j \in J_g$ , a utilidade condicional indireta do indivíduo  $i$  é dada por

$$u_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma)\varepsilon_{ij}$$

onde  $\delta_j = x'_j \beta - \alpha p_j + \xi_j$ , e  $\varepsilon_{ij}$  é, tal como antes, independente e identicamente distribuído seguindo a distribuição de valores extremos (ou Gumbel). Para o consumidor  $i$ , a variável  $\zeta$  é comum a todos os produtos no grupo  $g$ , tendo distribuição que depende de  $\sigma$ ,  $0 \leq \sigma < 1$ . Cardell (1997) mostra que a distribuição de  $\zeta$  é a única distribuição satisfazendo a propriedade segundo a qual, se  $\varepsilon$  é distribuída como Gumbel,  $[\zeta + (1 - \sigma)\varepsilon]$  também o será. A importância do parâmetro  $\sigma$

aqui se deve ao fato de, conforme seu valor se aproxima de 1, a correlação dos níveis de utilidade dentro do grupo tende a 1 e, conforme seu valor se aproxima de 0, tal correlação tende a 0.

Definindo-se  $d_{jg}$  como uma variável *dummy* tomando valor 1 quando  $j \in J_g$ , a utilidade condicional indireta do indivíduo  $i$  pode ser reescrita como:

$$u_{ij} = \delta_j + \sum_g [d_{jg} \cdot \zeta_{ig}] + (1 - \sigma) \varepsilon_{ij}$$

podendo então ser interpretada como um modelo de coeficientes aleatórios em que tais coeficientes  $\zeta_{ig}$  operam apenas em variáveis *dummies* grupo-específicas.

Para um produto  $j$  pertencendo ao grupo  $g$ , seu *market share* condicional a pertencer ao grupo é:

$$s_{j/g}(\delta, \sigma) = \frac{\exp(\delta_j / (1 - \sigma))}{D_g}$$

onde  $D_g := \sum_{j \in J_g} \exp(\delta_j / (1 - \sigma))$ . Já o *market share* do grupo  $g$  é dado por:

$$s_g(\delta, \sigma) = \frac{D_g^{(1-\sigma)}}{\sum_g D_g^{(1-\sigma)}}$$

resultando em um *market share*:

$$s_j(\delta, \sigma) = s_{j/g}(\delta, \sigma) \cdot s_g(\delta, \sigma) = \frac{\exp(\delta_j / (1 - \sigma))}{D_g^\sigma \cdot \sum_g D_g^{(1-\sigma)}}$$

Para o bem exterior, tem-se  $s_0(\delta, \sigma) = 1/\sum_g D_g^{(1-\sigma)}$ . Tomando-se os logaritmos das equações de *share* dos bens  $j$  e 0, e subtraindo-se uma equação da outra, chega-se a:

$$\log(s_j) - \log(s_0) = \delta_j / (1 - \sigma) - \sigma \log(D_g) \quad (26)$$

que depende, portanto, do valor  $D_g$ . No entanto, tomado-se o logaritmo do *market share* condicional,  $s_{j/g}$ , resolvendo para  $\log(D_g)$ , e plugando na equação (26), obtém-se:

$$\delta_j = [\log(s_j) - \log(s_0)] - \sigma \log(s_{j/g}) \quad (27)$$

que difere do valor de  $\delta_j$  do modelo *logit* pela presença do novo termo,  $-\sigma \log(s_{j/g})$ . Lembrando que a utilidade média é dada por  $\delta_j = x'_j \beta - \alpha p_j + \xi_j$ , substituindo-se na equação (27) via  $\delta_j$ , e resolvendo-se para as *shares* individuais, tem-se

$$[\log(s_j) - \log(s_0)] = x'_j \beta - \alpha p_j + \sigma \log(s_{j/g}) + \xi_j \quad (28)$$

onde é importante notar a simultaneidade entre a variável dependente e o termo  $\log(s_{j/g})$ , o que exige a instrumentação deste. Mais uma vez, tem-se a estimativa de um modelo potencialmente complexo usando-se técnicas lineares de variáveis instrumentais, em função dos resultados de Berry (1994). Tal formulação pode ser estendida para o caso de múltiplos estágios e diferentes parâmetros de correlação  $\sigma_g$  entre os grupos à custa de algum trabalho algébrico, sendo deixado como exercício.<sup>47</sup> Focamos aqui um modelo mais simples pela facilidade de interpretação. ■

Uma importante aplicação do NL foi feita por Goldberg (1995), que o aplica para o mercado americano de automóveis, trabalho que revisamos brevemente.<sup>48</sup> Usando dados de compras de automóveis novos por parte das unidades familiares (cujo uso para se estimar a demanda esperada pelas firmas é algo inédito para a época), a autora estima probabilidades de compra de carros novos ao nível domiciliar. Sua estratégia de estimativa consiste em três partes. Na primeira, estimam-se

47. Veja-se Verboven (1996a) e Fiúza (2002) para detalhes.

48. Para detalhes mais aprofundados, recomendamos a leitura do original, bem como o capítulo de Reiss e Wolak (2002).

as demandas em nível do domicílio. Na segunda, tais demandas são agregadas para formar curvas de demanda representativas esperadas pelas firmas. Na terceira, as curvas de demanda esperadas são usadas para calcular as condições de primeira ordem das firmas, sob a hipótese de os fabricantes de automóveis serem competidores Bertrand-Nash, o que permite recuperar as margens dos fabricantes para cada modelo de carro.

O modelo econômico subjacente à análise de Goldberg trata consumidores como maximizadores de utilidade (estáticos), e as demandas esperadas pelas firmas são computadas como:

$$q_{jt}^e = \sum_{i=1}^{M_t} P(i \text{ compra produto } j)$$

onde  $M_t$  é o tamanho do mercado potencial no período  $t$ . Ao estimar tal modelo, as probabilidades são obtidas a partir de um modelo de escolha discreta usando dados obtidos no US Bureau of Statistics Consumer Expenditure Survey (CES), uma pesquisa sobre as despesas dos consumidores americanos conduzida trimestralmente com 5 mil domicílios por trimestre. Tal base de dados permite a observação do veículo comprado pela família, bem como o preço da transação, base de dados à qual ela adiciona informações sobre características dos automóveis (potência, consumo de combustível, tamanho, e variáveis *dummies* para opcionais). Uma hipótese crucial no seu estudo é que a amostra obtida via CES é representativa do mercado potencial,  $M$ , o que permite que a demanda esperada possa ser substituída por uma soma ponderada das probabilidades de compra das famílias.

No que se refere ao lado da oferta, Goldberg supõe que fabricantes maximizam lucros esperados (estáticos) escolhendo um preço de atacado,<sup>49</sup> que ela supõe ter uma relação exata com os preços transacionados (por exemplo, na proporção de 3/4 para um carro médio) o que, apesar de facilitar o processo de estimativa, é passível de críticas, já que não fica claro por que fabricantes e intermediários se comportariam dessa forma.<sup>50</sup>

49. Goldberg não observa preços no atacado, apenas preços transacionados. No entanto, os incentivos dos revendedores de carros novos nos Estados Unidos não são modelados. Ver a Seção 1.5 para a diferença entre elasticidades no varejo e no atacado.

50. Um problema adicional do estudo de Goldberg (1995) é o fato de a base de dados ignorar compras de carros que não venham de unidades familiares. Com isso, compras governamentais e de firmas são excluídas *a priori*, podendo gerar viéses em, por exemplo, estimativas de custo, uma vez que tais agentes são obviamente considerados pelas firmas quando da fixação de preços.

Quanto à demanda, Goldberg começa por partitionar o conjunto de veículos em subconjuntos distintos de acordo com a idade do modelo, o segmento de mercado, o país de origem, e a marca do veículo. Assim, a árvore de decisão das unidades familiares tem os seguintes estágios: no primeiro estágio, escolhe-se entre comprar ao menos um carro e consumir o bem exterior (não comprar nenhum carro); no segundo estágio, decide-se entre comprar ao menos um carro novo e comprar apenas carro(s) usado(s); a seguir, decide-se entre modelos de nove segmentos distintos e, para cada um desses segmentos, se o carro a ser comprado é produzido nos Estados Unidos, ou no exterior. No quinto e último estágio, decide-se finalmente o modelo a ser comprado. Infelizmente, não é possível contar com o auxílio da teoria econômica no desenho da árvore de decisão, devendo-se basear tal escolha no conhecimento institucional da indústria para realizar tal tarefa. No entanto, é possível comparar mais de uma configuração (ordenação) de estágios consistente com RUM usando o teste de modelos não-aninhados de Vuong (a ser considerado adiante), apesar de não ser de nosso conhecimento que isso tenha sido feito.

É importante notar que, como demonstrado por McFadden (1978a), a estrutura de estágios é consistente com RUM se, e somente se, determinados coeficientes (correspondentes a  $\sigma$ , citados no exemplo) estão contidos no intervalo  $(0,1)$  — caso eles tendam a 1, o modelo se reduz a um MNL e, caso eles sejam maiores do que 1, há substituição entre ramos diferentes — isso implica que a especificação dos estágios considerada não é consistente com a hipótese RUM, e uma configuração alternativa da árvore de decisão deve ser considerada.

Um outro ponto a ser considerado é que a estimação de um modelo com muitos estágios não é factível — há mais de 200 coeficientes a se estimar em Goldberg (1995) —, de forma que a estimação é feita em seqüência, em cada estágio, separadamente. Apesar de se perder eficiência, o método gera ainda estimadores consistentes — no entanto, a estimação seqüencial requer que as estimativas da matriz de covariâncias sejam ajustadas [ver McFadden (1981) para as fórmulas, que são obtidas recursivamente].

O modelo de Goldberg obviamente é mais rico que um NL tradicional (em função do uso de dados amostrais), admitindo padrões de substituição complexos entre produtos — padrões esses que dependem da proximidade dos produtos no espaço de atributos, como esperado. Para isso é necessário introduzir uma série de hipóteses econométricas e funcionais, com o intuito de tornar o modelo mais tratável. Entretanto, é difícil avaliar como cada hipótese econômética impacta suas conclusões.

Apesar de combinar a tratabilidade do modelo *logit* com padrões de substituição mais flexíveis e realistas, a IIA persiste intra-ramos para unidades familiares, mas não persiste intra-ramos para o mercado como um todo (no entanto, não está claro quão flexíveis realmente são o sistema de demanda e as elasticidades cruzadas agregadas). Isso ocorre porque Goldberg interage renda e preços com características de consumidores e produtos [ver Goldberg (1995, p. 898-900) para detalhes e discussão], após o que é feita a agregação usando as ponderações para os dados CES.

Convém ainda ressaltar que a divisão dos produtos em grupos é um tanto arbitrária, e a hipótese de choques i.i.d. intra-ramo, passível de críticas. Além do problema da arbitrariedade da escolha dos elementos constituintes de cada estágio, deve-se atentar para o tamanho dos ramos de um dado estágio. De forma intuitiva, espera-se que algumas das características de produtos em um mesmo mercado estejam correlacionadas — por exemplo, potência, número de cilindros e dimensões do carro, tal como no caso dos automóveis, em Goldberg (1995). Adotando-se um ramo com poucos produtos, eles tenderão a ser mais homogêneos, gerando um alto grau de colinearidade, mas eliminando a necessidade de incluir todas as características dos produtos. Por outro lado, incluindo-se ramos com muitos produtos tende-se a mitigar a colinearidade, mas volta-se ao problema da IIA. Além disso, cabe ressaltar que, apesar de capturar correlações entre elementos de um mesmo ramo, o NL não consegue capturar correlações entre ramos distintos, de forma que, quando alternativas não podem ser separadas em diferentes ramos de modo satisfatório, o NL não é aplicável.

#### 1.2.4.6 Modelo *logit* com coeficientes aleatórios (*mixed logit*)

Nesta subseção abordamos uma generalização do modelo *logit*, inicialmente proposto em Berry, Levinsohn e Pakes (1995) — modelo BLP daqui em diante —, mas também utilizado em Nevo (2000 e 2001); Berry, Levinsohn e Pakes (2004); entre outros [para detalhes mais aprofundados, bem como uma análise crítica detalhada, recomenda-se a leitura dos originais, além de Reiss e Wolak (2002)].

Partindo-se de um modelo microeconômico de comportamento individual dos agentes, que é agregado, obtém-se a demanda de mercado. A microfundamentação do modelo permite que o mesmo arcabouço seja utilizado para dados de mercado (*market/product level data*), microdados, amostras estratificadas etc. Convém destacar que a heterogeneidade individual é modelada de forma a não restringir padrões de substituibilidade *a priori*, fazendo com que as elasticidades entre produtos estejam relacionadas à sua proximidade no espaço de características. Tal fato não apenas gera elasticidades “mais realistas”, como afeta cálculos posteriores feitos para se simular os efeitos de fusões.

Considere agora a utilidade indireta condicional  $u$  obtida pelo indivíduo  $i$  ao consumir o bem  $j$  como sendo dada por:

$$u_{ij} = \sum_{k=1}^K x_{jk} \beta_{ik} + \xi_j + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, J$$

com

$$\beta_{ik} = \lambda_k + \beta_k^{o'} z_i + \beta_k^{u'} v_i$$

onde:

$x_{jk}$  e  $\xi_j$  são, respectivamente, características observadas e não-observadas dos produtos;

$z_i$  e  $v_i$  são, respectivamente, vetores de atributos observados e não-observados dos consumidores;

$\lambda_k$  reflete o impacto da característica  $k$  sobre a utilidade média;

os vetores  $\beta_k^o$  e  $\beta_k^u$  medem, respectivamente, o impacto de características observadas ( $o$ ) e não-observadas ( $u$ ) do consumidor na característica  $k$ ; e

$\varepsilon_{ij}$  representa um componente idiosincrático de preferências individuais pelos diferentes produtos, sendo essas independentes tanto das características dos consumidores, como das características dos produtos sendo modeladas. A hipótese de o erro seguir uma distribuição de valores extremos é mantida, principalmente pela conveniência de se poder escrever as *market shares* seguindo a fórmula clássica para modelos *logit*.<sup>51</sup>

A especificação do sistema de demanda fica completa com a introdução de um bem exterior, tal como nos modelos anteriores.

Pode-se ver que o modelo acima difere dos anteriores por ter coeficientes aleatórios, ou seja, variam de indivíduo para indivíduo em virtude de características observadas e não-observadas destes. Intuitivamente, parece razoável que indivíduos diferentes tenham sensibilidades diferentes às características de um bem. Por exemplo, indivíduos mais ricos seriam menos sensíveis a preços, chefes de famílias numerosas

51. Tal hipótese, que facilita a estimação e fornece derivadas suaves, não é imune a críticas: Petrin (2002) mostra que mudanças de bem-estar decorrentes da introdução de novos produtos é superestimada pela hipótese a respeito do termo de erro. Trabalhos recentes, na linha de Berry e Pakes (2002) e Bajari e Benkard (2004), removem o erro *logit* inteiramente mas, apesar de promissores, não estão ainda estabelecidos na literatura.

prezariam mais o espaço de um automóvel do que uma pessoa solteira, e assim por diante, e os coeficientes aleatórios permitem racionalizar tal fato. Mantendo-se a hipótese de erros i.i.d. seguindo a distribuição de valores extremos, tem-se que as elasticidades próprias não serão mais necessariamente resultantes da forma funcional adotada, pois cada indivíduo terá um coeficiente-preço distinto, que será agregado, gerando uma sensibilidade-preço média, com pesos dados pelas probabilidades de compra dos indivíduos — em outras palavras, tem-se aí o abandono da hipótese de simetria e/ou consumidor representativo dos modelos *logit*. Como consequência de tal fato, haverá sensibilidades-preço diferentes para diferentes bens, e padrões de substituição mais flexíveis, agora, sim, resultantes das características dos produtos em questão. Além disso, consumidores que deixam de consumir um bem em virtude do aumento de seu preço tenderão a trocá-lo por um produto próximo dele no espaço de características, ao contrário do que ocorria com consumidores *logit*, cujo padrão de substituibilidade dependia apenas dos *market shares* (e preços). McFadden e Train (2000) demonstram que tal modelo aproxima tão próximo quanto se queira qualquer modelo de escolha.

Plugando-se a segunda equação na primeira, obtém-se:

$$u_{ij} = \delta_j + \sum_{k=1}^K \sum_{r=1}^R x_{jk} z_{ir} \beta_{kr}^o + \sum_{k=1}^K \sum_{l=1}^L x_{jk} v_{il} \beta_{kl}^u + \varepsilon_{ij} \quad (29)$$

onde  $\delta_j = \sum_{k=1}^K x_{jk} \lambda_k + \xi_j$  é a utilidade média.

A equação (29) permite identificar dois tipos de interações entre características de produtos e características de consumidores. Primeiro, há interações entre características observadas dos consumidores ( $z$ ) e características observadas dos produtos ( $x$ ). Além disso, há ainda interações entre características não-observadas dos consumidores ( $v$ ) e características observadas dos produtos. Tais interações são determinantes para gerar elasticidades-preço próprias e cruzadas que façam sentido econômico, eliminando a propriedade IIA. De modo intuitivo, o aumento do preço de um modelo de automóvel fará com que consumidores com determinadas características optem por outros modelos — modelos esses que devem estar próximos do modelo anterior no espaço de características. Adicionalmente, um aumento de preços vai ter impacto diferente sobre consumidores diferentes — consumidores de carros caros tendem a ser menos sensíveis a mudanças de preços do que consumidores de modelos populares, com a consequência de elasticidades próprias diferentes entre essas classes distintas — em termos de equilíbrio Nash-

Bertrand, por exemplo, ter-se-iam *markups* mais altos justamente para os modelos mais caros, para os quais os respectivos consumidores são menos sensíveis a preço.

*Estimação usando market-level data.* Infelizmente, nem sempre é possível contar com microdados — dados relacionando consumidores e produtos por eles comprados; ver a Seção 1.5 de forma que a melhor informação que se pode obter é sobre a distribuição das variáveis  $z$  (pelo censo, por exemplo). Mesmo no caso de microdados, é necessário certificar-se da riqueza das informações fornecidas, isto é, se os atributos observados dos consumidores são suficientemente ricos a ponto de capturar todas as fontes de heterogeneidade nas preferências por características. Caso isso não ocorra, é necessário incluir os não-observáveis. Berry, Levinsohn e Pakes (1995) consideram apenas a existência de dados de mercado (quantidades vendidas de automóveis novos,<sup>52</sup> tabelas de preços e características dos veículos), partindo de

$$u_{ij} = \delta_j + \sum_{k=1}^K \sum_{l=1}^L x_{jk} v_{il} \beta_{kl}^u + \varepsilon_{ij}$$

onde  $\delta_j = \sum_{k=1}^K x_{jk} \lambda_k + \xi_j$ . A título de comparação com a literatura anterior, vale notar que, fixando-se um único  $v_p$ , tem-se um modelo *logit* clássico. Comparando-se tal especificação com a do modelo *logit*, a introdução de  $v_i$ 's diversos resulta na quebra da hipótese de consumidor representativo — ou na introdução de heterogeneidade entre consumidores, que reagirão a mudanças de preços relativos de acordo com o seu tipo. Intuitivamente, espera-se que tal heterogeneidade nas respostas gere padrões de substituição mais realistas entre os produtos.

No caso específico de Berry, Levinsohn e Pakes (1995), a especificação adotada foi:

$$u_{ij} = \delta_j + \sum_{k=1}^K \sum_{l=1}^L x_{jk} v_{il} \beta_{kl}^u + \alpha \ln(y_i - p_j) + \varepsilon_{ij}$$

A especificação logarítmica da renda é adotada para que a renda de fato influa nas escolhas do consumidor<sup>53</sup> (ver a nota 31 para maiores detalhes). O bem

52. Ao contrário de Goldberg (1995), a decisão de comprar um carro usado em vez de um novo não é explicitamente considerada; assim, carros usados fazem parte da infinidade de produtos agregados como bem exterior.

53. Enquanto em BLP (usando dados americanos) se trunca a distribuição da renda quando  $y < p$ . Para um comentário sobre a aplicabilidade da especificação funcional de renda e preço numa situação de extrema desigualdade de renda, ver Fiúza (2002).

exterior tem utilidade condicional indireta seguindo  $u_{ij} = \alpha \ln(y_i) + \sigma_0 v_{i0} + \varepsilon_{i0}$ , valendo notar que o seu preço  $p_0$  é suposto ser 0, e  $\sigma_0$  representa o desvio-padrão do componente não-observado das preferências dos consumidores,  $v_{i0}$ . Nevo (2000 e 2001) introduz o uso de atributos observados dos consumidores — mais precisamente, ele extrai realizações de características demográficas de dados de censo, o que pode ser interpretado como o uso da função distribuição empírica dessas características, tendo uma interpretação não-paramétrica.

A motivação para o modelo de BLP pode ser obtida voltando-se a Berry (1994), onde se considerava um sistema  $s^{obs} = s(\xi, \dots; \theta)$ . De modo intuitivo, o problema de BLP pode ser entendido como a minimização de uma distância  $\|s^{obs} - s(\xi, \dots; \theta)\|$  entre *shares* observadas e *shares* geradas pelo modelo (alternativamente,  $\|q^{obs} - M \cdot s(\xi, \dots; \theta)\|$  onde  $M$  é o tamanho do mercado). No entanto, ao contrário de um modelo de consumidor representativo, o que se tem agora é um modelo mais complexo, que se espera que possa refletir padrões de substituição mais realistas entre os produtos do que, por exemplo, um modelo *logit* clássico.

O lado da oferta do modelo é bastante simples. Fabricantes têm conhecimento das funções demanda, dos custos marginais uns dos outros, não discriminam preços, e maximizam funções lucroestáticas escolhendo preços, agindo como competidores Bertrand-Nash. Ou seja, o modelo é essencialmente estático, não se modelando a escolha de atributos por parte das firmas. Analogamente, além de se supor que consumidores conheçam preços e atributos de todos os carros novos, também se supõe que eles não trocam preços e características presentes por futuras (por se tratar de um bem durável — ver a Seção 1.5).

No que se refere à implementação do modelo, suponha a existência apenas de dados agregados de mercado, para fixar idéias [ou seja,  $\beta_k^u = 0$ , tal como em Berry, Levinsohn e Pakes (1995)]. O primeiro passo consiste em obter os *market shares* de mercado condicionais a  $(\theta, \delta)$ ,

$$s_j(\theta, \delta) = \frac{\exp\left(\delta_j + \sum_{k=1}^K \sum_{l=1}^L x_{jk} v_{il} \beta_{kl}^u\right)}{1 + \sum_{h=1}^J \exp\left(\delta_h + \sum_{k=1}^K \sum_{l=1}^L x_{hk} v_{il} \beta_{kl}^u\right)} f(v) d(v) \quad (30)$$

onde a fórmula (30) pode ser entendida como uma soma ponderada (com pesos dados pela densidade  $f(\cdot)$  das preferências dos consumidores) de consumidores *logit* (por isso o termo *mixed logit* também é utilizado para denotar o modelo). No

caso em que há apenas um tipo de consumidor, a fórmula se reduz à fórmula de *market shares* do modelo *logit*, em função da hipótese feita sobre o erro  $\epsilon$ .

A complexidade da integral sugere o uso de métodos simulados para agregar os *market shares* individuais. Usam-se então  $ns$  simulações para agregar em  $v$

$$s_j^{ns}(\theta, \delta) = \frac{1}{ns} \sum_{r=1}^{ns} \frac{\exp\left(\delta_j + \sum_{k=1}^K \sum_{l=1}^L x_{jk} v_{il} \beta_{kl}^u\right)}{1 + \sum_{h=1}^J \exp\left(\delta_h + \sum_{k=1}^K \sum_{l=1}^L x_{hk} v_{il} \beta_{kl}^u\right)}$$

A forma das *market shares* mostra a importância da hipótese distribucional sobre o erro  $\epsilon$  — de modo intuitivo, tem-se agora uma agregação entre  $ns$  indivíduos *logit*. Caso tal hipótese não fosse feita, a complexidade envolvida no cálculo dos *market shares* seria ainda maior. Além disso, é preciso cuidado com o uso de métodos simulados, já que o seu uso introduz mais um erro — o erro de simulação. Como aprofundamos adiante, o número de simulações,  $ns$ , está ligado ao número de produtos, e é preciso que  $ns$  seja suficientemente grande para obter boas estimativas. O uso de métodos simulados permite que se introduza informações adicionais no processo de estimação — por exemplo, Berry, Levinsohn e Pakes utilizam técnicas de amostragem sobre renda a partir do censo norte-americano.

O segundo passo relativo à implementação do modelo nos remete novamente a Berry (1994), por consistir em obter  $\xi(\beta, \lambda)$  a partir dos *market shares* obtidos no primeiro estágio. De modo intuitivo, esse passo consiste em resolver o sistema que liga *shares* observadas e *shares* geradas pelo modelo, ou seja, resolver o sistema  $s_j^{obs} = s_j^{ns}(\theta, \delta)$  para  $j = 1, \dots, J$ . Infelizmente, a solução de tal sistema não é tão simples como no caso *logit*, mas é possível obtê-la de forma iterativa usando

$$\delta_j^k(\beta) = \delta_j^{k-1}(\beta) + \log(s_j^{obs}) - \log[s_j^{ns}(\beta, \delta^{k-1})]$$

partindo-se de um valor inicial  $\delta_j^0$  (por exemplo, os valores do modelo *logit*). Felizmente, pode-se mostrar que  $\{\delta^k\}$  é uma contração tendo, portanto, um ponto fixo e, consequentemente, uma única solução, que chamaremos  $\delta^* := \delta(\theta, s^{obs}, P^{ns})$ ,

onde  $P^{ns}$  é a distribuição de probabilidade obtida via simulação. Uma vez obtido  $\delta^*$  e lembrando-se que  $\delta_j = \sum_{k=1}^K x_{jk} \lambda_k + \xi_j$ , pode-se resolver para  $\xi_j$ , escrevendo então:

$$\xi_j(\theta, s^{obs}, P^{ns}) = \delta^* - \sum_{k=1}^K x_{jk} \lambda_k$$

onde  $\xi_j$  pode ser interpretado como um erro de estimação de uma regressão (estimada por, digamos, 2SLS) de  $\delta^*$  em  $\{x_{jk}\}_{k=1}^K$ .

O terceiro estágio relativo à implementação do modelo consiste em tomar o erro  $\xi_j(\theta, s^{obs}, P^{ns})$  e obter uma condição de momento com uma função de  $(x, w)$ , onde  $x$  são as características dos produtos e  $w$  os componentes de custo, obtendo o valor que leva a condição de momento tão próxima de 0 quanto possível.<sup>54</sup> Ou seja, dada a condição de momento

$$G_{J, n, ns}(\theta) = \frac{1}{J} \sum_{j=1}^J \xi_j(\theta, s^o, P^{ns}) \cdot f_j(x, w)$$

onde  $f_j(.,.)$  é uma função de todas características de produtos (e custos) que são ortogonais a  $\xi_j$ , obter o valor de  $\theta$  que minimize  $\|G_{J, n, ns}(\theta)\|$ .

As propriedades assintóticas do estimador de  $\theta$  podem ser obtidas de modo análogo às do GMM, devendo-se levar em conta agora o erro advindo da simulação. A matriz de covariância é dada por:

$$\text{cov}(\hat{\theta}) = \frac{1}{J} (\Gamma' \Gamma)^{-1} (\Gamma' V_0 \Gamma) (\Gamma' \Gamma)^{-1}$$

onde:

$J$  é o número de observações;

$\Gamma$  é a derivada do valor esperado da condição de momento com relação ao vetor de parâmetros, avaliada em  $\theta = \theta_0$ ; e

---

54. Remetemos o leitor à Subseção 1.2.5, onde instrumentos são tratados em detalhe.

$V_0 = V_1 + V^{\text{sim}}$  é a matriz de covariância das condições de momento avaliada em  $\theta = \theta_0(V_1)$ , somada à variância das simulações ( $V^{\text{sim}}$ ); esses termos são ortogonais entre si e o segundo deles pode ser obtido via *bootstrap*.<sup>55</sup>

*Estimação usando microdados.* Em Berry, Levinsohn e Pakes (1995), não se tem acesso a microdados, ou seja, variáveis representando atributos observados dos consumidores, ao contrário de Berry, Levinsohn e Pakes (2004), onde é possível combinar características individuais com informações tais como *a*) a probabilidade de comprar um carro; *b*) as características do automóvel, em caso de compra; além de *c*) informações sobre a segunda escolha do indivíduo, isto é, o automóvel que o indivíduo teria comprado caso não tivesse comprado a sua primeira escolha, obtida através de uma pesquisa de preferências declaradas (*stated preference survey*). Os itens *(a)* e *(b)* recuperam informações valiosas com relação ao bem exterior, e com relação às interações entre características observadas de consumidores e produtos. O item *(c)* recupera informações valiosas no que se refere a características não-observadas individuais e, intuitivamente, espera-se que dê mais precisão à estimação de elasticidades cruzadas.

Mais precisamente, os seguintes dados são utilizados em Berry, Levinsohn e Pakes (2004) — ver também a Seção 1.5:

- preços, quantidades vendidas, e características dos veículos, essencialmente os dados agregados de mercado de Berry, Levinsohn e Pakes (1995);
- para cada veículo vendido, as características da unidade familiar que o comprou (idade, renda, tamanho etc.); e
- segundas escolhas, obtidas a partir da pergunta “Que veículo compraria caso não tivesse comprado esse?”.

Como antes, os atributos individuais não-observados ( $v_j$ ) diferenciam esse modelo dos modelos *logit* tradicionais, mas há que se atentar para a complexidade adicionada ao incorporá-los ao processo de estimação envolvido.

Deve-se atentar para o fato de os microdados permitirem que se estime um intercepto  $\delta_j$  para cada escolha, e que a estimação de  $\beta := (\beta^o, \beta^u)$  não necessita de hipóteses distribucionais sobre  $\xi_j$ . No entanto, o termo  $\delta_j = \sum_{k=1}^K x_{jk} \lambda_k + \xi_j$  deverá ser necessariamente decomposto, uma vez que se necessita do valor de  $\lambda$  ao se calcular as elasticidades com relação às características  $x$ , inclusive preço.

---

55. O procedimento de *bootstrap* para calcular o impacto das simulações na matriz de covariância consiste em gerar realizações de  $v$  e recalcular a condição de momento  $G(\cdot)$  várias vezes no ponto  $\theta = \hat{\theta}$ . Ver Berry, Levinsohn e Pakes (1995) e Berry, Linton e Pakes (2004) para maiores detalhes, e Wooldridge (2002, cap. 12) para um tratamento de livro-texto.

Várias hipóteses com relação a  $\xi_j$  podem ser feitas, e a identificação desse modelo é idêntica ao de Berry, Levinsohn e Pakes (1995), em que não se tem microdados: tipicamente,  $E[\xi | x_1, w] = 0$ , onde  $x_1$  são as características exógenas dos produtos e  $w$  os componentes de custo.

No que concerne ao método de estimação, o micro BLP em muito se assemelha ao clássico BLP. No primeiro estágio, tal como em BLP, para cada  $\beta$  dado, recorre-se a Berry (1994) para resolver numericamente o sistema  $s(\beta, \delta, P^{ns}) = s^{obs}$  (que iguala *market shares* gerados pelo modelo a *market shares* observados) para  $\delta^* = \delta^{N, ns}(\beta)$ . O segundo estágio consiste em plugar o valor de  $\delta^*$  obtido na contração em  $\xi$  e, dada uma hipótese de identificação, definir uma restrição de covariância nula entre erro e instrumentos — define-se o estimador de  $\beta$  como o argumento minimizando uma norma da condição de momento empregada.

#### 1.2.4.7 Modelos da família GEV

Uma alternativa a modelos NL e *logit* com coeficientes aleatórios é a família de modelos baseada na distribuição Generalized Extreme Value (GEV), inicialmente proposta em McFadden (1978a), que nada mais é do que uma generalização da hipótese de o erro seguir uma distribuição de valores extremos. Nesta subseção daremos ênfase ao modelo Principles of Differentiation Generalized Extreme Value (PD-GEV), proposto e implementado em Bresnahan, Stern e Trajtenberg (1997).

Ao contrário do NL, o modelo PD-GEV permite tratar potenciais segmentações de mercado diferentes de forma simétrica e não-aninhada, contornando assim a estrutura hierárquica (e consequentes limitações) do NL. Tal como os modelos anteriores, o PD-GEV é baseado no arcabouço RUM, e suas demandas de mercado são obtidas como agregações das demandas individuais. Além disso, ao contrário do NL, ele também permite o tratamento tanto de novos produtos, como de produtos não mais produzidos (bastando-se definir os “princípios de diferenciação” correspondentes — ver adiante). Finalmente, apesar da conveniência de usar os métodos propostos em Berry (1994), à semelhança de modelos *logit* em geral, o modelo PD-GEV não usa de integração numérica para a obtenção das *market shares*.

Seguindo BST, considere um mercado com  $J + 1$  produtos (sendo o bem  $j = 0$  o bem exterior) em que o consumidor  $i$  se defronta com o problema

$$\max_{j \in \{0, \dots, J\}} V_{ij} = X'_j \beta + \alpha p_j + \xi_j + \varepsilon_{ij}$$

onde:

$V_{ij}$  é o valor do produto  $j$  para o indivíduo  $i$ ;

$X_j$  é um vetor de características observadas do produto  $j$ ; e

$\xi_j$  é o termo de características não-observadas do produto  $j$ .

Assim, define-se  $\delta_j = X'_j \beta + \alpha p_j + \xi_j$  como a valoração média do produto  $j$  por parte dos consumidores desse mercado, e  $\varepsilon_{ij}$  como a diferença entre a valoração do indivíduo  $i$  e valoração média  $\delta_j$ . Desse modo, cada consumidor recebe uma realização do vetor  $\varepsilon_i$  de dimensão  $(J+1) \times 1$ , que é uma realização da variável aleatória  $\varepsilon$  — apesar das realizações serem independentes entre indivíduos  $i$ ,  $\varepsilon_{ij}$  não precisa ser independente de  $\varepsilon_{ij'}$  se os produtos  $j$  e  $j'$  têm características em comum. A função distribuição  $F(\varepsilon; X, p)$  (de dimensão  $J+1$ ) de  $\varepsilon$  depende das características  $X$  e do parâmetro  $p$  (que controla o grau de dependência entre produtos com características em comum), a ser estimado com  $\alpha$  e  $\beta$ .

A aplicação de BST é feita para o mercado de computadores pessoais (PC's) no final da década de 1980. Os autores classificam PC's como seguindo dois princípios de diferenciação (PD's): *a*) sendo ou não de fronteira (F ou NF — modelos avançados ou não); e *b*) sendo ou não de marca (B ou NB — do inglês, *branded* ou *non-branded*). Eles então parametrizam a função distribuição  $F(\cdot)$  de forma que vários PD's possam ser incorporados simetricamente.

Tal formulação depende do seguinte resultado adaptado de McFadden (1978a):

Teorema. Se  $G := \mathbb{R}^{J+1} \Rightarrow \mathbb{R}$  é uma função não-negativa, homogênea de grau 1 satisfazendo condições de regularidade,<sup>56</sup> então:

$$F(\varepsilon_{i0}, \varepsilon_{i1}, \dots, \varepsilon_{ij}) = \exp\left[-G\left(e^{\varepsilon_{i0}}, e^{\varepsilon_{i1}}, \dots, e^{\varepsilon_{ij}}\right)\right]$$

---

56. O limite de  $G(\cdot)$  tomado-se o limite de qualquer argumento tendendo a  $\infty$  deve ser igual a  $+\infty$ , ou seja,

$\lim_{r_j \rightarrow \infty} G(r_0, r_1, \dots, r_j) = +\infty$ ; além disso,  $\frac{\partial^{(k)} G}{\partial r_{j_1} \dots \partial r_{j_k}}$  deve ser não-negativa se  $k$  for ímpar, e não-positiva se  $k$  for par, para

$(j_1, \dots, j_k)$  distintos obtidos de  $[1, \dots, J]$ . Para detalhes e uma prova construtiva do resultado, ver McFadden (1978a).

é a função distribuição multivariada de uma distribuição de valores extremos, e a equação determinando a *market share* do produto  $j$  é dada por

$$s_j = \frac{e^{\delta_j} \cdot G_j(e^{\delta_0}, \dots, e^{\delta_J})}{G(e^{\delta_0}, \dots, e^{\delta_J})}$$

onde  $G_j$  é a derivada parcial de  $j$  com relação a  $e^{\delta_j}$ .

De acordo com McFadden (1978a), tal resultado teórico fornece formas de gerar equações de *market share* após a especificação da função  $G()$ , com a relação de dependência dos  $\epsilon_{ij}$ 's entre os produtos sendo estabelecida pelo analista. Ou seja, em última análise, tal resultado fornece uma forma de parametrizar a substituibilidade entre os produtos.

*Exemplo (NL com um nível como caso particular de PD-GEV).* Considere o caso em que PC's são descritos como sendo ou não de fronteira (F ou NF, respectivamente). Nesse caso,

$$G(e^\delta) = \left( \sum_{j \in F} e^{\delta_j / \rho_F} \right)^{\rho_F} + \left( \sum_{j \in NF} e^{\delta_j / \rho_F} \right)^{\rho_F} + e^{\delta_0}$$

onde  $\rho_F$  parametriza o grau de substituição entre produtos de mesmo nicho (F ou NF) em relação à substituição entre produtos de nichos diferentes. Convém ressaltar que, para que o modelo seja consistente com RUM, deve-se ter  $\rho$  tomando valores entre 0 e 1 — caso seu valor tenda a 0, a dependência entre produtos em um mesmo nicho aumenta, enquanto no caso de seu valor tender a 1, os elementos de  $\epsilon$  são independentes entre si, fazendo o NL resultar em um MNL (ou seja, o PD considerado não é relevante para a segmentação de mercado). ■

A contribuição de Bresnahan, Stern e Trajtenberg é parametrizar um modelo RUM que incorpora categorias não-aninhadas de produtos (ao contrário do que ocorre com o NL). Conforme descrito, Bresnahan, Stern e Trajtenberg consideram duas dimensões distintas: B/NB e F/NF. Nesse caso,  $G()$ , é obtida como a soma ponderada de duas funções  $G()$  de um NL com um nível apenas. Assim,

$$\begin{aligned} G(e^\delta) &= \alpha_F \cdot \left[ \left( \sum_{j \in F} e^{\delta_j / \rho_F} \right)^{\rho_F} + \left( \sum_{j \in NF} e^{\delta_j / \rho_F} \right)^{\rho_F} \right] + \\ &\quad + \alpha_B \cdot \left[ \left( \sum_{j \in B} e^{\delta_j / \rho_B} \right)^{\rho_B} + \left( \sum_{j \in NB} e^{\delta_j / \rho_B} \right)^{\rho_B} \right] + e^{\delta_0} \end{aligned}$$

onde  $\alpha_F = \frac{1 - \rho_F}{2 - \rho_F - \rho_B}$  e  $\alpha_B = \frac{1 - \rho_B}{2 - \rho_F - \rho_B}$ , fazendo com que  $\alpha_F + \alpha_B = 1$  e  $\rho_F, \rho_B \in (0, 1]$ .

Tal como no exemplo anterior, os parâmetros  $\rho$  parametrizam o grau de correlação intra-ramo entre os produtos e, caso seu valor seja igual a 1, o PD correspondente não é importante para a segmentação de mercado. Em particular,  $\rho_F = 1 (\rho_B = 1)$  implica que o modelo é um NL para *brands (frontier)*, apenas. As *market shares* associadas são dadas por

$$s_j = \frac{\alpha_F \cdot \frac{e^{\delta_j/\rho_F}}{\sum_{k \in F} e^{\delta_k/\rho_F}} \cdot \left( \sum_{k \in F(j)} e^{\delta_k/\rho_F} \right)^{\rho_F} + \alpha_B \cdot \frac{e^{\delta_j/\rho_B}}{\sum_{k \in B} e^{\delta_k/\rho_B}} \cdot \left( \sum_{k \in B(j)} e^{\delta_k/\rho_B} \right)^{\rho_B}}{G(e^\delta)}$$

onde  $B(j), F(j)$  são os grupos a que o produto  $j$  pertence. Convém notar que a equação de *share* é composta de dois termos, um para  $F$ , outro para  $B$ , e o numerador da equação é proporcional ao NL com um nível. Para um produto situado em  $\{B(j), F(j)\}$ , mudanças em preços ou características de produtos localizados no mesmo ramo terão um impacto relativamente forte na *market share*  $j$ ; mudanças em produtos que não estão em  $B(j)$  nem em  $F(j)$  terão um impacto muito menor, enquanto mudanças em produtos que se situam em  $B(j)$  ou  $F(j)$  terão um impacto intermediário — tal como antes, efeitos se tornam mais acentuados conforme  $\rho$  tenda a 0.

O modelo PD-GEV consiste então numa alternativa interessante diante de vários competidores. Primeiro, por permitir diversas dimensões no que se refere à diferenciação de produto, ele se mostra mais apropriado do que modelos clássicos de diferenciação vertical. Segundo, ao contrário do NL, vimos que aqui a segmentação ocorre em paralelo, sem uma hierarquia predefinida, o que torna o modelo mais desejável e menos passível de contestações em aplicações práticas. Terceiro, por ter *market shares* que podem ser escritas de forma analítica, o modelo se mostra uma alternativa menos demandante em termos computacionais do que o modelo *logit* com coeficientes aleatórios, em que é necessário simular as *market shares* a cada iteração da contração — vale ressaltar aqui que o procedimento de inversão de Berry (1994) se acelera significativamente, pela menor dimensão das matrizes consideradas do que no caso com simulação, e a estimação via GMM é análoga ao caso do *logit* com coeficientes aleatórios. Quarto, apesar de ainda não ser do nosso conhecimento haver implementações, generalizações de modelos da

família GEV surgem como alternativas promissoras na literatura teórica de demanda, conforme Davis (2001) — no caso do modelo PD-GEV, Peters (2001) analisa os efeitos de fusões e aquisições no setor de transporte aéreo americano.

### **1.2.5 Identificação de demanda**

#### **1.2.5.1 Motivação**

Em uma especificação típica de demanda, como a expressa em (12), as variáveis de preços são endógenas. Os preços e as quantidades observados pelo econometrista refletem as realizações de equilíbrio do mercado — uma realização para cada abertura de mercado (seja, por exemplo, período no tempo ou região geográfica) — na medida em que as curvas de demanda dos consumidores e as relações de oferta das firmas são deslocadas de forma exógena pelas variáveis exógenas de demanda e as variáveis exógenas de oferta. Intuitivamente, para identificarmos a curva de demanda de forma apropriada, é essencial distinguirmos as alterações de preços e de quantidades que resultam do deslocamento da relação de oferta (os pontos de equilíbrio resultantes encontram-se ao *longo* da curva de demanda, permitindo-nos medir sua inclinação) das alterações de preços e de quantidades que resultam do deslocamento da curva de demanda (ou seja, estas não mapeiam a curva de demanda, e sim a relação de oferta). Com relação à identificação da curva de demanda (12), todo esforço é feito para controlar essa variação nas condições de demanda, refletida no deslocamento das variáveis  $Y$ , observada pelo econometrista. No entanto, é natural que uma parcela não-trivial dessa variação nas condições de demanda não seja observada pelo econometrista, sendo capturada então pelo erro econométrico  $\epsilon$ . O problema de identificação de demanda surge justamente quando essa parcela de variação em  $\epsilon$ , apesar de não observada pelo econometrista, é observada pelas firmas. Dessa forma, os preços estarão correlacionados com o erro  $\epsilon$ : em determinado período, por exemplo, o preço estará alto não porque o custo terá subido, mas porque ocorreu um choque de demanda positivo, levando a firma a reajustar seu preço para cima (ocasionando também um aumento na quantidade). Consequentemente, estimação usando o método de mínimos quadrados ordinários não gerará estimativas consistentes dos parâmetros  $\alpha$  (ou seja, não identificará a curva de demanda). Em termos práticos, ignorar esse fenômeno tipicamente resulta em coeficientes-preço menos negativos do que na realidade, decorrente da correlação positiva entre preços e  $\epsilon$  em virtude da inclinação positiva da relação de oferta. No limite, pode-se inclusive obter coeficientes positivos (sugendo curvas de demanda positivamente inclinadas!), tal como em Trajtenberg (1989).

A solução para esse problema é utilizar a técnica de variáveis instrumentais.<sup>57</sup> Em termos econômicos, desejamos obter variáveis excluídas da função de demanda (12) que estejam correlacionadas com os regressores endógenos (preços  $p$ ) mas que não estejam correlacionadas com o erro  $\epsilon$  (choques de demanda não observados pelo econometrista). Dessa forma, podemos associar variações nesses chamados instrumentos a movimentações do ponto de equilíbrio *ao longo* da curva de demanda, em vez de deslocamentos, traçando assim a curva de demanda.

#### 1.2.5.2 “Classes” de instrumentos de demanda

Nesta subseção, tecemos comentários sobre os diferentes grupos, ou “classes”, de instrumentos que têm sido utilizados na literatura de estimação de demanda. Estes variam desde os tradicionais “deslocadores de oferta”, utilizados de forma direta ou implícita, a variáveis defasadas (em níveis ou diferenças) e a outras variáveis disponíveis de acordo com o caso particular sendo estudado.

*Deslocadores de oferta.* São os instrumentos clássicos na análise de demanda, tipicamente empregados na estimação de demanda por produtos homogêneos, ou como parte do conjunto de instrumentos utilizado na estimação de demanda por produtos diferenciados (por exemplo, na estimação da equação de nível superior (17) em um sistema de escolha em estágios múltiplos). A intuição é dada acima: na medida em que a relação de oferta se desloca exogenamente, esta traça (ou identifica) a curva de demanda. Deslocadores de oferta podem ser, por exemplo, preços de fatores, custos de venda, ou a tecnologia empregada (características de processo). A premissa identificadora é que tais deslocadores de oferta não são correlacionados com os choques de demanda não-observados.

O problema do uso de deslocadores de oferta é que, na melhor das hipóteses, os dados disponíveis são no nível da indústria ou mercado (por exemplo, preços de fatores), não havendo variação no nível dos produtos. Na estimação de demanda por produtos diferenciados, o número de deslocadores de oferta é tipicamente insuficiente, ou estes não estão disponíveis na mesma freqüência (periodicidade) que os dados de preços e quantidades das marcas. Assim, se apenas essa classe de instrumentos for usada, seu uso está limitado a sistemas de demanda muito restritivos; daí o seu uso ser feito junto com outras variáveis instrumentais, descritas a seguir.

*Preços em outras regiões.* Em lidando com dados em painel, em que um mesmo produto é vendido em diversas regiões, Hausman e seus co-autores têm defendido

57. Na prática, o que muitas vezes é usado é o método conhecido como mínimos quadrados de dois estágios (2SLS).

o uso de preços em outras regiões como instrumentos para o preço em determinada região. A proposição de uso dessa classe de instrumentos vai de encontro à necessidade de se dispor de instrumentos em número considerável ao estimar um sistema de demanda para produtos diferenciados, em que o número de variáveis endógenas é considerável. Na estimação de, por exemplo, uma equação de nível inferior (15) em um sistema de escolha em estágios múltiplos, o pesquisador dificilmente disporá de preços de fatores (o que foi discutido anteriormente em deslocadores de oferta) em quantidade igual ou superior ao número de marcas substitutas, ou estes não estarão disponíveis em periodicidade suficientemente alta. Sendo assim, o uso de instrumentos “tradicionalis” utilizando o deslocamento de oferta de forma direta não bastará. Nesse momento poderá convir utilizar o deslocamento de oferta (ou custos) de forma implícita, utilizando preços em outras regiões como instrumentos. A idéia é que os preços de um mesmo produto (ou marca) em diferentes regiões têm um componente de custos comum a todas as regiões e um componente não-observado de demanda específico a cada região. Na medida em que esses choques de demanda específicos a cada região são independentes entre si (condicional a outros fatores que podem ser controlados, como sazonalidade ou despesas de publicidade observadas), preços em uma região podem servir de instrumentos para preços em outra região: os preços em ambas as regiões estarão correlacionados em virtude do componente comum de custos, ao passo que o preço na região que está servindo de instrumento não estará correlacionado com o choque não-observado de demanda ocorrido na outra região cujo preço está sendo instrumentado.

Uma situação em que tal premissa falha é quando ocorrem despesas em publicidade comuns às diversas regiões, e essas despesas (ou uma parcela não-trivial delas) não são observadas pelo econometrista, porém são levadas em consideração pelas firmas quando estas estabelecem seus preços. Por outro lado, em indústrias de produtos diferenciados tipicamente há considerável variação exógena nos preços em decorrência de promoções realizadas pelo varejo (ou seja, produtos entram e saem de promoção). Como Hausman tem defendido, na medida em que esses preços são estabelecidos em momento anterior à realização das condições de demanda, o problema de publicidade comum não-observada pode ser atenuado (ou mesmo pode-se vislumbrar o método dos mínimos quadrados ordinários gerando estimativas consistentes, já que podemos tratar preços no momento de realização de demanda como sendo predeterminados).<sup>58</sup>

---

58. Ver, por exemplo, o tratamento exógeno dado a preços em Capps, Church e Love (2003).

Para fixar idéias, suponha que os dados disponíveis incluam diversos produtos (indexados por  $j$ ), em diferentes regiões (indexadas por  $n$ ) e em diferentes períodos (indexados por  $t$ ). Pode-se então escrever:

$$\begin{aligned} p_{jnt} &= cmg_{jnt} + markup_{jnt} \\ &= cmg_{jt} + \Delta cmg_{jnt} + markup_j + markup_n + markup_t + \Delta markup_{jnt} \end{aligned}$$

onde  $\Delta cmg_{jnt}$  é o componente do custo marginal específico da região  $n$ , e  $markup_j$ ,  $markup_n$ ,  $markup_t$  são, respectivamente, os *markups* específicos do produto  $j$ , da região  $n$  e do período  $t$ . A decomposição ilustra o fato de preços de regiões distintas serem correlacionados em virtude dos choques comuns no custo marginal ( $cmg_j$ ). Nesse caso a premissa identificadora é de que choques de demanda ( $\epsilon_{jnt}$ ) são independentes entre regiões  $n$ , após se ter controlado para efeitos fixos de produtos e regiões, despesas publicitárias etc.

*Preços defasados, em níveis ou em diferenças.* Em certas condições, preços defasados estarão correlacionados com preços correntes, em decorrência de correlação serial nas variáveis exógenas seja do lado da demanda ou do lado da oferta, ao passo que a correlação entre preços defasados e choques de demanda correntes pode tender a ser pequena, em virtude da baixa correlação serial entre os choques de demanda. Ou seja,  $p_{t-1}$  está correlacionado com  $p_t$  (além de estar correlacionado com  $\epsilon_{t-1}$ ), mas como não há correlação entre  $\epsilon_{t-1}$  e  $\epsilon_t$  (ou esta for relativamente baixa), não há correlação entre  $p_{t-1}$  e  $\epsilon_t$ . Nessas condições, preços defasados poderão ser utilizados como instrumentos para preços correntes. Intuitivamente, em vez da premissa de não-correlação entre os choques de demanda das diferentes regiões pertinente ao segundo grupo de instrumentos (preços em outras regiões), tem-se aqui a premissa de não-correlação entre os choques de demanda dos diferentes períodos.

*Características de produtos.* Em modelos de produtos diferenciados em que as características<sup>59</sup> são tratadas como exógenas (predeterminadas), é razoável supor que a competição varie com a localização, resultando em *markups* diferentes; ou seja, em última análise, os preços variam com a localização das características — idéia que aparece na literatura ao menos desde Bresnahan (1987). Seguindo tal linha, Berry, Levinsohn e Pakes (1995) supõem que  $E(\xi|X)=0$ , ou seja, que as características não-observadas, condicionadas às características observadas, têm

---

59. Ou a localização no espaço de características, no contexto de modelos discretos.

média 0. Isso implica que qualquer função genérica de  $X$  pode ser usada como instrumento, e eles propõem o uso:

- das próprias características do produto;
- das somas das características dos outros produtos produzidos pela *mesma firma*; e
- das somas das características dos produtos produzidos pelas *outras firmas*.

O primeiro conjunto de instrumentos decorre de modo imediato da exogeneidade de  $X$ . O uso de características de outros produtos pode ser racionalizado da seguinte forma: “produtos com bons substitutos tendem a ter *markups* baixos, enquanto produtos não sofrendo concorrência tão intensa tendem a ter *markups* mais altos, logo preços mais altos com relação aos custos” [Berry, Levinsohn e Pakes (1995, p. 855); tradução livre]. Quanto à distinção entre produtos produzidos pela própria firma e produtos produzidos pelas outras firmas, Berry, Levinsohn e Pakes justificam seu uso da seguinte forma: “Como *markups* (determinados por firmas maximizadoras se comportando segundo a hipótese de Nash) vão responder de forma diferente a produtos da própria firma e produtos de firmas rivais,<sup>60</sup> os instrumentos devem distinguir entre características de produtos produzidos pela mesma firma multiproduto e características de produtos produzidos por firmas rivais” (*loc. cit.*).

Características de produtos substitutos podem ser utilizadas como instrumentos na estimação de uma equação como a (25) [ver Berry (1994) e Berry, Levinsohn e Pakes (1995)], e têm sido variáveis instrumentais amplamente utilizadas na última década, ao menos entre modelos discretos. Observe que as características de produtos substitutos  $k \neq j$  estão excluídas da equação referente à demanda pelo produto  $j$ , mas estão correlacionadas com o preço  $p_j$  através do lado de oferta (decorrente da interação estratégica entre firmas, capturadas nas condições de primeira ordem).

Há, no entanto, alguns argumentos que geram dúvida quanto à validade de tais instrumentos:

- Mesmo que as características sejam determinadas antes de a firma estabelecer seus preços, não é improvável que elas tenham sido escolhidas já levando em consideração um valor esperado para as características não-observadas.
- Se as características mudam muito, por que elas não seriam correlacionadas com  $\xi$ ?

---

60. Ver, ainda, a discussão referente à ilustração da Subseção 1.1.1.

- Se as características não mudam, por que então, supondo que se tenham dados o suficiente, não usar efeitos fixos ?

Ainda supondo a exogeneidade das características  $X$ , Bresnahan, Stern e Trajtenberg (1997) propõem uma série de instrumentos baseados na segmentação de mercado de seu modelo PD-GEV para o preço do produto  $j$ :

- o número de produtos do segmento em que  $j$  se encontra;
- o número de produtos vendidos pela mesma firma que produz  $j$ ;
- o número de produtos vendidos pela mesma firma que produz  $j$  no mesmo segmento em que  $j$  se encontra;
- a soma e a média das características dos outros produtos produzidos pela *mesma firma* que produz  $j$ , tal como em Berry, Levinsohn e Pakes (1995);
- a soma e a média das características dos outros produtos localizados no mesmo segmento em que  $j$  se encontra; e
- a soma e a média das características dos outros produtos localizados no mesmo segmento em que  $j$  se encontra e produzidos pela *mesma firma* que produz  $j$ .

O primeiro conjunto de instrumentos se baseia no fato de que o aumento do número de produtos em um determinado segmento tende a tornar a curva de demanda mais horizontal e deslocá-la para dentro — ambos os movimentos afetando tanto o preço do produto, como seu *market share* no segmento. Ou seja, explora-se a hipótese de que a entrada em um segmento específico é exógena (em termos econometríticos) para construir uma medida local do número de substitutos de um produto. O segundo se baseia no fato de que uma firma com um número maior de produtos tende a aproveitar a substituibilidade entre eles aumentando seus preços. O terceiro conjunto combina os dois anteriores, explorando a substituibilidade intra-segmento e intrafirma. Tal como em Berry, Levinsohn e Pakes, o quarto conjunto de instrumentos se baseia no fato de a soma das características dos demais produtos ser positivamente correlacionada com preço. O quinto e sexto conjuntos são racionalizados de modo análogo ao primeiro e terceiro conjuntos, respectivamente.

*Deslocadores de custos interagidos com efeitos fixos de produtos.* A intuição subjacente ao seu uso é deixar que os dados determinem como produtos diferentes vão reagir de forma diferente a choques de oferta em função de, por exemplo, fabricantes diferentes usarem combinações de insumos diferentes para produzir seus produtos [ver Nevo (2000 e 2001) para detalhes]. Ou seja, em termos formais está se supondo que

$$cmg_{jt} = f_j(\text{deslocadores de oferta}_t)$$

O uso de tal classe de instrumentos pressupõe que se tenha controlado para produtos no lado da demanda (via efeitos fixos de marca) — do contrário, seriam capturadas diferenças no custo médio dos produtos.

*Variáveis disponíveis e adequadas de acordo com o caso em estudo.* Como ilustração, considere inicialmente o estudo de Genesove e Mullin (1998) sobre a indústria de refino de açúcar nos Estados Unidos na virada do século XX. Na época, a indústria importava o açúcar em forma bruta de Cuba e de outros países, procedendo ao seu refino. Os autores modelam a indústria como sendo de produtos homogêneos e na estimativa de demanda precisam instrumentar a variável endógena preços. Eles optam por não utilizar o preço de açúcar bruto importado (preço do principal fator) como instrumento pois assumem que a premissa identificadora decorrente, apesar de não-testável, não deve valer aqui: os Estados Unidos já correspondiam a 25% do consumo mundial de açúcar e, portanto, o preço internacional de açúcar bruto não pode ser tratado como não-correlacionado com os choques de demanda domésticos. Os autores optam sim por utilizar a *quantidade* de açúcar bruto importado de Cuba, uma série felizmente disponível a eles, como instrumento para o preço (de açúcar refinado), cuidadosamente justificando sua escolha. Tal justificativa resume-se em: *a)* Cuba era a fonte mais barata de açúcar bruto para os Estados Unidos, mas não a fonte única, nem tampouco uma fonte de volume constante; e *b)* o que determinava a quantidade de açúcar bruto embarcada de Cuba para os Estados Unidos era a colheita cubana, e não as condições de demanda nos Estados Unidos, Cuba embarcava praticamente a totalidade de sua colheita para os Estados Unidos, e estes supriam o restante de suas necessidades de outros países, como a Europa. Dessa forma, a quantidade de açúcar bruto importado de Cuba estaria correlacionada com o preço de açúcar refinado (em virtude de uma boa colheita, uma importação maior de Cuba, em detrimento da proveniente de outros países, deslocaria a relação de oferta para a frente), mas não estaria correlacionada com os choques de demanda do mercado consumidor nos Estados Unidos.<sup>61</sup>

Ao estudar o mercado de cereais, Nevo (2000 e 2001) considera *dummies* regionais, cujo intuito é capturar custos de transporte; densidade populacional das cidades (definida como o número de habitantes por milha quadrada), como *proxy* das diferenças do custo de espaço (ou *display* dos produtos); e renda média

---

61. Dada a proximidade entre Cuba e o mercado consumidor na costa leste dos Estados Unidos, além do consumo neste país ser sempre superior à produção cubana, os autores argumentam que mesmo uma realização de demanda muito baixa dificilmente redirecionaria o açúcar cubano para outros países consumidores. Dessa maneira, Cuba era uma fonte inframarginal de açúcar bruto para o mercado americano, ao menos no curto prazo e, portanto, sua produção era exógena à demanda desse mercado. Os autores também argumentam que choques de demanda não poderiam ocasionar variações especulativas nos níveis de armazenagem de açúcar bruto em Cuba [ver Genesove e Mullin (1998)].

do setor de supermercados das cidades consideradas. No entanto, tais instrumentos não variam entre produtos, sendo insuficientes para identificação, exceto se combinadas com outras classes de instrumentos.

### **1.3 INCORPORANDO OFERTA: CUSTOS E CONDUTA**

Um problema clássico em organização industrial consiste em determinar empiricamente o padrão de comportamento em um oligopólio a partir de dados observados. Por padrão de comportamento entende-se o(s) modelo(s) de conduta entre as firmas de maneira geral, em todas as dimensões nas quais essa interação estratégica pode ocorrer, como concorrência em preços, quantidades, investimentos em capacidade, investimentos em P&D ou gastos em mídia. Dado o escopo deste estudo, focaremos aqui a questão que mais tem atraído pesquisadores empíricos voltados ao estudo de oligopólios: como determinar econometricamente o modelo comportamental que está gerando os dados observados de preços e quantidades. Discutiremos também um estudo voltado à concorrência em outras dimensões. Uma aplicação relevante à prática antitruste, na situação comum em que custos são informação privada, trata de estimar o diferencial entre preços e custos marginais, e assim o poder de mercado de firmas no oligopólio, a partir da estatística comparativa do equilíbrio. Para tanto, iniciamos esta seção definindo um modelo estrutural estático que, segundo Bresnahan (1989), reúne vários estudos da (antes) chamada “nova organização industrial empírica”. Consideraremos a situação geral em que custos não são conhecidos, e parâmetros estruturais dos lados de demanda e de oferta — elasticidades, custos e conduta — são estimados. Na seqüência, apresentaremos duas metodologias econométricas que têm sido utilizadas para inferir o modelo estrutural, analisando a questão de identificação. Situações particulares onde o analista tem conhecimento (integral ou parcial) sobre custos e/ou conduta serão consideradas, com algumas aplicações seminais da literatura.

#### **1.3.1 Estimação de um modelo estrutural estático**

Seguindo Bresnahan (1989), convém escrever a função de demanda na forma inversa:

$$P_t = P(Q_t, Y_t, \alpha, \varepsilon_t^d) \quad (31)$$

onde, em uma indústria de produtos homogêneos,  $P_t$  denota preço (no período  $t$ ),  $Q_t$  denota quantidade agregada,  $Y_t$  são as variáveis exógenas que deslocam a curva de demanda,  $\alpha$  são parâmetros a serem estimados e  $\varepsilon_t^d$  é o erro econômétrico da

especificação de demanda. O lado de oferta nesse modelo estilizado é especificado no nível da firma, em que a função de custos da firma  $i$  é dada por

$$C_{it} = C(q_{it}, W_{it}, Z_{it}, \beta, \varepsilon_{it}^c) \quad (32)$$

onde  $q_{it}$  é a quantidade da firma  $i$ , sendo que  $\sum_i q_{it} = Q_t$ ,  $W_{it}$  são os preços de fatores pagos pela firma,  $Z_{it}$  são outras variáveis que deslocam a relação de oferta da firma,<sup>62</sup>  $\beta$  são parâmetros a serem estimados e  $\varepsilon_{it}^c$  é o erro econometrônico da especificação de custos. O custo marginal  $MC$  da firma  $i$  é obtido a partir da primeira derivada de (32) com relação a  $q_{it}$ , denotada por  $C_1(q_{it}, \cdot)$

$$MC = C_1(q_{it}, W_{it}, Z_{it}, \beta, \varepsilon_{it}^c) \quad (33)$$

(Vê-se por (33) que esta especificação admite o erro entrando em (32) de forma não-linear.)

A relação de oferta, ou equação de especificação (*pricing equation*), da firma  $i$  é então especificada de forma abrangente a partir de sua CPO

$$P_t + \theta_{it} Q_t P_1(Q_t, Y_t, \alpha, \varepsilon_t^d) = C_1(q_{it}, W_{it}, Z_{it}, \beta, \varepsilon_{it}^c) \quad (34)$$

onde  $P_1(Q_t, \cdot)$  é a primeira derivada de (31) com relação a  $Q_t$ , e  $\theta_{it}$  é um parâmetro que captura a conduta da firma  $i$ , ou um índice do “grau de concorrência da conduta oligopolística” [Bresnahan (1989, p. 1.016)].<sup>63</sup> A relação (34) abrange vários padrões alternativos de conduta. Em seus dois extremos, concorrência perfeita corresponde a  $\theta_{it} = 0$ , onde a receita marginal da firma competitiva é dada por  $P_t$ , ao passo que conluio perfeito (cartelização de firmas homogêneas) corresponde a

62. Como vimos ao se discutirem deslocadores de oferta na identificação de demanda,  $Z$  inclui, por exemplo, características de processo que são fixas no curto prazo (ou no prazo relevante do estudo) e impactam custos, como a tecnologia ou tamanho de um forno.

63. Observe que em nossa exposição optamos por especificar (34), e portanto  $\theta_{it}$ , de maneira diferente de Bresnahan (1989), que escreve a receita marginal como  $P_t + \theta_{it} q_{it} P_1(Q_t, \cdot)$ . Enquanto em nossa especificação  $\theta_{it}$  pode variar entre 0 e 1 (com  $\theta_{it} = 1$  correspondendo a conluio perfeito; ver adiante), na especificação de Bresnahan conluio perfeito corresponde a  $\theta_{it} = Q_t/q_{it} = 1/S_{it}$ . Pela especificação de Bresnahan, podemos usar o problema da firma,  $\max_{q_{it}} \geq 0 P(Q_t, \cdot) q_{it} - C(q_{it}, \cdot)$ , e a CPO dele resultante,  $P_t + q_{it} P_1(Q_t, \cdot) (dQ_t/dq_{it}) = C(q_{it}, \cdot)$  para motivar o parâmetro de conduta como uma “variação conjectural”, onde  $\theta_{it} = dQ_t/dq_{it}$ . Por essa interpretação, ao incrementar sua quantidade em  $dq_{it}$ , a firma  $i$  detém uma “conjectura” com respeito ao incremento agregado resultante da indústria  $dQ_t$ . Tal interpretação tem sido criticada por não estar sedimentada em um modelo teórico de oligopólio [ver Corts (1999) e Reiss e Wolak (2002)].

$\theta_{it} = 1$ , onde a receita marginal do cartel (sem a presença de uma franja competitiva, ou *competitive fringe*) é dada por  $P_t + Q_t P_1(Q_t, \cdot)$ . Observe que dessa forma o modelo estático dado por (31) e (34) retrata resultados (*outcomes*) — conluio entre firmas — que são apoiados apenas como equilíbrios de um jogo repetido, ou seja, de um modelo dinâmico. Já o comportamento de uma firma Cournot corresponde a  $\theta_{it} = \frac{q_{it}}{Q_t} = s_{it}$ , onde  $s_{it}$  denota a participação de mercado da firma; a receita marginal da firma Cournot é dada por  $P_t + q_{it} P_1(Q_t, \cdot)$ . A equação de precificação (34) pode ser escrita em termos da elasticidade-preço de demanda  $E(Q_t)$ :

$$P_t + \theta_{it} \frac{P_t}{E_t} = c_{it} \quad (35)$$

(onde suprimimos os argumentos das funções e reescrevemos custo marginal como  $c_{it} := C_1(q_{it}, \cdot)$ , por mera conveniência). Observe que esse modelo acomoda heterogeneidade entre firmas, associando variações em  $\theta_{it}$  entre firmas a variações em seus custos  $c_{it}$ . Podemos ainda reescrever a equação de precificação na forma de índice de Lerner (ou *markup* preço-custo) ajustado pela elasticidade-preço de demanda, tipicamente utilizado para definir “poder de mercado”:

$$\theta_{it} = -E_t \frac{P_t - c_{it}}{P_t} \quad (36)$$

Fica evidente por (36) que, no equilíbrio, uma margem preço-custo  $P_t - c_{it}$  elevada pode estar associada a uma baixa elasticidade de demanda  $E_t$  ou a um elevado poder de mercado  $\theta_{it}$ .

Discorremos agora sobre as duas metodologias que têm sido avançadas de forma a estimar o modelo estrutural dado por (31) e (34).

#### 1.3.1.1 Estimação 1: estimação de conduta como parâmetro livre

Vários artigos na literatura de estimação de conduta<sup>64</sup> estimam um sistema formado pela equação de demanda (31) e por equações de precificação (34) para cada

---

64. Ver, por exemplo, os artigos pioneiros de Iwata (1974), Gollop e Roberts (1979), Roberts (1984), Suslow (1986) e Bresnahan (1987), além dos trabalhos mencionados no texto.

firma, de forma a obter estimativas livres dos parâmetros de demanda  $\alpha$  (e portanto das elasticidades), custos  $\beta$  e conduta  $\theta_{it}$ . A estimativa livre de  $\theta_{it}$  pode ser interpretada como uma medida de quão próximo está o equilíbrio observado das previsões teóricas associadas a hipóteses de conduta alternativas. Por exemplo, uma estimativa de  $\theta_{it}$  próxima de 0 é interpretada como evidência de que a indústria tem comportamento que tende a competitivo. Na prática, apesar da especificação (34) permitir a variação de conduta ao longo do tempo, o parâmetro de conduta é normalmente restringido a um só valor médio ao longo do tempo (ou seja,  $\theta_{it} = \theta_{1i} \forall t$ ), ou a um número limitado de “quebras” (por exemplo,  $\theta_{it} = \theta_{1i}$  se  $0 \leq t \leq T$  e  $\theta_{it} = \theta_{2i}$  se  $t > T$ ), coincidentes com eventos relevantes à indústria.<sup>65</sup> Adicionalmente, é bastante comum ver o sistema especificado de forma agregada, no nível de indústria, por não se dispor de dados no nível da firma — omite-se portanto o subscrito  $i$  para firma na equação de especificação (34),  $C_i(\cdot)$  é então o custo marginal médio na indústria,<sup>66</sup> e o parâmetro de conduta  $\theta$ , recebe a interpretação do “grau médio de conluio na conduta” (*average collusiveness of conduct*) [Bresnahan (1989)].

Tal método procura identificar — como veremos na Subseção 1.3.1.2 — os parâmetros estruturais  $\alpha$ ,  $\beta$  e  $\theta$  a partir de dados observados de preços e quantidades, na medida em que o equilíbrio de mercado é deslocado de um ponto para outro por choques exógenos de demanda ( $Y$ ) e de oferta ( $W, Z$ ), observados também pelo econometrista. Apesar de as condições nas quais se identifica o modelo estrutural serem conhecidas, o “esforço” que se exige dos dados não é trivial [Gasmu, Laffont e Vuong (1990)]. Os dados estão ao mesmo tempo informando sobre demanda, custos e conduta. Obviamente o econometrista precisará fazer escolhas de forma a proceder à estimação paramétrica, escolhendo formas funcionais apropriadas para demanda e para custos marginais. Por exemplo, nas indústrias de açúcar [Genesove e Mullin (1998)] e de cimento [Salvo (2004)], o conhecimento institucional da indústria informará ao analista que, por ser a tecnologia de coeficientes fixos, o custo marginal poderá ser modelado como linear nos preços de fatores e constante em quantidade (até a capacidade das instalações). É claro que a consistência das estimativas dependerá sensivelmente da escolha de formas fun-

65. No estudo do cartel ferroviário de Porter (1983), tais quebras correspondem ao congelamento dos grandes lagos no inverno, impossibilitando o transporte hídrico. Ver também comentários abaixo sobre a aplicação de Genesove e Mullin (1998).

66. É fácil ver a partir de (34) que quando conduta corresponde a concorrência perfeita ( $\theta_{it} = 0$ ) ou a conluio perfeito ( $\theta_{it} = 1$ ),  $C_i(\cdot)$  é tanto o custo médio aritmético ( $\sum_i C_i(q_{it})/n$ ) para ver isso, somar as equações de especificação de todas as  $n$  firmas, dividindo então por  $n$ ) como o custo médio ponderado pela participação  $\sum_i s_{it} C_i(q_{it})$  (multiplicar a equação de especificação de cada firma  $i$  por  $s_{it}$  e somar todas as  $n$  equações resultantes), dada a simetria de custos marginais entre firmas implícita no *equilíbrio*. Já quando conduta na indústria corresponde a Cournot ( $\theta_{it} = s_{it}$ ), a primeira dessas manipulações algébricas (ou seja,  $C_i(\cdot)$  sendo o custo médio aritmético) gera o parâmetro médio de conduta ( $\sum_i \theta_{it}/n = (\sum_i s_{it})/n = 1/n$ , enquanto a segunda manipulação algébrica ( $C_i(\cdot)$  sendo o custo médio ponderado pela participação) gera o parâmetro médio de conduta  $\sum_i s_{it} \theta_{it} = \sum_i s^2_{it} = HHI_t$  (o índice de Herfindahl-Hirschman).

cionais. Para tanto, como veremos posteriormente, é importante incluir testes de robustez — por exemplo, estimar o lado da demanda usando formas funcionais alternativas, como linear, log-linear ou semi-log-linear, para testar a robustez das elasticidades estimadas — (ver a Seção 1.5 para uma discussão mais detalhada) — , assim como testes de especificação — utilizar um período em que conduta, ou custos, sejam conhecidos, comparando o valor conhecido ao valor estimado, como fazem Parker e Röller (1997) usando o período inicial de monopólio em telefonia celular nos Estados Unidos).

Uma variante desse método consiste em especificar uma equação adicional onde o parâmetro de conduta, em vez de ser estimado de forma livre, é “explicado” em termos de características observadas sobre a estrutura de mercado (o grau de concentração, por exemplo). Assim, utilizando um painel de mercados locais de telefonia móvel nos Estados Unidos, Parker e Röller (1997) incluem uma equação que expressa o parâmetro  $\theta$ , em termos do índice Herfindahl-Hirschman (HHI) e um índice que procura refletir o quanto freqüentemente as firmas que competem em um dado mercado local se encontram nos outros mercados.<sup>67, 68</sup> Vale mencionar o potencial problema de endogeneidade resultante dessa especificação, além das dificuldades citadas anteriormente. A conduta irá depender da estrutura de mercado assim como a estrutura de mercado irá depender da conduta. Os autores (implicitamente) assumem que a utilização de dados de uma mesma indústria e portanto um mesmo arranjo institucional, com variação entre mercados locais (variação na *cross-section*), os efeitos desse problema de endogeneidade — tão típicos de antigos estudos dentro do paradigma estrutura-conduta-performance de Bain — serão mitigados.

Por fim, vale ressaltar uma crítica que tem sido dirigida a essa metodologia empírica, na mesma linha ao que aludimos na nota 62. A teoria de oligopólio até a data presente não consubstancia um intervalo contínuo para os valores do parâmetro de conduta  $\theta_i$ , o que, de certa forma, não justifica sua estimativa livre. Citamos Reiss e Wolak (2002): “Até hoje, há muito pouca teoria econômica para guiar modelos estruturais a respeito de como  $\theta_i$  varia entre empresas. As especificações adotadas mais amplamente são *ad hoc*, com  $\theta_i$  variando de acordo com a produção da firma, sua participação de mercado ou o *ranking* da firma em termos de tamanho” (p. 57; tradução livre).

---

67. Ver Bernheim e Whinston (1990) sobre como o “ contato multimercado ” pode aumentar a sustentabilidade do conluio.

68. Outro exemplo é dado por Genesove e Mullin (1998), que a certa altura expressam  $\theta$  em função da participação (em termos de capacidade) da maior firma na indústria; ver a nota 73 adiante.

### 1.3.1.2 Identificação

Bresnahan (1982) analisa a identificação econométrica do modelo estrutural, com atenção particular ao parâmetro de conduta (o “conceito de solução do oligopólio”). Na versão mais simples, porém não menos geral, a questão que se coloca é se as observações de preços, quantidades, e das variáveis exógenas de demanda ( $Y$ ) e de oferta ( $W, Z$ ), podem distinguir entre duas hipóteses alternativas: a hipótese de que os dados observados são “gerados” por um cartel com custo baixo, contra a hipótese de que os dados correspondem a uma indústria competitiva com custo elevado.

Como vimos na seção sobre estimação de demanda, a curva de demanda (31) é identificada através dos deslocadores de oferta, ou seja, as variáveis exógenas  $W$  e  $Z$  são excluídas da função de demanda. Similarmente, a relação de oferta (34) é identificada através dos deslocadores de demanda, ou seja, a variável exógena  $Y$  é excluída da relação de oferta. Resta saber, porém, se a relação de oferta identificada corresponde: *a)* à relação de oferta de uma indústria competitiva de alto custo, traçando, portanto, a relação “preço igual a custo marginal” ( $\theta = 0$  em (34)); ou *b)* à relação de oferta de um cartel (com comportamento de monopólio) de baixo custo, traçando, portanto, a relação “receita marginal agregada igual a custo marginal” ( $\theta = 1$  em (34)). Bresnahan demonstra que a hipótese de conduta e de custos pode, sim, ser identificada pela flutuação da curva de demanda. No caso mais simples, quando o custo marginal é constante em quantidade, meras translações da curva de demanda bastarão para empiricamente distinguir entre as hipóteses de conduta: partindo de um ponto de equilíbrio, na medida em que a curva de demanda se translada, o novo ponto de equilíbrio segundo a hipótese comportamental de concorrência será distinto do novo ponto de equilíbrio segundo a hipótese comportamental de cartel. No caso mais geral de custos marginais que variam em função de quantidade, rotações da curva de demanda empiricamente possibilitam a distinção entre hipóteses de conduta.<sup>69</sup> Tais “rotadores de demanda” são naturais na medida em que, por exemplo, o tamanho de um mercado varia exogenamente, podendo ser implementados pela inclusão da interação entre variáveis exógenas  $Y$  e o regressor endógeno na especificação de (31) quando da estimativa de uma função linear, ou já figuram automaticamente pela inclusão de  $Y$  quando a função de demanda é não-linear.<sup>70</sup>

69. Intuitivamente, ao passo que uma rotação da curva de demanda não altera o equilíbrio competitivo, essa rotação irá deslocar o ponto de equilíbrio na presença de poder de mercado, por alterar a receita marginal [pois a inclinação da curva (inversa) de demanda  $P_1(Q_r)$  em (34) é alterada].

70. Ver, também, Lau (1982). Econometricamente, a identificação da relação de oferta decorre da premissa identificadora de ortogonalidade entre as variáveis exógenas de demanda  $Y$ , excluídas da equação de especificação (34), e o erro desta equação de especificação  $\epsilon^*$  (ou seja,  $E(Y\epsilon^*) = 0$ ).

A identificação dos parâmetros de conduta  $\theta$  em uma indústria de bens diferenciados é estudada por Nevo (1998). Ele confirma a generalidade do resultado de Bresnahan (1982): em princípio os parâmetros de conduta podem ser identificados junto com os parâmetros de demanda e de custo. Na prática, porém, Nevo alerta para a dificuldade de se dispor de quantidade suficiente de variáveis exógenas de demanda. Devido ao maior número de parâmetros de conduta em uma indústria de bens diferenciados, a identificação de conduta na ausência de dados sobre custo (ou lucros) fica comprometida devido às restrições de exclusão que dificilmente serão satisfeitas. Enquanto o número mínimo de variáveis endógenas em cada equação de precificação é 1 em uma indústria de bens homogêneos — o número mínimo correspondendo ao caso mais simples de custos marginais constantes em quantidade<sup>71</sup> —, este número mínimo em uma indústria de bens diferenciados sobe para o número de bens — novamente o número mínimo correspondendo ao caso de custos marginais constantes em quantidade — (veja-se a demonstração no próximo parágrafo). Como solução para o problema de inferir conduta juntamente com os parâmetros de demanda e de custo, Nevo propõe que seja utilizada a metodologia de seleção de modelos, apresentada na Subseção 1.3.1.4.

*Demonstração do número mínimo de variáveis endógenas de ordem J em uma indústria de bens diferenciados.* Considere uma indústria com  $J$  bens diferenciados, de propriedade de  $F$  firmas, onde  $F \leq J$  (ou seja, firmas podem ser proprietárias de mais do que uma marca). Suponha-se, para fixar idéias, que o custo marginal é invariante à quantidade; caso contrário, o número de variáveis endógenas em cada equação de precificação será ainda maior. Escreva o conjunto de marcas  $j \in \{1, 2, \dots, J\}$  de propriedade da firma  $f$  como  $O_f$ ; ou seja, se a firma 1 é dona das marcas 1, 2 e 4, então  $O_1 = \{1, 2, 4\}$ . Inicialmente supondo que o modelo de conduta seja o de Bertrand-Nash, a CPO é dada por uma generalização da equação (3) apresentada na introdução:

$$D_j(p) + \sum_{k \in O_f} (p_k - c_k) \frac{\partial D_k(p)}{\partial p_j} = 0, \quad \forall j \in O_f \quad (37)$$

Podemos reescrever o sistema de CPO (37) usando notação matricial, onde  $D(p)$ ,  $p$  e  $c$  são vetores com dimensão  $J$ , e portanto  $\nabla D(p)$  é uma matriz de

71. A variável  $\frac{P_i}{E}$  na equação (35), por exemplo, é endógena. A depender da especificação do custo marginal  $c_i = C_i(q_i, \cdot)$ , pode haver outras variáveis endógenas — se o custo marginal variar em função da quantidade.

dimensão  $J \times J$  com elemento  $\nabla D_{jk}(p) = \frac{\partial D_j(p)}{\partial p_k}$ . Definimos uma matriz de propriedade (*ownership matrix*)  $\Theta^B$  (o sobrescrito B denotando a conduta Bertrand), com elemento

$$\Theta_{jk}^B = \begin{cases} 1 & \text{se } \exists f \mid \{j, k\} \subseteq O_f \\ 0 & \text{caso contrário} \end{cases}$$

Por exemplo, se há duas firmas na indústria ( $F = 2$ ) e quatro marcas ( $J = 4$ ), a firma 1 sendo proprietária das marcas 1, 2 e 4, e a firma 2 proprietária da marca 3, então temos que

$$\Theta^B = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

O sistema de CPO segundo o modelo de conduta Bertrand é então dado por

$$D(p) + \Theta^B \cdot \times \cdot (\nabla D(p))' (p - c) = 0$$

onde o operador “ $\cdot \times \cdot$ ” denota a multiplicação matricial elemento-por-elemento e o operador “ $\cdot'$ ” denota a transposição matricial. A equação de especificação é dada então por:

$$p = c + \left[ -\Theta^B \cdot \times \cdot (\nabla D(p))' \right]^{-1} D(p) \quad (38)$$

Mais geralmente, essa especificação pode englobar diferentes modelos de conduta, alterando-se a matriz de propriedade (ou matriz de *conduta*). Por exemplo,

em um modelo de pleno conluio, a matriz de conduta conterá o valor 1 em cada elemento; em nossa indústria ela é dada por

$$\Theta^M = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Observe que nesse modelo de cartel (monopólio), cada firma, ao estabelecer o preço de uma marca, internaliza o efeito resultante sobre todas as marcas no mercado, sejam elas de sua propriedade ou não.<sup>72</sup> No caso mais geral onde se deseja inferir conduta a partir de uma matriz de parâmetros livres a ser estimada em (38), observe que há  $J$  parâmetros de conduta na equação de precificação de cada marca — isso decorre dos  $J$  elementos em cada fileira da matriz de conduta  $\Theta$ .

### 1.3.1.3 Ilustração 1: estimação de conduta como parâmetro livre

Consideramos o estudo de Genesove e Mullin (1998), sobre a indústria de refino de açúcar nos Estados Unidos na virada do século XX, assim como fizemos ao ilustrarmos instrumentos de demanda na Subseção 1.2.5.2. Os autores estimam um modelo estrutural estático composto por uma função de demanda (31) e uma equação de precificação (35) [equivalente a (34)] para a indústria como um todo (suprimindo portanto o subscrito  $i$ , como já mencionamos). Parâmetros de demanda, de custo e de conduta são estimados. Como mencionado nas Seções 1.2 e 1.5, diferentes formas paramétricas para a função de demanda são empregadas (confirmando a robustez dos resultados), e conhecimento institucional implica a escolha da forma paramétrica do custo marginal como sendo linear nos preços de fatores e constante em quantidade (até a capacidade das instalações). Em princípio, a sazonalidade do consumo de açúcar (o mercado cresce exogenousmente no verão) identifica a conduta.

Os autores alegam que seu estudo serve como um teste da estimação conjunta de demanda, custos e conduta — o que tem dado um caráter seminal ao trabalho — pelo seguinte motivo. Dadas a tecnologia de coeficientes fixos na produção de

---

72. Um terceiro exemplo de modelo de conduta é dado pelo modelo de Bertrand produto unitário (*single-product Bertrand*) onde uma firma detentora de múltiplas marcas estabelece preços de cada marca individualmente, ou seja, nem sequer internalizando o efeito de uma variação no preço de dada marca sobre a demanda por outras marcas de sua propriedade. A matriz de conduta, nesse modelo, seria então a matriz identidade de dimensão  $J, J$ .

açúcar refinado e a observação do preço de açúcar bruto importado (a principal matéria-prima), os autores estão na feliz posição de poder diretamente calcular o custo marginal com elevada acurácia. Esse custo marginal diretamente calculado pode então ser utilizado para *calcular* o parâmetro de conduta através do índice de Lerner ajustado pela elasticidade-preço de demanda estimada, conforme (36). Dessa maneira, o custo marginal e a conduta *diretamente calculados* podem ser comparados ao custo marginal e a conduta *estimados* na estimação conjunta de (31) e (35), onde não se assume o conhecimento de custos. Como a distância entre os parâmetros estimados e os parâmetros calculados é modesta, os autores concluem que em sua aplicação à indústria de açúcar “a metodologia (da nova organização industrial empírica) tem um desempenho razoavelmente satisfatório (*performs reasonably well*) ao estimar  $\theta$  (o parâmetro de conduta)” (p. 370; tradução livre exceto palavras em itálico; outros parênteses adicionados). Ou seja, os autores interpretam seus resultados como evidência a favor da estimação empírica do poder de mercado em uma indústria, sem o conhecimento de custos.<sup>73</sup> Genesove e Mullin também analisam *a*) a estimação de conduta, ao se observar custos somente parcialmente (ou seja, alguns parâmetros da especificação de custos são restringidos aos valores observados), e *b*) a estimação de custos, ao se assumir o modelo de conduta, na linha de uma segunda metodologia à qual voltamos agora nossa atenção.<sup>74</sup>

#### 1.3.1.4 Estimação 2: seleção estatística de modelos

Gasmi, Laffont e Vuong (1990) propõem uma segunda metodologia onde o pesquisador, dado o seu conhecimento institucional da indústria e sua análise de fatos estilizados, parte de um conjunto de modelos alternativos de conduta, de dimensão finita, que são plausíveis *a priori*. Por exemplo, esse conjunto de modelos, ou “família de formulações”, poderia conter  $K$  hipóteses alternativas de conduta, sendo dado por  $\Theta := \{\theta^1, \theta^2, \dots, \theta^K\}$ . Tomando uma hipótese candidata a cada

73. Aliás Genesove e Mullin (1998) relatam que sua estimação conjunta de demanda, custos e conduta *subestima* o poder de mercado, ao passo que *superestima* o custo marginal na indústria de açúcar refinado. Salvo (2004) argumenta que esses vieses podem não ser coincidentes. Ele demonstra que em indústrias que estão sujeitas à ameaça de entrada (por exemplo, de produto importado) e onde essa ameaça restringe preços pré-entrada, a conduta poderá não ser identificável pela estatística comparativa. Em tais situações, a premissa identificadora convencional de ortogonalidade entre as flutuações de demanda e o resíduo da equação de precificação (34) não é satisfeita, gerando inconsistência nas estimativas. A observação de custos pelo analista poderá ser necessária para se inferir conduta. De fato, Genesove e Mullin (1998) admitem que os preços de açúcar refinado na época estavam pressionados para baixo pela concorrência externa (potencial, pois não se observava importação em equilíbrio), o que sugere, portanto, a possibilidade de inconsistência nas estimativas. Nessa mesma linha, Corts (1999) mostra que a estimativa do parâmetro de conduta poderá medir o poder de mercado incorretamente em um modelo dinâmico de oligopólio.

74. Ilustrando outras colocações que fizemos no item 1.3.1.1, na seção final do artigo Genesove e Mullin (1998) também permitem a) que o parâmetro de conduta varie entre períodos caracterizados por “guerra de preços” e períodos “normais”; ou b) que o parâmetro de conduta seja função linear da participação (em termos de capacidade) da maior firma na indústria. Em relação a (b), os autores tomam o cuidado de alertar para a natureza descritiva, em vez de estrutural, dessa regressão, pelos motivos já expostos.

vez (digamos  $\theta^k$ ), estima-se o modelo estrutural dado pela equação de demanda (31) e por equações de especificação (34), ou CPO equivalentes caso o modelo de conduta não esteja aninhado em (34), com  $\theta = \theta^k$ , computando-se a verossimilhança de que tal modelo teria gerado os dados observados. Cada modelo é então testado em relação a cada modelo alternativo usando-se testes de verossimilhança relativa para modelos não-aninhados [Vuong (1989)], conforme exposto no Apêndice.

Em vez de basear a inferência do padrão de conduta na “distância” do parâmetro estimado livremente  $\hat{\theta}$  em relação a diferentes valores (discretos) vinculados à teoria de oligopólio, esse método seleciona o melhor modelo de  $\Theta$  (os modelos de conduta considerados pelo analista) com base na qualidade de ajustamento (*goodness-of-fit*) dos dados. Tal método, conhecido também como o “método do menu” (*menu approach*), oferece a vantagem de vincular a estimativa diretamente à teoria [Nevo (1998)]. Ou seja, são os dados que decidem diretamente pelo modelo que melhor os explica. É evidente que o sucesso da inferência dependerá da qualidade da seleção prévia dos elementos em  $\Theta$  pelo analista. No entanto, Gasmi, Laffont e Vuong (1990) enfatizam “a eficiência adicional ao incorporar informações prévias através de restrições decorrentes de estruturas de mercado específicas” (p. 519; tradução livre).<sup>75</sup> Em outras palavras, exige-se menos dos dados ao *supor*, a cada estimativa, um modelo-candidato alternativo de conduta, restando a estimativa dos parâmetros de demanda  $\alpha$  e de oferta  $\beta$ , no lugar de ter de estimar os parâmetros de demanda, de oferta e de conduta juntos. Adicionalmente, na inferência de indústrias de bens diferenciados, o método requer uma quantidade menor de restrições de exclusão, conforme apontado por Nevo (1998).

### 1.3.1.5 Ilustração 2: seleção estatística de modelos

Gasmi, Laffont e Vuong (1992) ilustram a metodologia ao estudar o padrão de comportamento em um modelo dinâmico aplicado à indústria de refrigerantes (especificamente ao submercado de sabor cola) nos Estados Unidos entre 1968 e 1986 [ver também Gasmi e Vuong (1991)]. Eles consideraram um jogo repetido entre firmas duopolistas (Coca-Cola e Pepsicola) nas variáveis estratégicas de investimentos em mídia e de preços. A demanda pela cola de uma firma é função do preço próprio, do preço cruzado, da mídia própria e da mídia cruzada. Entre os

---

75. A análise dos autores sugere que a estimativa livre de conduta através da equação de especificação (34) — que aninha vários padrões de conduta e pode, portanto, ser descrita como um “modelo mais amplo” — pode apresentar estimativas não-confiáveis: “diante da qualidade dos dados geralmente utilizados em trabalhos empíricos em organização industrial, e devido ao número elevado de parâmetros, nossa experiência demonstra que a estimativa do modelo mais amplo freqüentemente produz estimativas não-confiáveis dos parâmetros” (p. 518; tradução livre).

modelos de conduta considerados (ou seja, os elementos de  $\Theta$ ) estão: *a)* ambas as firmas estipulam preços e mídia simultaneamente segundo o padrão de comportamento Nash (ou seja, cada firma toma o preço e o nível de mídia de sua rival como dados ao estabelecer seu preço e sua mídia); *b)* modelos em que uma das firmas atua como líder de Stackelberg ao estabelecer preço e mídia, a rival atuando como seguidora; *c)* conluio (tácito) em que as firmas maximizam uma média ponderada de seus respectivos lucros; *d)* conluio em mídia e concorrência em preços; e *e)* conluio em mídia e coordenação em preços. Além disso, os autores consideraram uma mudança entre modelos supostamente ocorrida em 1976, aproximadamente na metade do período considerado (caracterizando uma “mudança entre regimes”, ou *regime switch*). Em uma primeira análise, ao estimar cada modelo-candidato e efetuar testes de seleção entre modelos, verifica-se que os modelos que incorporam alguma forma de coordenação — candidatos (*c*) a (*e*) — têm desempenho melhor que modelos calcados somente em não-coordenação — candidatos (*a*) a (*b*). Eliminação consecutiva de modelos-candidatos sugere que o padrão de comportamento que melhor se ajusta aos dados é caracterizado pela Coca-Cola atuando como líder Stackelberg até 1976, com conluio em mídia e concorrência em preços a partir dessa data. As elasticidades de preço e de mídia calculadas a partir dos parâmetros estimados do modelo comportamental selecionado também parecem ser, segundo os autores, de qualidade superior às alternativas. A flexibilidade da metodologia de seleção estatística de modelos fica evidente a partir dessa ilustração, em que modelos não-aninhados são testados entre si.

#### **1.4 A EXPERIÊNCIA INTERNACIONAL NO ANTITRUSTE**

Nesta seção buscamos compatibilizar os resultados já descritos com as restrições impostas na prática aos SDCs, com especial ênfase aos Estados Unidos (DOJ/FTC) e Europa (EC).<sup>76</sup> Se, por um lado, os métodos abordados mostram sua adequação em inúmeros estudos acadêmicos, por outro, restrições de informações e tempo para análise impõem aos SDCs a necessidade de dispor de métodos que não demandem muito em termos de dados, e que sejam de fácil e rápida implementação. Assim, abordaremos um pequeno conjunto de métodos propostos na última década que, em troca de uma tratabilidade maior, pragmaticamente adotam hipóteses simplificadoras para efetuar estudos de antitruste.

De acordo com Ivaldi *et alii* (2003a), há pelo menos duas formas de ameaça à livre concorrência, supondo-se a inexistência de dominância de uma única firma. Apesar de bastante distintas em teoria, em termos práticos é muitas vezes difícil

---

76. Para uma análise de políticas de concorrência, ver Motta (2004), bem como a análise detalhada de vários casos.

distingui-las. A primeira forma é conhecida como colusão tácita,<sup>77</sup> que pode ser melhor entendida como o resultado de as firmas se comportarem como se houvesse uma única firma dominante no mercado. A segunda forma, às vezes chamada de rivalidade individual,<sup>78</sup> ocorre quando a concentração do mercado é alta o suficiente para fazer com que uma situação não-competitiva ocorra mesmo que nenhuma das firmas possa ser considerada dominante.

Motta (2004, seção 5.7) propõe duas questões-chave que um SDC deve investigar antes de autorizar uma fusão:<sup>79</sup>

1. Quais os efeitos unilaterais da fusão?

Ou seja, se as *firms envolvidas na fusão* irão aumentar os preços de forma significativa após a autorização da mesma. Em caso negativo, o SDC deve responder se

2. A fusão tende a tornar as condições mais propícias para a ocorrência de efeitos coordenados (pró-colusivos)?

Ou seja, se a fusão alteraria as condições de concorrência na *indústria* investigada a ponto de tornar a colusão (tácita ou explícita) entre as firmas mais provável.

Tal compartmentalização não é unânime, nem está imune a críticas. De acordo com a “EU Merger Regulation”, por exemplo, a primeira questão a ser investigada seria se a fusão resulta em dominância de uma única firma (*single-firm dominance*). Caso o SDC adote o padrão de bem-estar total (o somatório dos efeitos sobre o consumidor e sobre o produtor), em vez do bem-estar somente do consumidor, ao analisar a primeira questão, alguma ênfase poderá ser colocada na possível redução de custos fixos. Lembre que, por não afetar o preço de equilíbrio, um SDC que adota o padrão de bem-estar do consumidor tenderá a ignorar a geração de economias fixas [ver Farrell e Shapiro (1990)]. Quanto à segunda questão, observe que a existência de condições favoráveis ao conluio não implica que essa ação concertada irá necessariamente ocorrer; tipicamente o equilíbrio colusivo não será o único, podendo, por exemplo, permanecer o equilíbrio estático. Competirá ao SDC avaliar o peso que deverá atribuir à possibilidade de ação coordenada, advinda da análise da segunda questão.

77. Também aludida como dominância coletiva (*joint dominance*) ou interação por efeitos coordenados (*coordinated effects*). No Brasil, costuma-se usar a expressão “exercício coordenado de poder de mercado”.

78. Também aludida como interação por efeitos unilaterais (*unilateral effects*). No Brasil, costuma-se usar “exercício unilateral de poder de mercado”.

79. A esse respeito, vejam-se as recentes contribuições de Ivaldi *et alii*(2003a e b) para a Comissão Européia.

Em termos técnicos, o estudo de efeitos unilaterais consiste na comparação entre os equilíbrios não-cooperativos estáticos na indústria em questão pré e pós-fusão. Intuitivamente, supondo-se a ausência de eficiências (variáveis), espera-se que uma fusão entre firmas produtoras de marcas substitutas leve a um aumento de preços, resultando na diminuição do excedente do consumidor (assumindo também a ausência de impacto sobre a inovação). No entanto, se a fusão gerar eficiências ou promover a inovação, o efeito final sobre o bem-estar é ambíguo, pelo fato de o poder de mercado poder ser compensado pela redução de preços advinda dos ganhos de eficiência ou o aumento da utilidade decorrente da melhoria de qualidade. Alternativamente, o estudo de efeitos coordenados, cuja análise foge do escopo deste trabalho,<sup>80</sup> consiste na avaliação de como o ato de concentração poderá alterar as características da indústria no sentido de favorecer ou incentivar a adoção de algum equilíbrio cooperativo, em detrimento do equilíbrio não-cooperativo.

A forma mais direta de detectar o exercício de poder de mercado por parte das firmas envolvidas em uma fusão é investigar se elas terão como impor aumentos de preços após a fusão. Em tese, tal análise poderia ser feita com os métodos econométricos descritos anteriormente, mas por vezes torna-se difícil, se não inviável, a realização de tal análise, principalmente por falta de dados e de tempo para sua coleta e análise.<sup>81</sup> Adicionalmente, há fatores de difícil avaliação por parte do analista, tais como os ganhos de eficiência necessários para compensar o aumento de preços decorrente da fusão.<sup>82</sup>

Muito em função da dificuldade de tratamento, os SDC's, especialmente o DOJ americano, tendem a se concentrar no estudo dos efeitos unilaterais das fusões; apesar do interesse recente do DOJ<sup>83</sup> e, principalmente, da Comissão Européia sobre o assunto,<sup>84</sup> o fato de implementações empíricas do estudo de efeitos coordenados não estarem devidamente desenvolvidas é um fator limitante ao seu estudo.<sup>85</sup>

80. Ivaldi *et alii* (2003 b) é um interessante *survey* sobre tal assunto.

81. Tal observação traz à tona a necessidade, à qual nem sempre é dada a devida importância, de uma "generosidade" maior para com os analistas tanto em termos de prazo, como em termos de dados, para que estes tenham melhores condições de realizar seu trabalho.

82. A esse respeito, como notamos adiante, é possível estimar os ganhos de eficiência necessários para que a variação de preços seja nula, então comparando o valor das eficiências implicitamente obtidas com o apresentado de forma convincente pelas partes; evita-se, assim, o estabelecimento arbitrário de ganhos de eficiência.

83. Através de *tender offer* para o seu estudo, em 2003.

84. Tal interesse em efeitos coordenados de fusões recebeu recentemente atenção especial em função da fusão proposta entre as firmas Airtours e First Choice, no mercado britânico de pacotes turísticos — rejeitado pela Comissão de Competição Européia. A fusão criaria um "triopólio" simétrico, o qual, segundo a comissão, aumentaria a sustentabilidade do conluio.

85. Para uma contribuição recente à literatura teórica de efeitos coordenados com punições ótimas, ver Kühn (2004) e Davis, Huse e van Reenen (2005), para uma inovadora implementação empírica de efeitos coordenados.

### 1.4.1 Modelos calibrados de simulação de efeitos unilaterais

Uma importante característica institucional do SDC americano é o pouco tempo concedido à análise dos efeitos de uma fusão, o que em muito dificulta a implementação de métodos mais demandantes em termos de técnica e de dados. Tal fato levou à proposição de modelos parcimoniosos e de cálculo rápido, ainda que sob o risco de serem considerados demasiadamente simples.

#### 1.4.1.1 O modelo PCAIDS

Dentre os métodos parcimoniosos já referidos, um dos que têm recebido considerável atenção nos últimos anos é o chamado PCAIDS, proposto em Epstein e Rubinfeld (2002). A hipótese fundamental desse modelo é a proporcionalidade entre as elasticidades-preço cruzadas e os *market shares* das firmas. Isso permite o cálculo dos efeitos unilaterais de uma fusão necessitando-se apenas conhecer o valor de *market shares*, da *elasticidade-preço própria da indústria* e da *elasticidade-preço própria de um dos produtos*, para que então se obtenha a matriz de elasticidades.

A intuição econômica subjacente à hipótese de proporcionalidade é que a participação de mercado perdida por uma firma como consequência de um aumento de preços é absorvida pelas demais firmas em proporção às suas próprias participações de mercado. Segundo seus próprios propositores, a hipótese de proporcionalidade parece mais apropriada em mercados com diferenciação de produtos limitada — o que excluiria, por exemplo, mercados de automóveis e tecnologia em geral.

*Especificação da demanda.* Para melhor ilustrar o método, considere um mercado de produtos diferenciados com  $J$  firmas uniproduto agindo como competidores Nash-Bertrand. A função demanda (em termos de *shares*) é dada por:

$$s_j = \alpha_j + \sum_{k=1}^J \beta_{jk} \cdot \ln(p_k)$$

e pode ser escrita na forma matricial por:

$$s = A + B \cdot p$$

No modelo AIDS,<sup>86</sup> a matriz  $B$  deve satisfazer as propriedades de homogeneidade e *adding-up* — a primeira implicando que as *market shares* não se alteram em

---

86. Considerado em detalhe na Seção 1.2.

decorrência de um aumento uniforme nos preços de todos os produtos do mercado (e na renda do consumidor), e a segunda, implicando que as *market shares* somem um. Em conseqüência disso, é possível escrever as elasticidades como função dos *market shares* ( $s$ ), elementos da matriz  $B$ , e da elasticidade da indústria ( $\eta$ ),

$$\eta_{jj} = \frac{\beta_{jj}}{s_j} + s_j \cdot (1 + \eta) - 1$$

$$\eta_{jk} = \frac{\beta_{jk}}{s_j} + s_j \cdot (1 + \eta)$$

O modelo PCAIDS de Epstein e Rubinfeld (2002) impõe uma estrutura adicional sobre a matriz  $B$ , logo, sobre a matriz de elasticidades, a saber,

$$\beta_{jk} = -\frac{s_j}{1-s_j} \cdot \beta_{kk}$$

Supondo que um dos elementos da diagonal seja conhecido — o termo  $\beta_{11}$ , sem perda de generalidade — a relação entre ele e os demais termos da diagonal da matriz  $B$  é dada por:

$$\beta_{jj} = \frac{s_j}{1-s_1} \cdot \frac{1-s_j}{s_1} \cdot \beta_{11}$$

*Exemplo.* Suponha o caso  $J = 3$  para fixar idéias. Plugando-se os elementos fora da diagonal, obtém-se:

$$B = \begin{bmatrix} \beta_{11} & -\frac{s_1}{1-s_1} \cdot \beta_{22} & -\frac{s_1}{1-s_1} \cdot \beta_{33} \\ -\frac{s_2}{1-s_2} \cdot \beta_{11} & \beta_{22} & -\frac{s_2}{1-s_2} \cdot \beta_{33} \\ -\frac{s_3}{1-s_3} \cdot \beta_{11} & -\frac{s_3}{1-s_3} \cdot \beta_{22} & \beta_{33} \end{bmatrix}$$

usando-se a relação entre os elementos da diagonal, chega-se em:

$$B = \begin{bmatrix} \beta_{11} & -\frac{s_1}{1-s_1} \cdot \beta_{22} & -\frac{s_1}{1-s_1} \cdot \beta_{33} \\ -\frac{s_2}{1-s_2} \cdot \beta_{11} & \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_1} \cdot \beta_{11} & -\frac{s_2}{1-s_2} \cdot \beta_{33} \\ -\frac{s_3}{1-s_3} \cdot \beta_{11} & -\frac{s_3}{1-s_3} \cdot \beta_{22} & \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_1} \cdot \beta_{11} \end{bmatrix}$$

e, finalmente, a matriz  $B$  é obtida apenas como função dos observáveis (*market shares*) e de  $\beta_{11}$ :

$$B = \begin{bmatrix} \beta_{11} & -\frac{s_1}{1-s_1} \cdot \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_1} \cdot \beta_{11} & -\frac{s_1}{1-s_1} \cdot \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_1} \cdot \beta_{11} \\ -\frac{s_2}{1-s_2} \cdot \beta_{11} & \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_1} \cdot \beta_{11} & -\frac{s_2}{1-s_2} \cdot \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_1} \cdot \beta_{11} \\ -\frac{s_3}{1-s_3} \cdot \beta_{11} & -\frac{s_3}{1-s_3} \cdot \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_1} \cdot \beta_{11} & \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_1} \cdot \beta_{11} \end{bmatrix}$$

A matriz de elasticidades correspondente é então obtida como:

$$E = \begin{bmatrix} \frac{\beta_{11} + s_1 \eta^* - 1}{s_1} & -\frac{s_1}{1-s_1} \cdot \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_1} \cdot \frac{\beta_{11} + s_1 \eta^*}{s_1} & -\frac{s_1}{1-s_1} \cdot \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_1} \cdot \frac{\beta_{11} + s_1 \eta^*}{s_1} \\ -\frac{s_2}{1-s_2} \cdot \frac{\beta_{11} + s_2 \eta^*}{s_2} & \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_1} \cdot \frac{\beta_{11} + s_2 \eta^*}{s_2} - 1 & -\frac{s_2}{1-s_2} \cdot \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_2} \cdot \frac{\beta_{11} + s_2 \eta^*}{s_2} \\ -\frac{s_3}{1-s_3} \cdot \frac{\beta_{11} + s_3 \eta^*}{s_3} & -\frac{s_3}{1-s_3} \cdot \frac{s_2}{1-s_1} \cdot \frac{1-s_2}{s_3} \cdot \frac{\beta_{11} + s_3 \eta^*}{s_3} & \frac{s_3}{1-s_1} \cdot \frac{1-s_3}{s_1} \cdot \frac{\beta_{11} + s_3 \eta^*}{s_3} - 1 \end{bmatrix}$$

onde  $\eta^* := (1 + \eta)$ , o que demonstra a necessidade de se ter a elasticidade da indústria. ■

De posse da especificação da demanda, remetemos o leitor à Subseção 1.1.1, que descreve a obtenção dos equilíbrios pré e pós-fusão, ou mesmo ao Capítulo 2 deste livro.

#### 1.4.1.2 O modelo ALM

Revisitaremos brevemente o modelo *logit*, discutido em detalhe na Subseção 1.2.4, para então abordar o modelo ALM, proposto em Werden e Froeb (1994).

Tal como no caso do PCAIDS, a motivação desse modelo reside na sua parcimônia e tratabilidade, por gerar fórmulas analíticas das quantidades de interesse e ter um requisito pequeno de dados: apenas os *preços* (pré-fusão) e *market shares* das firmas são necessários.

*Especificação da demanda.* Werden e Froeb (1994) supõem que a utilidade condicional indireta do consumidor é dada por:

$$u_{ij} = \delta_j + \varepsilon_{ij}, \\ i = 1, \dots, I_t; j = 1, \dots, J; t = 1, \dots, T$$

onde  $\delta_j = \alpha_j - \beta p_j$ , o parâmetro  $\beta$  é suposto constante entre indivíduos e produtos, e o termo idiossincrático, que pode tanto ser interpretado por características como preferências individuais subjetivas não-observadas (pelo econometrista), segue a distribuição de valores extremos do tipo I, de forma que diferenças de qualidade entre produtos são capturadas por diferenças no termo de intercepto,  $\alpha_j$ . Como já vimos, o *market share* do produto  $j$  é dado por:

$$s_j = \frac{\exp(\alpha_j - \beta p_j)}{\sum_{k=1}^J \exp(\alpha_k - \beta p_k)}$$

Sem perda de generalidade, considere o bem  $J$  como sendo o bem exterior, e suponha que  $p_J = 0$ , de forma que sua utilidade seja constante, e denote por  $\bar{p}$  uma média ponderada (por exemplo, pelos *market shares* em valor) de preços pré-fusão. A elasticidade-preço da demanda da indústria (bens internos) é dada por:

$$\eta_I := \left. \frac{\partial s_I(\lambda p)}{\partial \lambda} \cdot \frac{\bar{p}}{s_I(p)} \right|_{\lambda=1} = \beta \cdot \bar{p} \cdot s_J(p)$$

onde  $s_j := 1 - s_I(p)$  reflete a soma dos *market shares* dos bens interiores, e  $\lambda$  é um número real cuja derivada é avaliada no ponto  $\lambda = 1$  [ver Werden e Froeb (1994), para detalhes]. O termo  $\eta_I$ , tomado por Werden e Froeb (1994) e Werden, Froeb e Tardiff (1996) como o primeiro elemento fundamental do modelo, mede a variação na demanda dos bens interiores em face de um aumento de seus preços por um fator (uniforme entre eles)  $\lambda$ .

Realizando cálculos análogos aos feitos em 1.2.4.4, obtém-se fórmulas para as elasticidades-preço próprias e cruzadas dos produtos do mercado, que são dadas por, respectivamente,

$$\eta_{jj} = -\beta \cdot p_j (1 - s_j)$$

$$\eta_{jk} = \beta \cdot p_k \cdot s_k$$

O segundo elemento fundamental do modelo são os *market shares* condicionados ao fato de um bem interior ter sido consumido, obtido como uma probabilidade condicional,

$$s_{j/I} := \frac{s_j}{s_I} = \frac{s_j}{1 - s_J} = \frac{s_j}{1 - \frac{\eta_I}{\beta \bar{p}}}$$

Resolvendo-se a equação acima para  $s_j$  e plugando-se nas fórmulas das elasticidades, obtém-se:

$$\eta_{jj} = -\beta \cdot p_j (1 - s_j) = -\frac{p_j \cdot [\beta \cdot \bar{p} \cdot (1 - s_{j/I}) + \eta_I \cdot s_{j/I}]}{\bar{p}}$$

$$\eta_{jk} = \beta \cdot p_k \cdot s_k = \frac{p_k \cdot s_{k/I} \cdot (\beta \cdot \bar{p} - \eta_I)}{\bar{p}}$$

Convém notar a necessidade de se impor  $\beta \cdot \bar{p} - \eta_I > 0$ . Caso isso não ocorra, as elasticidades cruzadas entre *qualsquer bens j e k* serão negativas, o que implicaria que os bens em um dado mercado são todos complementos.

É ainda importante notar que as variáveis a serem calibradas se resumem a  $\eta_I$ , que controla a substituibilidade entre os bens internos e o bem exterior, e  $\beta$ , que controla a substituibilidade entre os bens internos, enquanto se supõe que os preços e *market shares* são observados, e as elasticidades próprias e cruzadas podem ser computadas. Os parâmetros  $\beta$  e  $\eta_I$  são quantidades da *indústria*, e não da firma, podendo ser estimados, obtidos de estudos anteriores, ou calibrados para que os *markups* gerados pelo modelo aproximem os *markups* observados (supondo-se que estes sejam conhecidos, obtidos, ou estimáveis de alguma forma).

De posse do valor de  $\beta$ , volta-se à fórmula de *shares* e, tomando-se o logaritmo da razão entre  $s_j$  e  $s_J$  (e lembrando-se que  $p_J = 0$ ), tem-se:

$$\ln\left(\frac{s_j}{s_J}\right) = \ln\left(\frac{\exp(\alpha_j - \beta p_j)}{\exp(\alpha_I)}\right)$$

Resolvendo-se a equação acima para  $\alpha_j$  chega-se a:

$$\alpha_j = \alpha_I + \ln(s_j) - \ln(s_J) + \beta p_j, \quad j = 1, \dots, J-1$$

onde o valor de  $\alpha_j$  pode ser fixado como uma constante ( $\alpha_j = 0$ , sem perda de generalidade), fazendo com que as quantidades do lado direito da equação tenham sido todas previamente determinadas. Plugando-se  $s_j = \eta_I / \beta \bar{p}$  e  $s_j = s_{J/I} (1 - \eta_I / \beta \bar{p})$  na equação acima obtém-se:

$$\alpha_j = \alpha_I + \ln(s_{J/I}) - \ln\left(\frac{\beta \bar{p}}{\eta_I} - 1\right) + \beta p_j, \quad j = 1, \dots, J-1$$

Assim, pode-se obter  $\{\alpha_j\}_{j=1}^{J-1}$  analiticamente conhecendo-se preços, *shares*, o parâmetro  $\beta$ , e a elasticidade  $\eta_I$ , da indústria.

*Os equilíbrios pré e pós-fusão.* Tendo resolvido o lado da demanda, Werden e Froeb (1994) fazem uma série de hipóteses sobre o lado da oferta, a saber:

- antes da fusão, cada firma produz apenas um produto diferenciado;
- as firmas têm apenas custos marginais  $c_j$ , supostos constantes em quantidade, não tendo custos fixos;
- firmas competem em preços; e
- todas as demais características dos produtos são tomadas como fixas (ou seja, a fusão não pode estimular a entrada, nem o reposicionamento de produtos pelas firmas existentes).

Assim, recordando a expressão (2), a CPO de Nash-Bertrand pré-fusão é dada por:

$$\frac{p_j - c_j}{p_j} = \frac{1}{\eta_{jj}}, \quad j = 1, \dots, J-1$$

que, plugando-se o valor das elasticidades próprias, pode ser reescrita como:

$$p_j - c_j = \frac{\bar{p}}{\beta \cdot \bar{p} \cdot (1 - s_{j/I}) + \eta_I \cdot s_{j/I}}, \quad j = 1, \dots, J-1$$

implicando que a margem da firma é crescente no seu *market share* ( $s_{j/I}$ ) — lembre que  $\beta \cdot \bar{p} - \eta_I > 0$  — e decrescente tanto na substituibilidade entre os bens inteiros ( $\beta$ ), como na substituibilidade entre bens inteiros e o bem exterior ( $\eta_I$ ). Assim, a equação acima permite recuperar os custos marginais pré-fusão em termos de variáveis observadas ou previamente determinadas.

Para se obter a estimativa de preços para o equilíbrio pós-fusão, utilizam-se as estimativas de custo marginal pré-fusão (podendo-se ainda incluir estimativas de ganhos de eficiência). Suponha, sem perda de generalidade, que as firmas 1 e 2 se fundem, formando a firma  $m$ . A CPO da firma  $m$  é dada por:

$$p_1 - c_1 = p_2 - c_2 = \frac{\bar{p}}{\beta \cdot \bar{p} \cdot (1 - s_{m/I}) + \eta_I \cdot s_{m/I}}$$

onde  $s_{mI}$  é o *market share* combinado das firmas 1 e 2, pré-fusão.<sup>87</sup> Ou seja, a margem da firma é decrescente na substituibilidade entre os bens interiores ( $\beta$ ) e na substituibilidade entre bens interiores e o bem exterior ( $\eta_I$ ), e crescente no *market share* pré-fusão combinado das firmas 1 e 2 ( $s_{mI}$ ), em função de se ter  $\beta \cdot \bar{p} - \eta_I > 0$ .

#### 1.4.1.3 Análise crítica dos modelos calibrados de simulação

Os modelos abordados nesta seção são especialmente úteis quando se trabalha sob restrições de tempo e obtenção de dados, e aí residem sua maior virtude e sua maior fragilidade. Por serem simplificações de modelos estudados anteriormente, além das propriedades indesejáveis dos modelos originais, eles herdam características indesejáveis adicionais. Seu uso é recomendado apenas em estudos de caráter exploratório, ou quando restrições severas de tempo e dados forem observadas.

### 1.5 CONSIDERAÇÕES PRÁTICAS NA ESTIMAÇÃO ESTRUTURAL

Nesta seção oferecemos comentários de natureza prática relevantes ao analista que tem ao seu alcance dados sobre alguma indústria e deseja estimar a estrutura dessa indústria. Procuramos tecer a discussão da forma mais intuitiva possível.

#### 1.5.1 Requisitos de dados

Tipicamente — e deixando de lado por agora a existência de efeitos dinâmicos — identifica-se cada observação como refletindo o ponto de equilíbrio em um mercado estático. Observações podem corresponder: *a*) a um mesmo mercado geográfico ao longo do tempo, sendo que cada período no tempo é então tratado como uma realização distinta desse mercado (por exemplo, dados referentes ao Distrito Federal em janeiro de 2005 correspondem a uma realização de mercado, ao passo que dados referentes ao Distrito Federal em fevereiro de 2005 correspondem a outra realização); *b*) a diferentes mercados geográficos em um mesmo período no tempo (dados referentes a diferentes mercados geográficos são tratados como diferentes realizações de um mesmo processo de geração de dados ou de distintos processos com parâmetros em comum a serem estimados); ou *c*) a uma combinação dos dois anteriores. O valor de dados em painel — onde, por exemplo, dados referentes a diferentes mercados geográficos são coletados ao longo do tempo — provém da disponibilidade de observações tanto na dimensão temporal como na *cross-section*.

<sup>87</sup> Para uma discussão mais detalhada, incluindo aplicações e discussões referentes a excedente do consumidor e bem-estar, ver Werden e Froeb (1994) e Werden, Froeb e Tardiff (1996).

De maneira geral, quanto maior o número de observações à disposição do analista, menor será o erro-padrão das estimativas. No entanto, não é necessariamente verdade que mais dados (ou dados mais freqüentes) sejam sempre melhores. Por exemplo, Genesove e Mullin (1998) optam por agregar dados para uma periodicidade trimestral, apesar de terem dados semanais disponíveis, em função da alta correlação serial observada nos últimos, que precisaria ser tratada de forma satisfatória (ver a seguir). Além disso, uma premissa implícita em muitos trabalhos empíricos envolvendo estimação estrutural (apesar de poucas vezes explicitada pelo pesquisador) é que as observações ao longo do tempo são realizações de um mesmo processo de geração de dados, cujos parâmetros são invariantes no tempo. Tal premissa pode fazer sentido quando o período de tempo ao qual os dados correspondem é razoavelmente curto, digamos três a quatro anos. No entanto, quando o período de tempo for mais longo, ou a indústria sob estudo estiver em franca evolução (por exemplo, telefonia móvel), o analista deverá considerar a possibilidade de que esteja ocorrendo uma variação ou quebra estrutural, com mudança nos parâmetros estruturais, como, por exemplo, uma alteração nas preferências dos consumidores. Se tal possibilidade for concreta, o analista deverá incorporar a variação ou quebra estrutural em sua especificação, adicionando — para citar um exemplo simples — efeitos de tempo ou uma variável de tempo.

Como exposto acima, para efetuar a estimação estrutural o analista irá requerer séries de preços e quantidades, seja na dimensão temporal e/ou na dimensão da *cross-section*, junto com deslocadores exógenos de demanda (incluídos na especificação de demanda, mas tipicamente excluídos da especificação de oferta, utilizados, portanto, na identificação do lado de oferta) e deslocadores exógenos de oferta (incluídos na especificação de oferta, mas tipicamente excluídos da especificação de demanda, podendo, portanto, ser utilizados na identificação de demanda).

Ilustramos os requisitos de dados com referência à estimação de modelos discretos de demanda, pelo fato de estes modelos necessitarem de uma disponibilidade maior de dados (lembra que, nesses modelos, como o consumidor demanda características de produtos em vez de demandar produtos em si, são necessários dados referentes a características da indústria em estudo).

*Dados agregados de mercado (market level data).* O caso clássico do uso de dados de mercado para estimação de modelos discretos é Berry, Levinsohn e Pakes (1995). Os requisitos mínimos de dados são:

- Preços, quantidades vendidas e características (podendo incluir dados de publicidade) dos produtos no mercado.

Em termos da notação acima, tais informações nos fornecem  $p_j$ ,  $x_j$ , e  $q_j$  para todo produto  $j = 1, \dots, J$  dentro do mercado.

Para que se obtenham os *market shares* dos produtos, necessita-se do tamanho do mercado. No caso de automóveis, tal como em Berry, Levinsohn e Pakes (1995 e 2004), para o mercado americano, uma primeira aproximação é dada pelo número de unidades familiares (que não seria tão satisfatória para o caso brasileiro, pelo fato de o país ser bem mais pobre); para Bresnahan, Stern e Trajtenberg (1997), o tamanho do mercado é tomado como sendo o número de trabalhadores de escritório no país. Em resumo, tem-se que:

- Tamanho do mercado = dados + hipóteses.

De posse do tamanho do mercado, tem-se agora  $s_j$ ,  $j = 0, 1, \dots, J$ . Não é incomum que pesquisadores façam experimentos, alterando o tamanho do mercado a fim de checar a robustez dos resultados obtidos.

Opcionalmente, tal como de início feito em Nevo (2000 e 2001), podem-se incluir características demográficas (tamanho da unidade familiar, existência de crianças etc.) no modelo [variáveis  $z$ , em termos de notação adotada, por exemplo, em (29)]. Uma forma de fazê-lo é extrair realizações de dados do censo, o que tem uma interpretação não-paramétrica. Alternativamente, poder-se-ia estimar uma distribuição paramétrica sobre os dados disponíveis e, de posse das estimativas dos parâmetros, gerar variáveis aleatórias com essa mesma distribuição.

As principais vantagens dos dados em nível de mercado são duas:

- São mais fáceis de serem obtidos.
- O risco de viés de seleção é relativamente pequeno (em contraste com o que acontece com microdados — ver adiante).

As principais desvantagens giram em torno do tratamento menos satisfatório de heterogeneidade do consumidor:

- O fato de a estimação ser potencialmente mais difícil (há pouca informação sobre as características do consumidor que comprou determinado produto).
- O fato de se usar informações sobre a distribuição de, por exemplo, características demográficas, e não as características associadas a cada unidade familiar.

*Microdados — micro data.* No caso de microdados, os requisitos mínimos são:

- Preços e características (podendo incluir dados de publicidade) de todos os produtos no mercado.

- As escolhas feitas pelos consumidores.

Em termos da notação acima, tais informações nos fornecem  $p_{ij}$ ,  $x_{ij}$  e  $q_{ij}$  para todo produto  $j = 1, \dots, J$  dentro do mercado e consumidor  $i = 1, \dots, I$ .

Tal como no caso anterior, opcionalmente, pode-se incluir características demográficas de cada família no modelo.

As principais vantagens dos microdados referem-se ao tratamento de heterogeneidade:

- Oportunidade de avaliar em detalhe o impacto das características demográficas nas escolhas.
- Oportunidade de observar o consumidor/unidade familiar em vários períodos, tendo acesso a seu comportamento dinâmico.
- Em tese, a estimação é mais simples, pela maior informação fornecida (confronte-se com o caso anterior).

As principais desvantagens são:

- Dificuldade de conseguir os dados.
- Viés de seleção, ou seja, a possibilidade de a amostra considerada não ser representativa da população para a qual se deseja extrapolar o estudo (considere os efeitos de se modelar a demanda por automóveis para uma amostra de classe média alta e as consequências de seu uso para toda a população brasileira).
- Finalmente, a potencial inconsistência das estimativas, pela dificuldade de se controlar a endogeneidade.

Um comentário importante diz respeito à distribuição (ou cobertura) de produtos em mercados de produtos diferenciados. A situação ideal é aquela em que todas as marcas de interesse para o estudo encontrem-se sempre disponíveis no ponto de varejo no momento em que o consumidor toma sua decisão de compra. Afinal, é o que se assume ao procurar inferir a substituibilidade entre marcas. Na prática, poucas marcas estão disponíveis o tempo todo no ponto de varejo, seja porque os estoques acabaram no estabelecimento ou porque o estabelecimento opta por não distribuir aquela marca. É possível que os erros decorrentes de tal ocorrência se anulem no agregado, ao trabalharmos com dados agregados de mercado (*market level data*) ou ao agregarmos as decisões individuais de cada consumidor quando os dados em questão sejam microdados. Mas é pouco provável que tal distorção seja trivial quando da introdução de uma nova marca, em que sua

distribuição esteja em evolução ao longo de um período mais longo de tempo. Sempre que possível valerá a pena para o analista procurar obter informações a respeito de distribuição, incorporando-as à estimativa. Tais dados são coletados por agências de coleta de dados como a ACNielsen e a IRI<sup>88</sup> [para algumas aplicações ver Mariuzzo, Walsh e Whelan (2004) e Slade (2004)].

Vale mencionar ainda o cuidado que deverá ser tomado pelo analista quando a prática de descontos ao consumidor através de cupons ou outras vantagens for difundida em determinada indústria sob estudo (ainda que tal prática no Brasil seja tímida em comparação com, por exemplo, os Estados Unidos). O desafio que se apresenta aqui é que dados sobre cupons poderão não ser observados pelo analista, ou observados somente imperfeitamente: como então obter os preços líquidos pagos pelos consumidores? Vale também mencionar que os preços coletados por agências como ACNielsen podem corresponder não aos preços médios no período (ou seja, receita no período dividida por quantidade vendida no período) mas sim a preços praticados *no momento em que a coleta é feita* (por exemplo, uma vez ao mês, ou uma vez por semana), podendo esse preço variar substancialmente entre um momento de coleta e o próximo. Essa distorção não estará presente quando o analista estiver trabalhando com microdados.

### **1.5.2 Nível de agregação de dados**

Uma consideração prática importante diz respeito ao nível de agregação dos dados. A agregação pode se dar em várias dimensões: *a)* agregação no espaço geográfico; *b)* agregação por embalagem, versão ou sabor; *c)* agregação por canal de vendas (e, para cada canal de vendas, por estabelecimento); e *d)* agregação no tempo. A título de ilustração, um analista hipotético que estivesse examinando a indústria de bebidas carbonatadas (refrigerantes), com toda sorte de dados à sua disposição, poderia optar por um nível de (des)agregação tal que uma observação equivaleria a um vetor de preços, quantidades vendidas e características para cada sabor (laranja), para cada marca (Fanta), para cada tamanho de embalagem (garrafa familiar de 2 litros em PET), correspondente a vendas através de supermercados com cinco ou mais caixas (*checkouts*) no município de Campinas na semana de 5 de fevereiro de 2005. Em outro extremo, uma observação poderia ser definida como um vetor de preços, quantidades vendidas e características (médias) por marca (Fanta), para

---

88. Um ponto semelhante deve ser feito com respeito à tentativa pelo analista de obter dados sobre promoções realizadas por produtores e/ou varejistas junto a consumidores. A literatura em *marketing* demonstra claramente a eficácia de promoções em aumentar o volume de vendas. Apesar da teoria econômica não informar de maneira explícita como essa atividade promocional e publicitária deve ser modelada, faz sentido conceber que a posse pelo analista de dados sobre promoções poderá melhorar sensivelmente a acurácia das estimativas.

todos os sabores e tamanhos de embalagem, através de todos os canais de venda (varejistas *self-service*, varejistas tradicionais com balcão de vendas e bares), para todo o Brasil, no ano de 2005.

Naturalmente, quanto mais desagregados estiverem os dados, maior é o conjunto de possibilidades que o analista terá em mãos na hora de partir para a modelagem e a estimativa (desde que a coleta de dados de maneira mais desagregada não tenha introduzido maiores erros de medida). No entanto, a utilização dos dados na forma mais desagregada possível não necessariamente será a mais indicada. Em alguns casos, os dados em nível demasiadamente desagregado, em virtude do processo de coleta, poderão conter muito ruído. Será aconselhável que o analista então proceda à agregação dos dados. Mas qual será o nível ótimo de agregação? A resposta aqui dependerá de questões institucionais e muito provavelmente não haverá uma resposta definitiva, devendo o analista testar a robustez de seus resultados estimando, quando possível, dados com diferentes níveis de agregação. Uma boa precaução consiste em considerar se os níveis de agregação não ignoram importante variação que possa estar ocorrendo em níveis menos agregados. Voltando ao exemplo de refrigerantes, conhecimento institucional poderá sugerir que a estimativa de elasticidades-preço deverá utilizar dados suficientemente desagregados por embalagem e por canal de vendas. A sensibilidade a preços do consumidor de Guaraná Antarctica na embalagem familiar de 2 litros, vendida através de grandes supermercados, certamente será distinta da sensibilidade a preços do consumidor de Guaraná Antarctica na embalagem de consumo individual de 350 ml, vendida através de bares (assim como a distribuição dessa mesma marca poderá ser tradicionalmente mais forte no canal bar do que no canal varejo). Ao estimar uma equação de demanda, o analista poderá então utilizar: *a*) todos os dados, porém de forma desagregada por embalagem e canal de vendas, incluindo, portanto, efeitos fixos em sua especificação (assim como interações desses efeitos fixos com, por exemplo, preços); ou *b*) dados referentes a cada tamanho de embalagem e cada canal de vendas separadamente, estimando uma equação distinta para cada combinação de embalagem e canal. Similarmente, a agregação na dimensão geográfica deverá considerar a possibilidade de que as preferências de consumidores, assim como o comportamento das firmas servindo esses consumidores, poderão ser distintos em diferentes mercados geográficos. Por exemplo, por motivos institucionais, a penetração de bebidas dietéticas (por exemplo, Coca *Light*) é sabidamente maior na cidade do Rio de Janeiro do que em outras grandes cidades como São Paulo. De posse de dados em painel, um analista, estimando uma equação como a (15), deverá então permitir tal variação através da inclusão em sua especificação de

efeitos fixos de cidade-marca (e possivelmente interagindo esses efeitos com preços, sendo a identificação do coeficiente de preços dada, então, pela variação temporal).<sup>89</sup>

Quanto à agregação no tempo, pode haver um *trade-off* entre eficiência (quanto maior o número de observações, maior o número de graus de liberdade) e correlação serial: é concebível que um analista, disporo de observações com variação temporal semanal, opte por agregar os dados na dimensão temporal para períodos mensais. Apesar de estar reduzindo o número de observações em 3/4, a correlação serial entre observações mensais poderá ser menor do que entre observações semanais, podendo ser tratada de modo mais satisfatório, utilizando-se, por exemplo, o estimador de covariância de Newey e West (1987).

### 1.5.3 Escolha de variável dependente (regressão reversa)

Em algumas situações, particularmente na estimação de demanda por produtos homogêneos, o analista deve escolher não somente a forma funcional (ver subseção adiante) como também a variável dependente, a saber, preços ou quantidades — por exemplo, ao passo que Genesove e Mullin (1998) tomam quantidades como a variável dependente, Parker e Roller (1997) escolhem preços. Aqui, até onde sabemos, a literatura é escassa. A análise dos resíduos (*fitted residuals*) poderá dar subsídios ao analista com respeito a qual regressão é preferível. Vale ressaltar que a robustez das estimativas à escolha de determinada variável dependente pode ser verificada através da estimação da regressão reversa. Carlton (2003) comenta que “parece ser uma regularidade empírica que um (pesquisador) freqüentemente pode obter uma elasticidade alta ao simplesmente estimar preços em quantidades do que o inverso!” (p. 8; tradução livre).

---

89. Em diversos momentos é concebível que o analista trabalhe, mediante algumas premissas, em nível *menos* agregado do que o nível de agregação em que os dados estão disponíveis. Para citar um exemplo [Salvo (2005)], assuma que dados de demanda por cimento (preços, quantidades, atividade na construção civil etc.) estão disponíveis somente por *estado* (unidade da federação), mas o analista deseja trabalhar no nível de *município*. (Note que há em média 200 municípios por estado, totalizando aproximadamente 5.600 municípios no Brasil.) Uma possibilidade consiste em, inicialmente, estimar uma curva de demanda para cada estado a partir dos dados estaduais. Observe que a curva de demanda (inversa) de determinado estado é o resultado da soma horizontal das curvas de demanda (inversas) de todos os municípios localizados naquele estado. Agora assuma que o preço de reserva do consumidor que mais valoriza o produto (*o highest-valuation consumer*) em cada município não varia entre municípios naquele estado (ou seja, o intercepto no eixo de preço é o mesmo para a curva de demanda de cada município naquele estado). Assuma também que na medida em que o tamanho de mercado dos municípios (dado pelo tamanho de suas populações) varia, as curvas de demanda dos municípios alteram-se de maneira “multiplicativa”, ou seja, elas giram ao redor do intercepto de preço. A partir da curva de demanda (inversa) do estado *S* dada por, digamos,  $p = \alpha_S - \beta_S q$ , o analista pode obter uma estimativa da curva de demanda do município *M*, ou seja  $p = \alpha_M - \beta_M q$ , onde

$\alpha_M = \alpha_S$  e  $\beta_M = \frac{Y_S}{Y_M} \beta_S$ , sendo que  $Y_M$  denota a população do município *M* e  $Y_S := \sum_{M \in S} Y_M$  é a população total do estado *S*. É fácil verificar que (a um dado preço) a elasticidade-preço de demanda no nível do estado equivale à elasticidade-preço de demanda no nível do município.

### 1.5.4 Implementação prática

Tratamos da implementação prática de modelos discretos, visto que as versões mais sofisticadas exigem a programação de rotinas complexas por parte do analista.

#### 1.5.4.1 Implementação de modelos discretos<sup>90</sup>

Ao se implementar modelos discretos como o *logit*, a variável dependente usada é a diferença entre o logaritmo do *market share* de cada produto subtraído do logaritmo do *market share* do bem exterior,  $\ln(s_j) - \ln(s_0)$  (obtidos após a definição do mercado potencial,  $M$ ) e regressores mais usados — além do preço — são despesas com publicidade, características dos produtos estudados, variáveis *dummies* para marcas e períodos distintos. Para as versões clássicas do modelo *logit*, propriedades teóricas dos estimadores podem ser obtidas em Amemiya (1985), ou ainda em livros-texto como Wooldridge (2002).

Nosso tratamento se baseia principalmente no GMM. Para chegar à formulação do modelo, considere as fórmulas dos *market shares logit*,

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^J \exp(\delta_k)}, \quad j = 1, \dots, J$$

$$s_0 = \frac{1}{1 + \sum_{k=1}^J \exp(\delta_k)}$$

divida a primeira equação pela segunda, e tome o logaritmo para obter  $\ln(s_j) - \ln(s_0) = \delta_j$ . Recorde que no modelo *logit* com características não-observadas [Berry (1994)], tem-se  $\delta_j = x'_j \theta + \xi_j$  ( $\beta \subseteq \theta$ ), onde  $x_j$  engloba características observadas mais o preço do produto  $j$ . Nesse caso, a solução do sistema relacionando *shares* geradas pelo modelo a *shares* observadas é trivial, fornecendo o que interpretamos como sendo o termo de erro

$$\xi_j(\theta, x, s) = \ln(s_j) - \ln(s_0) - x'_j \theta$$

---

90. A maior parte dos estimadores discutidos nesta subseção encontra-se implementada no pacote econômétrico *Stata* — os modelos e os respectivos comandos são: *Logit* - *logit*; *Probit* - *probit*; *Multinomial logit* - *mlogit*; *Nested logit* - *nlogit*. Além disso, o comando para o método de variáveis instrumentais é o *ivreg*.

Uma vez ilustrado esse passo fundamental, recordamos e aprofundamos a discussão sobre aspectos teóricos do modelo, que valem independentemente da especificação do mesmo — usaremos o *logit* como exemplo pela sua tratabilidade.

Primeiro, vale notar que as propriedades assintóticas do modelo estão relacionadas ao número de produtos  $J$ , bem como ao tamanho da amostra de consumidores a partir da qual se calculam os *market shares*. Segundo, algumas das características observadas — tipicamente preço — podem estar correlacionadas com  $\xi$ , requerendo o uso de instrumentos. Terceiro, ressaltamos que os *shares*  $s(\theta, x, \xi, P)$  são gerados pela agregação das demandas individuais dos agentes (indivíduos, unidades familiares etc.) com distribuição de preferências  $P$ , por exemplo, a função distribuição empírica de uma amostra aleatória gerada pela distribuição verdadeira de preferências,  $P^0$ . Quarto, recordamos o sistema relacionando *market shares* observados e *market shares* gerados pelo modelo,  $s(\theta_0, x, \xi, P^0) = s^{obs}$ , onde  $\theta_0$  é o valor verdadeiro do vetor de parâmetros. A estratégia a ser seguida para estimação é exatamente aquela descrita na seção referente a Berry, Levinsohn e Pakes (1995), onde a minimização da distância entre *shares* observados e *shares* gerados pelo modelo,  $\|s(\theta_0, x, \xi, P^0) - s^{obs}\|$  é mostrada como não sendo factível (tanto em função das não-linearidades envolvidas, como pelo fato de os argumentos serem funções uns dos outros), levando à proposição da adoção do GMM [Berry (1994)].

Voltando-se ao caso *logit*, interprete-se o termo  $\xi_j$  como um erro de um modelo de regressão de *shares* observados em características observadas. Dada a possibilidade de correlação entre componentes de  $x$  e o erro  $\xi$ , considere a decomposição  $x := (x_1, x_2)$ , onde  $x_1$  são características de produtos exógenas, e  $x_2$  são características endógenas (o preço, para fixar idéias). A hipótese de identificação clássica é supor a existência de um conjunto de instrumentos  $z$ , tipicamente constituído por  $x_1$  (instrumentos incluídos) e instrumentos (excluídos) para  $x_2$  (preço), e a condição de momento amostral associada é dada por:

$$\begin{aligned} G_J(\theta, s, P^0) &= \frac{1}{J} \sum_{j=1}^J z_j \cdot \xi_j(\theta, x, s, P^0) \\ &= \frac{1}{J} \sum_{j=1}^J z_j \cdot [\ln(s_j) - \ln(s_0) - x'_j \theta] \end{aligned}$$

É interessante notar que o estimador acima pode ser melhor entendido como um estimador IV (variáveis instrumentais) linear, de acordo com os resultados de Berry (1994), podendo ser estimado usando-se 2SLS.

Em termos teóricos, a condição de momento resulta da hipótese de covariância nula entre o conjunto exógeno de instrumentos,  $z$ , e características não-observadas,  $\xi$ , via

$$E[G_J(\theta)] := E\left[\frac{1}{J} \sum_{j=1}^J z_j \cdot \xi_j(\theta, x, s, P^0)\right] = 0$$

em  $\theta = \theta_0$  (valor verdadeiro do vetor de parâmetros), e a estimativa de  $\theta$  é obtida pela minimização de uma norma da condição de momento amostral,

$$\frac{1}{J} \sum_{j=1}^J z_j \cdot \xi_j(\theta, x, s, P).$$

As hipóteses sobre  $\{\xi_j\}_{j=1}^J$  são muito semelhantes às feitas sobre o erro num modelo de regressão linear. Primeiro, supõe-se que sejam variáveis aleatórias independentes. Além disso, supõe-se que  $E[\xi | x_i] = 0$  e  $\sup_j E[\xi_j^2 | x_{1j}] < \infty$  com probabilidade igual a 1. Vale ressaltar que tais hipóteses permitem que tal processo exiba heterocedasticidade condicional.

Como mencionado anteriormente, nem sempre é possível avaliar os *market shares* gerados pelo modelo,  $s(x, \xi, \theta, P^0)$ , analiticamente, de forma que é preciso simular a distribuição  $P^0$ , calculando-se  $s(x, \xi, \theta, P^m)$ . O acréscimo na variância decorrente do erro de simulação pode, como mencionado, ser estimado por meio de um procedimento de *bootstrap*, tal como destacado na seção sobre BLP.<sup>91</sup>

### 1.5.5 Comparação e escolha de formas funcionais (demanda e oferta)

A imposição de uma forma funcional é um mal necessário em virtude do problema de dimensionalidade. Idealmente poderíamos utilizar métodos não-paramétricos caso dispuséssemos de quantidade suficiente de dados, mas na prática há de se adotar estrutura. No entanto, conforme discutido anteriormente, a premissa de forma funcional não deverá estar informando o padrão de substituição,<sup>92</sup> papel

91. Para uma discussão sobre *bootstrap*, ver Wooldridge (2002, cap. 12).

92. A título de ilustração, vale voltar à Subseção 2.4.4, onde é feita uma análise crítica do modelo *logit*, em que o segundo exemplo compara as elasticidades das especificações lineares e logarítmicas em preços. A lição, no caso do *logit*, é que as elasticidades são consequência direta da forma funcional empregada. Assim, sabe-se que um modelo linear em preços vai dar uma elasticidade mais alta que um modelo logarítmico *a priori*, para preços maiores que a unidade.

este que deve ser desempenhado pelos dados: nesse sentido, é desejável que o modelo de demanda seja suficientemente “flexível”.<sup>93</sup>

Vale lembrar que a robustez das estimativas à escolha da forma funcional pode (e deve) ser verificada através da estimação de diferentes formas funcionais para verificar a robustez. Por exemplo, na estimação de demanda por um produto homogêneo, Genesove e Mullin (1998) estimam quatro diferentes formas funcionais, a saber: linear, log-linear, exponencial e quadrática. Afortunadamente, as elasticidades-preço de demanda estimadas pelos autores segundo as quatro especificações são próximas, sugerindo que a escolha de forma funcional não está gerando a estimativa de elasticidade. Menos óbvia é a estratégia que o analista deverá seguir caso as estimativas dependam sensivelmente da escolha de forma funcional, e em que a análise dos resíduos não recomende uma forma funcional (ou um subconjunto destas) em detrimento de outras. Nesse caso, o analista deverá reportar o achado e possivelmente ampliar o intervalo de confiança de suas estimativas, levando em consideração um intervalo que englobe os intervalos estimados pelas diferentes formas funcionais.

Já em uma indústria de produtos diferenciados, Crooke *et alii* (1999) mostram como a escolha de forma funcional pode afetar sensivelmente a magnitude dos efeitos de um ato de concentração sobre preços. Através de experimentos de Monte Carlo, eles obtêm aumentos de preços consideravelmente maiores com funções de demanda de elasticidade constante do que com funções de demanda lineares ou *logit*.<sup>94</sup> Werden, Froeb e Tschantz (*forthcoming*) argumentam, à luz do Antitrust Logit Model (ALM), que o modelo *logit* tende a gerar previsões de aumentos de preços modestos em face de atos de concentração.<sup>95</sup>

Reproduzimos (além de elaborarmos) aqui alguns comentários feitos por Hosken *et alii* (2002) e Werden *et alii* (1994) a respeito das vantagens e desvantagens de diferentes formas funcionais de demanda (no que tange, particularmente, à estimação de demanda de produtos diferenciados):

- Sistema de demanda linear: A principal vantagem é a simplicidade no cálculo dos efeitos de uma fusão [ver Werden, Froeb e Tardiff (1996) para maiores

93. Uma excelente discussão sobre a escolha de formas funcionais na modelagem de demanda e a flexibilidade oferecida por diferentes modelos é dada por Deaton e Muellbauer (1980b).

94. Intuitivamente, na medida em que o preço aumenta, a demanda se torna mais elástica *mais rapidamente* nas funções de demanda linear e *logit*, em comparação com a função de demanda log-linear (onde a elasticidade permanece constante). Esse aumento rápido da elasticidade atua como um freio no aumento de preços decorrente da fusão.

95. Werden (1996) propõe uma maneira de analisar um ato de concentração que passa ao largo da escolha da forma funcional e, portanto, do efeito desta sobre a magnitude de variação de preços: esta análise consiste em computar a redução de custos marginais necessária para exatamente compensar o efeito unilateral da fusão sobre preços. Como esse cálculo envolve, por definição, preços pós-fusão iguais a preços pré-fusão, a forma como as elasticidades variam em função de preços — e portanto a escolha da forma funcional — não mais é importante.

detalhes]. Entre as suas desvantagens está a falta de garantia de que as elasticidades-preço cruzadas tenham o sinal “esperado”: por exemplo, a elasticidade-preço cruzada entre marcas substitutas (assumindo que o efeito renda seja baixo<sup>96</sup>) deveria ser positiva, mas nem sempre o é.

- Sistema de demanda log-linear (elasticidade constante): A vantagem desse sistema é que as elasticidades podem ser lidas imediatamente a partir dos coeficientes de preço estimados pela especificação. Entre as desvantagens, está *a*) a restrição de que as elasticidades permaneçam constantes — em vez de aumentar, o que seria mais natural no caso da elasticidade própria — na medida em que o preço aumenta; *b*) a inversão comum de sinais em relação ao esperado; e *c*) restrições provenientes de teoria econômica, como a soma das fatias de gastos do consumidor sendo igual a 1, não são necessariamente satisfeitas.

- Sistema de demanda AIDS: Entre suas vantagens está a importante propriedade de flexibilidade. Hosken *et alii* (2002) afirmam que “muitas das estimativas de sistemas de demanda apresentados às Agências (de antitruste, como FTC e DOJ) usam o sistema de escolha em estágios múltiplos com a especificação de demanda AIDS” (p. 26; tradução livre). Entre as desvantagens, está a necessidade de estimar um elevado número de parâmetros e, portanto, a necessidade de o analista dispor de uma base de dados satisfatória para manter a variância de suas estimativas dentro de um patamar aceitável (além da necessidade de o analista adotar premissas identificadoras que podem suscitar alguma polêmica em certas situações — ver a Subseção 1.2.5).<sup>97</sup>

- Sistema de demanda *logit* (MNL): Entre as vantagens, estão a facilidade de estimativa e a baixa exigência de dados (devido ao reduzido número de parâmetros). Como vimos, essa facilidade vem a um custo — a falta de flexibilidade: a propriedade IIA (independência das alternativas irrelevantes) pode ser irrealista. Lembrar que no modelo *logit*, por exemplo, se o preço de um produto sobe, os consumidores substituem-no por outros produtos em proporção aos *shares* desses outros produtos.<sup>98</sup> Isso decorre de a elasticidade-preço cruzada de demanda de cada produto  $j \neq i$  com relação ao preço do produto  $i$  ser a mesma — não levando em conta informações

96. Lembrar a equação de Slutsky.

97. Outra limitação de sistemas desse tipo (contínuos) é o tratamento da heterogeneidade individual. O analista poderá ter acesso a dados ao nível do consumidor (micródados) mas não fica clara a forma como esses dados devem ser incorporados ao sistema.

98. Em defesa do modelo *logit* (ALM), Werden *et alii* (1994) citam o argumento de Willig (1991) de que a propriedade IIA é justificada em uma situação em que as marcas, sendo fusionadas, *não* estão situadas a uma distância particularmente próxima ou distante no espaço de características. Werden e seus co-autores afirmam que, em virtude do papel central de *shares* no modelo *logit*, “o ALM serve, portanto, como um filtro comparável ao filtro oferecido pelos *market shares* na análise antitruste tradicional, sendo o ALM uma melhoria substancial por oferecer previsões quantitativas sobre aumentos de preços” (p. 9; tradução livre).

sobre a proximidade desses produtos no espaço de características. Note que como nesse modelo a substituição é proporcional aos *shares* de mercado, uma fusão entre marcas com *shares* elevados levará a um aumento de preços maior do que uma fusão entre marcas com *shares* pequenos. Além disso, vale relembrar que as elasticidades são consequência direta da forma funcional adotada para preços, como demonstrado na Subseção 1.2.4.4.

- Generalização do *logit* 1 — Sistema de demanda logit aninhado (NL): Ao passo que no modelo *logit* o componente aleatório de utilidade está distribuído de forma idêntica e independente entre consumidores, o que não possibilita a correlação de preferências por subgrupos de marcas, no *logit* aninhado marcas “mais próximas” são colocadas em um mesmo ninho pelo analista, permitindo que haja correlação de preferências por subgrupos de marcas. A vantagem do *logit* aninhado, portanto, é que quando o preço de determinada marca sobe, é possível que boa parte da substituição ocorra para outras marcas do mesmo ninho (a depender de um parâmetro que é estimado, indexando tanto a proximidade entre marcas de um mesmo ninho, quanto a distância entre as marcas desse e de outros ninhos). Assim, a fusão entre marcas de um mesmo ninho poderá ter um impacto muito maior em preços do que a fusão entre marcas de ninhos distintos. As desvantagens desse modelo, como já discutimos, são a arbitrariedade na definição de ninhos (contornável, em princípio, via teste de Vuong), a persistência da IIA intra-ramos, e o fato de elasticidades próprias exibirem os mesmos problemas das do modelo *logit* clássico.

- Generalização do *logit* 2 — Sistema de demanda PD-GEV (BST): Outra forma de adicionar flexibilidade ao modelo *logit* é especificar dimensões de diferenciação. Marcas com características parecidas estão mais próximas nesse espaço de diferenciação multidimensional. Assim como no *logit* aninhado, onde a proximidade entre marcas de um ninho é estimada, no BST a importância de cada dimensão de diferenciação em explicar as escolhas de consumidores também é estimada (através do parâmetro  $\rho$ ). Assim, o efeito sobre preços será maior em uma fusão envolvendo marcas com classificações próximas. Um aspecto positivo desse sistema em comparação com o *logit* aninhado, é que ele permite tratar potenciais segmentações de mercado diferentes de forma simétrica, ao passo que o *logit* aninhado impõe uma estrutura hierárquica. Isso torna o modelo menos passível de críticas em aplicações práticas. No entanto, não surpreende que o esforço de implementação ante o *logit* aninhado seja maior.

- Métrica de distância: Sua principal vantagem é a flexibilidade. No entanto, tal modelo tem um componente de arbitrariedade na definição de vizinhança,

além de haver críticas quanto a sua aplicação para a análise de bem-estar [ver Nevo (2004)].

- Generalização do *logit* 3 — Sistema de demanda *logit* com coeficientes aleatórios (*mixed logit*, BLP): Trata-se da versão mais geral e flexível do *logit*, que se acredita fornecer elasticidades mais condizentes com a intuição econômica. Pode ser aplicado tanto a dados agregados, como a microdados, mas seu custo de implementação e o tempo gasto no processo de estimação, tendem a ser bastante altos.

### 1.5.6 Testes de especificação

Uma prática crescente e sá na literatura empírica em organização industrial é aquela em que o analista reporta o resultado de testes de especificação. Em vez de oferecermos um tratamento comprehensivo desse teste de especificação, remetemos o leitor ao livro-texto de Wooldridge (2002).<sup>99</sup> Visto que o contrário foge ao escopo do presente trabalho, nos limitamos aqui a um breve comentário do objetivo de alguns desses testes:

- Testes de sobreidentificação [ver Wooldridge (2002, subseções 6.2.2 e 8.5.2)]: A exogeneidade das variáveis instrumentais quando o número destas é igual ao número de variáveis endógenas (ou seja, a equação está *exatamente* identificada) é por definição uma hipótese mantida, ou seja, não-testável. No entanto, quando a equação estiver sobreidentificada, dado um número de instrumentos superior ao número de variáveis endógenas, é possível que a redução da variância das estimativas, obtida através do aumento no número de condições de momento (cada instrumento representa uma condição adicional que tende a reduzir a variância das estimativas), esteja sendo obtida à custa de consistência. Este é mais um exemplo do *trade-off* recorrente em econometria entre eficiência e consistência dos estimadores. Daí decorre a importância de que se teste a sobreidentificação.<sup>100</sup>

- Testes de endogeneidade [ver Wooldridge (2002, subseção 6.2.1)]: Esses testes visam avaliar a correlação entre a variável supostamente endógena e o resíduo da estimação. Ignorar a endogeneidade dos preços pode gerar resultados tais como elasticidades-preço próprias de demanda inesperadamente baixas (demandas inelásticas) e, no limite, coeficientes associados ao preço positivos [Trajtenberg (1989)]. A endogeneidade dos preços pode ser modelada com o uso de técnicas econômétricas clássicas, como 2SLS — ver a seção sobre identificação de demanda.

99. Um tratamento comprehensivo, incluindo comandos do pacote econômético Stata, é oferecido por Baum, Schaffer e Stillman (2003).

100. Davidson e MacKinnon (1993) comentam que “testes de sobreidentificação deveriam ser realizados rotineiramente quando se computam estimativas por variáveis instrumentais” (p. 236; tradução livre).

A instrumentação dos preços tende a gerar elasticidades com magnitudes mais próximas do que se esperaria em um dado mercado. Adicionalmente, espera-se que o uso de variáveis demográficas — e o controle de heterogeneidade em geral, com o uso de efeitos fixos — traga melhores resultados.

- Heterocedasticidade e autocorrelação dos resíduos [ver Wooldridge (2002, seção 7.8.5)]: A análise dos resíduos estimados poderá informar sobre a validade das premissas adotadas pelo analista quanto à distribuição e correlação temporal nos erros.

### **1.5.7 Análise de sensibilidade e comparação entre modelos**

Dada a complexidade de muitos dos modelos de demanda considerados, nem sempre a análise das implicações das hipóteses feitas é imediata, sejam elas feitas sobre o modelo estrutural, ou de cunho estocástico. Assim, uma forma muito adotada na literatura para checar a razoabilidade dessas hipóteses e/ou comparar modelos alternativos é por meio de análises de sensibilidade e robustez. Casos típicos incluiriam, como já descrito, verificar se os valores das elasticidades cruzadas para especificações lineares e logarítmicas em preço se equivalem. Mais geralmente, é importante checar quão robustos são os resultados a definições de variáveis como o *market share* do bem exterior, forma funcional — especialmente a forma como variáveis demográficas interagem com as características, hipóteses distribucionais, e assim por diante. Além disso, convém confrontar os resultados com a intuição econômica do pesquisador.

Modelos simétricos (tais como o *logit*) são os menos indicados em termos de estimação de demanda e análise de fusões, principalmente por não levarem em conta uma dimensão fundamental no que se refere a produtos diferenciados — a proximidade entre alternativas no espaço de características, com consequências indesejáveis nas elasticidades cruzadas e, consequentemente, na definição de mercado relevante. Apesar disso, modelos *logit* e derivados têm sido usados com freqüência [Werden e Froeb (1994)]. Entretanto, exceto quando tempo for a única variável relevante, não recomendamos o seu uso e, ainda assim, que sejam feitos testes de especificação.

Comparar modelos contínuos multiestágios e modelos discretos com coeficientes aleatórios é algo bem mais complexo. Modelos multiestágio (e mesmo discretos, como o *logit* aninhado) requerem a definição *a priori* dos segmentos em cada estágio, que não podem ser nem muito pequenos, nem muito grandes, e são freqüentemente difíceis de definir — uma saída seria definir especificações alternativas e verificar a robustez dos resultados para especificações diferentes.

Muitas vezes é de interesse comparar, por exemplo, modelos alternativos apresentados pelas partes envolvidas em um ato de concentração. Fazemos aqui alguns comentários sobre como comparar modelos alternativos, que devem ser usados em conjunto com as considerações feitas nas seções anteriores:

- Comparando MNL e NL: No que se refere a testes de especificação para o modelo *logit*, Hausman e McFadden (1984) propõem dois tipos de testes, sendo o primeiro uma aplicação do “Teste de Hausman” [Hausman (1978a)], em que o MNL é o modelo eficiente sob a hipótese nula e inconsistente sob a alternativa, e o NL o modelo consistente sob a hipótese alternativa; e o segundo obtido através da estimação de um modelo NL, do qual o MNL é um caso particular, e da aplicação da “santíssima trindade” de testes (Razão de Verossimilhança, Wald, Multiplicadores de Lagrange). O segundo teste decorre de um resultado demonstrado por McFadden (1978), de acordo com o qual a estrutura de estágios é consistente com RUM se, e somente se, determinados coeficientes (correspondentes a  $\sigma$  na equação (28)) estão contidos no intervalo (0,1) — caso eles tendam a 1, o modelo se reduz a um MNL e, caso eles sejam maiores do que 1, há substituição entre ramos diferentes. Isso implica que a especificação dos estágios estimada não é consistente com a hipótese RUM, e uma configuração alternativa da árvore de decisão deve necessariamente ser considerada. Trajtenberg (1989) compara MNL e NL em seu estudo de tomógrafos.

- Especificações alternativas em NL e escolha em estágios múltiplos (*multi-stage budgeting*): A divisão de produtos em grupos tanto em modelos NL, como no caso de modelos de escolha em estágios, é um tanto arbitrária.<sup>101</sup> Adicionalmente, no caso do NL, a hipótese de choques i.i.d. intra-ramo, é passível de críticas. Além do problema da arbitrariedade da escolha dos elementos constituintes de cada estágio, deve-se atentar para o tamanho dos ramos de um dado estágio. De forma intuitiva, espera-se que algumas das características de produtos em um mesmo mercado estejam correlacionadas — por exemplo, potência, número de cilindros e dimensões do carro, tal como no caso dos automóveis, em Goldberg (1995). Adotando-se um ramo com poucos produtos, eles tenderão a ser mais homogêneos, gerando um alto grau de colinearidade, mas eliminando a necessidade de incluir todas as características dos produtos. No caso do NL, incluir ramos com muitos produtos tende a mitigar a colinearidade, mas volta-se ao problema da IIA presente no MNL. Além disso, cabe ressaltar que, apesar de capturar

---

101. Rubinfeld (2000) lembra que a substituição entre marcas pertencentes a segmentos (nínhos) distintos em um sistema de escolha em estágios múltiplos opera através da alocação de gastos entre os segmentos, em um estágio anterior, em decorrência da variação no índice de preços de cada segmento. Isso gera um viés para baixo na elasticidade-preço cruzada entre essas marcas pertencentes a distintos segmentos.

correlações entre elementos de um mesmo ramo, o NL não consegue capturar correlações entre ramos distintos, de forma que, quando alternativas não podem ser separadas em diferentes ramos de modo satisfatório, o NL não é aplicável. Para ambos os modelos, sugerimos que, primeiro, especificações alternativas sejam testadas, a fim de comprovar a robustez dos resultados obtidos. A seguir, e apesar de não termos conhecimento de sua implementação empírica, recomendamos o teste de Vuong (descrito no Apêndice) para comparar especificações alternativas de seqüências de “nínhos” para modelos NL e de escolha em estágios múltiplos.<sup>102</sup> Tal teste seria incorporado quando, por exemplo, duas estruturas de ninhos passassem pelos testes de especificação do NL, sendo ambas consistentes com a hipótese RUM, e não tendo uma resposta sobre qual delas é preferível.

- Comparando MNL e BLP: Supondo-se que o econometrista tenha acesso ao método de estimação BLP, uma estratégia de teste óbvia é impor a hipótese nula de coeficientes não-aleatórios (ou seja,  $\beta^o = \beta'' = 0$ ) e comparar modelos BLP e MNL por meio do teste de razão de verossimilhança [ver Wooldridge (2002), ou Newey e McFadden (1994)]. Nesse caso, se a diferença entre as funções de verossimilhança não for substancial, rejeitaria o modelo BLP em favor do MNL. No caso de não se ter acesso ao método de estimação BLP, é teoricamente possível (mesmo que de implementação não-trivial) construir um teste de multiplicadores de Lagrange e testar-se a hipótese nula de coeficientes não-aleatórios ( $\beta^o = \beta'' = 0$ ) ou, equivalentemente, de um MNL. Tal como para a aplicação do teste de Vuong para a comparação de especificações NL ou multiestágios, não é de nosso conhecimento que tal teste tenha sido implementado.

### **1.5.8 Estimação simultânea de demanda e de oferta, ou estimação em estágios**

O *trade-off* entre eficiência e consistência manifesta-se novamente na opção pelo analista de estimar simultaneamente os lados de demanda e de oferta, ou de proceder à estimação em estágios. A vantagem de estimar simultaneamente equações de demanda e de oferta (por exemplo, equações de precificação, além de possivelmente outras equações tais como as de demanda por fatores<sup>103</sup>) é que o maior número de restrições (condições) na estimação deve resultar em menor variância das estimativas dos parâmetros estruturais. No entanto, caso haja erro de especificação no lado de oferta — o que tende a ser mais comum relativamente à

102. Capps, Church e Love (2003) sugerem o critério de dominância de verossimilhança de Pollak e Wales (1991), também um teste não-aninhado.

103. Para a inclusão de equações de demanda por fatores, ver, por exemplo, Gollop e Roberts (1979) e Appelbaum (1979), ou a discussão em Bresnahan (1989).

especificação de demanda — a estimativa de oferta simultaneamente à estimativa de demanda poderá “contaminar” esta última, levando a estimativas inconsistentes de demanda.

Por esse motivo, muitas vezes é preferível proceder à estimativa em estágios: primeiro estima-se o lado de demanda, possivelmente utilizando variáveis de oferta na condição de instrumentos para lidar com o problema de endogeneidade. De posse dos parâmetros estimados de demanda, estes são embutidos na especificação de oferta que é então estimada, em um segundo estágio. Naturalmente, como os parâmetros de demanda a serem incorporados à especificação de oferta neste segundo estágio são estimativas procedentes da inferência no primeiro estágio, e não os verdadeiros valores da população, no segundo estágio deve-se calcular erros-padrão pelo método de *bootstrap*, ou utilizar a teoria de estimadores de dois estágios [ver 2-step estimators, em Wooldridge (2002, cap. 12)].<sup>104</sup> Isso levará em consideração o erro de amostragem (*sampling error*) no primeiro estágio.

Slade (2004), por exemplo, cita que “uma vantagem do procedimento em duas etapas é que a especificação incorreta da CPO não irá contaminar as estimativas de demanda, nas quais tipicamente depositamos maior confiança” (p. 15; tradução livre).

### 1.5.9 Elementos dinâmicos

As especificações de demanda e de oferta apresentadas são estáticas: assume-se que não operem efeitos dinâmicos importantes na indústria tanto pelo lado da demanda quanto pelo lado da oferta, ou que estes não impactem a consistência das estimativas. É fácil conceber situações em que tal premissa pode falhar, tais como:

- Empilhamento de estoques pelo consumidor (*consumer stockpiling*): Quando o consumidor pode acumular estoques de algum produto não-durável (*non-durable*), mas armazenável (*storable*). Por exemplo, ao se deparar com um preço promocional de detergente em pó, o consumidor adquire estoques para várias semanas de consumo. Naturalmente, a elasticidade-preço de demanda de curto prazo, que leva em consideração a reação imediata das vendas à flutuação de preços de acordo com o calendário promocional, poderá ser maior do que a elasticidade-preço de demanda de longo prazo, que poderá ser relevante segundo a finalidade do analista (por exemplo, querendo estimar os efeitos sobre preços de um ato de concentração).<sup>105</sup> Remetemos o leitor a Hendel e Nevo (2001 e 2002).

104. Para uma discussão não-técnica do *bootstrap*, ver Brownstone e Valletta (2001).

105. Lembrar que, *caeteris paribus*, elasticidades-preço de demanda próprias mais elevadas limitam o aumento de preços após a fusão. Portanto, um viés para cima (em termos absolutos) na estimativa de elasticidade significa um viés para baixo no efeito estimado de preços de um ato de concentração.

- Formação de hábito (*habit formation*) ou efeitos de rede (*network effects*): Quando o consumo de amanhã depende sensivelmente do consumo de hoje, tais como no consumo de bens geradores de dependência (*addictive goods*, como narcóticos, tabaco e álcool) dizemos que ocorre formação de hábito. De forma semelhante, operam efeitos de rede em uma indústria quando externalidades estão presentes, tais como tipicamente é o caso em bens de informação, comunicação e entretenimento — a minha utilidade de consumo de *software* irá depender do número de outros consumidores, visto que quanto maior for o mercado, maior serão a variedade e a qualidade ofertadas de *software*; por outro lado, quanto maiores a variedade e a qualidade ofertadas maior será o número de usuários. Uma leitura sugerida é Pollak e Wales (1992), mas ver também Sutton (1998) para uma discussão interessante, acompanhada de exemplos, de efeitos de rede em ação.

- Aprendendo ao fazer (*learning by doing*): Em algumas indústrias, como a indústria química, custos são decrescentes na produção histórica acumulada. Quanto maior for a experiência acumulada de produção, maior é a produtividade. Este é outro exemplo de externalidade, matematicamente isomorfo ao de efeitos de rede, só que operando pelo lado de oferta e não de consumo [ver, novamente, Sutton (1998)].

- Bens duráveis: A demanda presente por alguns bens, como o de equipamentos industriais e agrícolas, irá depender de condições futuras, como a expectativa de renda ou o preço futuro de aço ou de soja. De maneira semelhante, a utilidade de um automóvel é usufruída não somente no período presente como também em períodos futuros [ver Hendel e Nevo (2004)].

Como capturar esses efeitos dinâmicos na estimação estrutural, identificando os parâmetros de forma satisfatória, não é uma questão trivial e é matéria de pesquisa atualmente, apesar de sua complexidade. Algumas vezes variáveis defasadas (em níveis e/ou em diferenças) são adicionadas à especificação, na tentativa de capturar esses efeitos. Entretanto deve-se atentar para o problema de identificação, assim como a possível multicolinearidade entre o valor presente da variável em questão e os valores defasados.

### **1.5.10 Relação entre elasticidades no varejo e elasticidades derivadas do produtor**

É comum o analista estar interessado em avaliar, por exemplo, um ato de concentração entre *produtores* (por exemplo, os fabricantes de chocolate Nestlé e Garoto) mas dispor de dados coletados por agências de pesquisa de mercado (como a ACNielsen) apenas junto a *varejistas* (através, por exemplo, de *scanners*). Apesar

de ter sido tipicamente ignorado na literatura empírica e na prática antitruste, o relacionamento entre as elasticidades-preço de demanda estimadas com dados de varejo e os efeitos sobre preços de uma fusão entre produtores não é trivial.<sup>106</sup> As elasticidades-preço da demanda “final” que os varejistas como um todo observam (e que o analista estima) não são necessariamente as mesmas que as elasticidades enfrentadas pelo produtor, a não ser que certas premissas restritivas sejam satisfeitas (por exemplo, que os varejistas sejam perfeitamente competitivos, ou que os varejistas apliquem um *markup* percentual fixo sobre o preço de compra<sup>107</sup>). Ver, por exemplo, Hausman e Leonard (2002), que abordam esse assunto, mas sobretudo Hosken *et alii* (2002), para uma discussão intuitiva.<sup>108</sup>

Hosken *et alii* (2002) também discutem o importante aspecto de que várias vezes o contrato de fornecimento entre produtores de bens diferenciados e os varejistas de tais produtos envolve formatos de preço (termos comerciais) mais complicados do que meramente preços lineares em quantidades (ou seja, onde o preço para uma unidade é R\$ 5, e o preço para 100 unidades é ainda R\$ 5 por unidade). Por exemplo, é comum observar a prática de descontos por volume, ou bonificações em dinheiro caso um volume de vendas preestabelecido entre produtor e varejista seja atingido. Novamente, esse formato não-linear de preços pode ter implicações não-triviais para a análise dos efeitos de uma fusão entre produtores, utilizando-se dados de varejo (vendas e condições, normalmente lineares, do varejista para o consumidor).

Sugerimos ao leitor interessado nesses contratos verticais de fornecimento e formas de lidar com os problemas citados, os trabalhos empíricos de Villas-Boas (2001); Manuszak (2002); Asker (2004); Bonnet, Dubois e Simioni (2004); Brenkers e Verboven (2004); e Villas-Boas e Zhao (2005). Esses trabalhos fazem parte de uma literatura ainda incipiente que trata de analisar como dados de *scanner*, coletados ao nível de varejo, podem nos informar sobre a estrutura da indústria fornecedora que fica a montante na cadeia vertical de suprimentos.

106. Observe que os dados de varejo normalmente disponíveis ao analista estão agregados para todos os estabelecimentos (dentro de determinado canal, como supermercados com mais de dez *checkouts*) em uma mesma região geográfica. Caso o analista estivesse trabalhando com dados desagregados por estabelecimento, ele teria de lidar com a substituição entre estabelecimentos, questão que não será necessariamente relevante na análise de ato de concentração entre produtores.

107. Hosken *et alii* (2002) lembram que tal comportamento por parte de um varejista será otimizante somente sob condições muito restritivas (por exemplo, varejistas vendendo somente um produto, enfrentando demanda de elasticidade constante e incorrendo em custos marginais 0 além do preço de custo).

108. Os autores especulam que “para alguns produtos onde a lealdade (do consumidor) à marca não seja particularmente forte, a ameaça por parte de varejistas de tirar esses produtos das prateleiras irá disciplinar os aumentos de preço de produtores. Nesses casos, as elasticidades ao nível de varejo poderão subdimensionar as elasticidades ao nível do produtor. Por outro lado, algumas marcas fortes não seriam descontinuadas por varejistas a não ser que o produtor praticasse um aumento substancial de preços. Aí então, as elasticidades ao nível de varejo poderão superdimensionar as elasticidades ao nível do produtor” (p. 26; tradução livre).

### 1.5.11 Calculando custos marginais diretamente

Em algumas situações é possível ao analista calcular o custo marginal diretamente, em vez de inferi-lo a partir da estimação estrutural. Este tipicamente poderá ser o caso quando a tecnologia de produção for simples (por exemplo, a tecnologia de coeficientes fixos) e houver disponibilidade de dados (por exemplo, quando o analista observa os coeficientes fixos bem como o preço de fatores). Exemplos são oferecidos por Genesove e Mullin (1998), que estudam a indústria de refinamento de açúcar (em que uma dada quantidade de açúcar bruto é convertida em açúcar refinado e os autores observam o preço de açúcar bruto importado), e Salvo (2004), que estuda a indústria de cimento (onde os coeficientes fixos e preços de fatores também são observados, mas crucialmente o autor observa os embarques de cimento de cada fábrica para cada mercado local, possibilitando o cálculo do custo de transporte). Apesar do considerável esforço exigido do analista, que terá de conhecer detalhes institucionais da indústria, o cálculo do custo marginal — quando possível — oferece a grande vantagem de substituir a estimação deste, ou mesmo possibilitar a verificação do custo marginal estimado.

Quando a tecnologia for mais complexa, pode ser possível que o analista calcule, a partir de algumas premissas, um limite superior do custo marginal. Essa informação também poderá ser utilizada pelo analista para conferir as estimativas econométricas de custo marginal, ou para aumentar a eficiência da estimação (ao incorporar as restrições decorrentes do limite superior na estimação).

## 1.6 RECOMENDAÇÕES DE POLÍTICA

Apesar de o presente trabalho ter trazido à tona uma variedade de aspectos técnicos da fronteira da literatura, os aspectos institucionais envolvidos no antitruste merecem tanta ou mais atenção do que a técnica. Assim, nossas principais recomendações são no sentido do fortalecimento institucional do SBDC.

Em primeiro lugar, recomendamos o fortalecimento de mecanismos de obtenção de dados de qualidade. Além disso, os dados requeridos e utilizados em estudos feitos por todas as partes em um ato de concentração, ou na instauração de um processo administrativo, devem ser depositados junto ao SBDC, permitindo, assim, a replicação dos resultados obtidos.

Em segundo lugar, recomendamos que análises feitas pelo SBDC sejam feitas antes da concretização do ato. Mesmo uma análise tecnicamente correta e bem fundamentada pode sofrer contestações por implicar a reversão de uma operação, contribuindo para o desgaste da imagem do SBDC diante do grande público, assim como contestações na justiça da parte perdedora.

Em terceiro lugar, recomendamos a valorização da carreira de analista da concorrência. Entendemos que um SBDC não pode prescindir de quadros bem formados e em processo de constante aperfeiçoamento: o material coberto neste trabalho mostra a importância do conhecimento de Organização Industrial e Econometria uma vez que se tenha optado pela utilização de modelos estruturais. Em particular, vemos cursos de extensão e pós-graduação, bem como intercâmbios com SDCs estrangeiros, como fontes sadias de aperfeiçoamento de pessoal.

Do ponto de vista da adoção de modelos descritos no trabalho, entendemos que, num primeiro estágio, o SBDC veja como mais conveniente adotar métodos parcimoniosos como o PCAIDS (utilizado pelo SDC da Nova Zelândia) e o ALM (utilizado pelo SDC americano). No entanto, sob os critérios de *precisão de previsão, tempestividade, parcimônia, operacionalidade e simplicidade de interpretação*, entendemos que, num horizonte de longo prazo, métodos mais adequados seriam, por exemplo, o NL e o sistema de escolha em estágios, desde que adotados com as precauções detalhadas no texto. Tal recomendação encontra-se alinhada com o seguinte comentário de Hosken *et alii* (2002, p. 26), em texto para discussão da Federal Trade Comission: “Many of the demand system estimations presented to the Agencies use the multi-stage budgeting approach with the AIDS demand specification.”

Como mencionamos no texto, os métodos sugeridos encontram-se implementados em pacotes econométricos disponíveis no mercado, não exigindo a programação de rotinas complexas por parte do analista, como no caso de métodos como o modelo logit com coeficientes aleatórios: estimar um modelo de tal complexidade exige uma série de escolhas feitas caso a caso, teste de especificações alternativas, além da adequação aos dados disponíveis para o estudo, requerendo conhecimento e maturidade que seriam obtidos, por exemplo, ao longo de um curso de Doutorado em Economia.<sup>109</sup>

---

109. Isso não significa que modelos dessa natureza não possam ser — ou não tenham sido — implementados no âmbito da análise antitruste brasileira. É de nosso conhecimento, por exemplo, que modelos dessa complexidade têm sido estimados pela Secretaria de Direito Econômico em análise recente de ato de concentração no mercado brasileiro de tratores.

## BIBLIOGRAFIA

- AKAIKE, H. Information theory and an extension of the likelihood ratio principle. In: PETROV, B. N., CSAKI, F. (eds.). *Second International Symposium of Information Theory*. Minnesota Studies in the Philosophy of Science, Budapest: Akademinai Kiado, 1973.
- ALSTON, J. M., FOSTER, K. A., GREEN, R. D. Estimating elasticities with the linear approximate almost ideal demand system: some Monte Carlo results. *Review of Economics and Statistics*, v. 76, p. 351-356, 1994.
- APPELBAUM, E. Testing price taking behaviour. *Journal of Econometrics*, v. 9, p. 283-294, 1979.
- ANDERSON, S. P., DE PALMA, A., THISSE, J.-F. *Discrete choice theory of product differentiation*. The MIT Press, 1992.
- ASKER, J. W. *Diagnosing foreclosure due to exclusive dealing*. NYU Stern, 2004, mimeo.
- BAJARI, P., BENKARD, C. L. *Demand estimation with heterogeneous consumers and unobserved product characteristics: a hedonic approach*. Stanford University, 2004, mimeo.
- BAKER, J., BRESNAHAN, T. F. The gains from Merger or collusion in product-differentiated industries. *Journal of Industrial Economics*, v. 33, n. 4, p. 427-444, 1985.
- BARTEN, A. P. *Theorie en empirie van een Volledig Stelsel van Vraagvergelijkingen*. Rotterdam: University of Rotterdam, 1966 (Doctoral Dissertation).
- BAUM, C. F., SCHAFFER, M. E., STILLMAN, S. *Instrumental variables and GMM: estimation and testing*. Department of Economics, Boston College, 2003 (Working Paper, 545).
- BEN-AKIVA, M. *Structure of passenger travel demand models*. MIT, 1973 (Ph.D. Dissertation).
- BERNDT, E. R., FUSS, M. A., WAVERMAN, L. *Dynamic models of the industrial demand for energy*. Electric Power Institute, EA-580, Palo Alto, CA, 1977.
- BERNHEIM, B. D., WHINSTON, M. D. Multimarket contact and collusive behavior. *Rand Journal of Economics*, v. 21, p. 1-26, 1990.
- BERRY, S. Estimating discrete-choice models of product differentiation. *Rand Journal of Economics*, v. 25, p. 242-262, 1994.
- BERRY, S., LEVINSOHN, J., PAKES, A. Automobile prices in market equilibrium. *Econometrica*, v. 63, p. 841-890, 1995.
- \_\_\_\_\_. Differentiated products demands systems from a combination of micro and macro data: the new car market. *Journal of Political Economy*, v. 112, n. 1, p. 68-105, 2004.
- BERRY, S., LINTON, O., PAKES, A. Limit theorems for estimating the parameters of differentiated product demand systems. *Review of Economic Studies*, 2004 (forthcoming).
- BERRY, S., PAKES, A. Some applications and limitations of recent advances in empirical industrial organization: Merger analysis. *American Economic Review*, v. 83, n. 2, p. 247-252, 1993.
- \_\_\_\_\_. *The pure characteristics demand model*. Harvard University, 2002, mimeo.
- BONNET, C., DUBOIS, P., SIMIONI, M. *Two-part tariffs versus linear pricing between manufacturers and retailers: empirical tests on differentiated products markets*. Toulouse, 2004, mimeo.
- BRENKERS, R., VERBOVEN, F. *Liberalizing a distribution system: the European car market*. K. U. Leuven, 2004, mimeo.

- BRESNAHAN, T. F. The oligopoly solution concept is identified. *Economic Letters*, v. 10, p. 87-92, 1982.
- \_\_\_\_\_. Competition and collusion in the American automobile oligopoly: the 1955 price war. *Journal of Industrial Economics*, v. 35, p. 457-482, 1987.
- \_\_\_\_\_. Empirical studies of industries with market power. In: SCHMALENSEE, R., WILLIG, R. D. (eds.). *Handbook of Industrial Organization*. New York: North-Holland, v. II, p. 1.011-1.057, 1989.
- \_\_\_\_\_. *The apple cinnamon cheerios war: valuing new goods, identifying market power, and economic measurement*. Stanford University, 1997, mimeo.
- BRESNAHAN, T. F., STERN, S., TRAJTENBERG, M. Market segmentation and the sources of rents from innovation: personal computers in the late 1980s. *Rand Journal of Economics*, Special Issue, 1997.
- BROWNSTONE, D., VALLETTA, R. The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. *Journal of Economic Perspectives*, v. 15, p. 129-141, 2001.
- BUSE, A., CHAN, W. H. Invariance, price indices and estimation in almost ideal demand systems. *Empirical Economics*, v. 25, p. 519-539, 2000.
- CAPPS Jr., O., CHURCH, J., LOVE, H. A. Specification issues and confidence intervals in unilateral price effects analysis. *Journal of Econometrics*, v. 113, p. 3-31, 2003.
- CARDELL, N. S. *Extensions of the multinomial logit: the hedonic demand model, the non-independent logit model, and the ranked logit model*. Harvard University, 1989 (Ph.D. Dissertation).
- \_\_\_\_\_. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, v. 13, p. 185-213, 1997.
- CARLTON, D. W. Using economics to improve antitrust policy. Milton Handler Lecture, *Columbia Business Law Review*, 2003 (forthcoming).
- CHAMBERLAIN, G. Multivariate regression models for panel data. *Journal of Econometrics*, v. 18, n. 1, p. 5-46, 1982.
- CHRISTENSEN, L. R., JORGENSON, D. W., LAU, L. J. Transcendental logarithmic utility functions. *American Economic Review*, v. 65, p. 367-383, 1975.
- CORTS, K. S. Conduct parameters and the measurement of market power. *Journal of Econometrics*, v. 88, p. 227-250, 1999.
- CYSNE, R. P. *et alii*. Demanda por cerveja no Brasil: um estudo econométrico. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 31, n. 2, p. 249-268, 2001.
- DAVIS, D. D., WILSON, B. J. *Differentiated product competition and the antitrust logit model: an experimental analysis*. Virginia Commonwealth University, 2003, mimeo.
- DAVIS, P., HUSE, C., VAN REENEN, J. *Evaluating incentives for collusion in differentiated products markets: measuring the coordinated effects of Mergers*. 2005 (in preparation).
- DAVIS, P. J. *Demand models for market level data*. MIT Sloan School of Management, 2001, mimeo.
- \_\_\_\_\_. *Lecture notes in economics of industry*. Department of Economics, London School of Economics, 2004.

- DEATON, A., MUELLBAUER, J. An almost ideal demand system. *American Economic Review*, v. 70, p. 312-326, 1980a.
- \_\_\_\_\_. *Economics and consumer behavior*. Cambridge University Press, 1980b.
- DIXIT, A., STIGLITZ, J. E. Monopolistic competition and optimum product diversity. *American Economic Review*, v. 67, p. 297-308, 1977.
- DUBÉ, J.-P. *Product differentiation and Mergers in the carbonated soft drink industry*. University of Chicago, 2004, mimeo.
- ELLISON, S. F. *et alii*. Characteristics of demand for pharmaceutical products: an examination of four cephalosporins. *Rand Journal of Economics*, v. 28, n. 3, 1997.
- ENCAOUA, D., JACQUEMIN, A. Degree of monopoly, indices of concentration and threat of entry. *International Economic Review*, v. 21, p. 87-105, 1980.
- EPSTEIN, R. J., RUBINFELD, D. L. Merger simulation: a simplified approach with new applications. *Antitrust Law Journal*, v. 69, p. 883-919, 2002.
- FARRELL, J., SHAPIRO, C. Horizontal Mergers: an equilibrium analysis. *American Economic Review*, v. 80, n. 1, p. 107-126, 1990.
- FIUZA, E. P. S. *Automobile demand and supply in Brazil: effects of tax rebates and trade liberalization on markups in the 1990s*. 2002, mimeo.
- FREEDMAN, D. A., PETERS, S. Bootstrapping a regression equation: some empirical results. *Journal of the American Statistical Association*, v. 79, p. 97-106, 1984.
- GASMI, F., LAFFONT, J. J., VUONG, Q. H. A structural approach to empirical analysis of collusive behavior. *European Economic Review*, v. 34, p. 513-523, 1990.
- \_\_\_\_\_. Econometric analysis of collusive behaviour in a softdrink market. *Journal of Economics and Management Strategy*, v. 1, p. 277-311, 1992.
- GASTRI, F., VUONG, Q. An econometric analysis of some duopolistic games in frices an advertising. In: RHODES, R., FOMBY, T. *Advances in econometrics: econometric methods and models for industrial organization*. JAI pres Inc., p. 225-254, 1991.
- GENESOVE, D., MULLIN, W. P. Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *Rand Journal of Economics*, v. 29, p. 355-377, 1998.
- GOLDBERG, P. K. Product differentiation and oligopoly in international markets: the case of the US automobile industry. *Econometrica*, v. 63, n. 4, p. 891-951, 1995.
- GOLLOP, F., ROBERTS, M. Firm interdependence in oligopolistic markets. *Journal of Econometrics*, v. 10, p. 313-331, 1979.
- GORMAN, W. M. A possible procedure for analysing quality differentials in the egg market. *Review of Economic Studies*, v. 47, p. 843-856, 1955, 1980.
- \_\_\_\_\_. Separable utility and aggregation. *Econometrica*, v. 27, p. 469-481, 1959.
- \_\_\_\_\_. The structure of utility functions. *Review of Economic Studies*, v. 5, p. 369-390, 1968.
- \_\_\_\_\_. Separability and aggregation. In: BLACKORBY, C., SHORROCKS, A. F. (eds.). *Collected works of W. M. Gorman*, 1995.

- GREEN, R., ALSTON, J. Elasticities in AIDS models. *American Journal of Agricultural Economics*, v. 72, p. 442-445, 1990.
- HAUSMAN, J. A. Specifications tests in econometrics. *Econometrica*, v. 46, p. 1.251-1.272, 1978.
- \_\_\_\_\_. Specification and estimation of simultaneous equations models. In: GRILICHES, Z., INTILIGATOR, M. (eds.). *Handbook of econometrics*. Amsterdam: North-Holland, 1983.
- \_\_\_\_\_. Valuation of new goods under perfect and imperfect competition. In: BRESNAHAN, T. F., GORDON, R. (eds.). *The economics of new goods. Studies in Income and Wealth*, v. 58. Chicago: National Bureau of Economic Research, 1997a.
- \_\_\_\_\_. *Reply to prof. Bresnahan*. MIT, 1997b, mimeo.
- HAUSMAN, J. A., LEONARD, G. K. Economic analysis of differentiated products mergers using real world data. In: HAUSMAN, J. A., LEONARD, G. K., MASON, G. *Law Review*, v. 5, n. 3, 1997.
- \_\_\_\_\_. The competitive effects of a new product introduction: a case study. *Journal of Industrial Economics*, v. L, n. 3, p. 235-261, 2002.
- HAUSMAN, J. A., LEONARD, G. K., ZONA, J. D. Competitive analysis with differentiated products. *Annales D'Economie et de Statistique*, v. 34, p. 159-180, 1994.
- HAUSMAN, J. A., MCFADDEN, D. Specification tests for the multinomial logit model. *Econometrica*, v. 52, n. 5, p. 1.219-1.240, 1984.
- HENDEL, I., NEVO, A. *Sales and consumer inventory*. NBER , 2001 (Working Paper, 9.048).
- \_\_\_\_\_. *Measuring the implications of sales and consumer stockpiling behavior*. Berkeley: University of California, 2002, mimeo.
- \_\_\_\_\_. Intertemporal substitution and storable products. *Journal of the European Economic Association*, v. 2, n. 2, p. 536-547, 2004.
- HICKS, J. R. *Value and capital*. Oxford University Press, 1936.
- HORIZONTAL MERGER GUIDELINES*. US Department of Justice and the Federal Trade Commission, 1992.
- HOSKEN, D. *et alii*. *Demand system estimation and its application to horizontal Merger analysis*. Federal Trade Commission, 2002, mimeo.
- IVALDI, M. *et alii*. *The economics of unilateral effects*. European Commission, Interim Report for DG Competition, 2003a.
- \_\_\_\_\_. *The economics of tacit collusion*. European Commission, Interim Report for DG Competition, 2003b.
- IVALDI, M., VERBOVEN, F. Quantifying the effects from horizontal Mergers in European competition policy. *International Journal of Industrial Organization*, 2004 (forthcoming).
- IWATA, G. Measurement of conjectural variations in oligopoly. *Econometrica*, v. 42, p. 947-966, 1974.
- KULLBACK, S., LEIBLER, R. On information and sufficiency. *Annals of Mathematical Statistics*, v. 22, n. 1, p. 79-86, 1951.

- KÜHN, K.-U. *The coordinated effects of Mergers in differentiated products markets*. University of Michigan, 2004, mimeo.
- LANCASTER, K. J. A new approach to consumer theory. *Journal of Political Economy*, v. 74, n. 2, p. 132-157, 1966.
- LEE, M.-J. *Methods of moments and semiparametric econometrics for limited dependent variable models*. Berlin: Springer-Verlag, 1966.
- LEONTIEF, W. Composite commodities and the problem of index numbers. *Econometrica*, v. 4, p. 39-59, 1936.
- MANSKI, C. F. The structure of random utility models. *Theory and Decision*, v. 8, p. 229-254, 1977.
- MANUSZAK, M. *The impact of upstream Mergers on retail gasoline markets*. Carnegie Mellon University, 2002, mimeo.
- MARIUZZO, F., WALSH, P. P., WHELAN, C. *Embedding consumer taste for location convenience into a structural model of equilibrium*. Dublin: Trinity College, 2004, mimeo.
- MCFADDEN, D. Conditional logit analysis of qualitative choice behavior. In: ZAREMBKA, P. (ed.). *Frontiers of econometrics*. New York: Academic Press, 1974.
- \_\_\_\_\_. Modeling the choice of residential location. In: KARLGVIST, A. et alii (eds.). *Spatial interaction theory and planning models*. Amsterdam: North-Holland, 1978a.
- \_\_\_\_\_. The general linear profit function. In: FUSS, M. A., MCFADDEN, D. (eds.). *Production economics: a dual approach to theory and applications*, Amsterdam: North-Holland, v. 1, 1978b.
- \_\_\_\_\_. Econometric models of probabilistic choice. In: MANSKI, C., MCFADDEN, D. (eds.). *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press, 1981.
- \_\_\_\_\_. Econometric analysis of qualitative response models. In: GRILICHES, Z., INTILLIGATOR, M. (eds.). *Handbook of econometrics*, Amsterdam: North-Holland, v. III, 1984.
- \_\_\_\_\_. A method of simulated moments for estimation of discrete response models without numerical Integration. *Econometrica*, v. 57, n. 5, p. 995-1.026, 1989.
- MCFADDEN, D., TRAIN, K. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, v. 15, p. 447-470, 2000.
- MOTTA, M. *Competition policy: theory and practice*. Cambridge University Press, 2004.
- NEVO, A. Identification of the oligopoly solution concept in a differentiated-products industry. *Economic Letters*, v. 59, p. 391-395, 1998.
- \_\_\_\_\_. Mergers with differentiated products: the case of the ready-to-eat cereal industry. *Rand Journal of Economics*, v. 31, p. 395-421, 2000.
- \_\_\_\_\_. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, v. 69, p. 307-342, 2001.
- \_\_\_\_\_. *Lecture notes in industrial organization*. Cemmap Masterclass, 2004.

- NEWEY, W. K., MCFADDEN, D. Large sample estimation and hypothesis testing. In: ENGLE, R. F., MCFADDEN, D. (eds.). *Handbook of econometrics*, Amsterdam: North-Holland, v. IV, 1994.
- NEWEY, W. K., WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, v. 55, p. 703-708, 1987.
- PAKES, A. *Lecture notes in graduate industrial organization*. Department of Economics, Harvard University, 2003.
- PAKES, A., MCGUIRE, P. Computation of Markov perfect equilibria: numerical implications of a dynamic differentiated product model. *Rand Journal of Economics*, v. 25, n. 4, p. 555-589, 1994.
- PAKES, A., POLLARD, D. Simulation and the asymptotics of optimization estimators. *Econometrica*, v. 57, n. 5, p. 1.027-1.057, 1989.
- PARKER, P. M., RÖLLER, L. H. Collusive conduct in duopolies: multimarket contact and cross-ownership in the mobile telephone industry. *Rand Journal of Economics*, v. 28, p. 304-322, 1997.
- PASHARDES, P. Bias in estimating the almost ideal demand system with the stone index approximation. *Economic Journal*, v. 103, p. 908-915, 1993.
- PETERS, C. *Evaluating the performance of Merger simulation: evidence from the US airline industry*. Northwestern University, 2001, mimeo.
- PETRIN, A. Quantifying the benefits of new products: the case of the Minivan. *Journal of Political Economy*, v. 110, p. 705-729, 2002.
- PINKSE, J., SLADE, M. E. Mergers, brand competition, and the price of a pint. *European Economic Review*, v. 48, n. 3, p. 617-643, 2004.
- PINKSE, J., SLADE, M. E., BRETT, C. Spatial price competition: a semi-parametric approach. *Econometrica*, v. 70, p. 1.111-1.155, 2002.
- POLLAK, R. A., WALES, T. J. The likelihood dominance criterion: a new approach to model selection. *Journal of Econometrics*, v. 47, p. 227-242, 1991.
- . *Demand system specification and estimation*. Oxford: Oxford University Press, 1992.
- PORTER, R. H. A study of cartel stability: the joint executive committee, 1880-1886. *The Bell Journal of Economics*, v. 14, n. 2, p. 301-314, Autumn 1983.
- REISS, P. C., WOLAK, F. Structural econometric modeling: rationales and examples from industrial organization. *Handbook of Econometrics*, v. 6, 2002 (manuscript of 2002).
- RIVERS, D., VUONG, Q. Model selection tests in nonlinear dynamic models. *The Econometrics Journal*, v. 5, n. 1, p. 1-39, 2002.
- ROBERTS, M. Testing oligopolistic behavior: an application of the variable profit function. *International Journal of Industrial Organization*, v. 2, p. 367-383, 1984.
- RUBINFELD, D. Market definition with differentiated products: the post/Nabisco cereal Merger. *Antitrust Law Journal*, v. 68, p. 163-182, 2000.
- SALVO, A. E. *Inferring conduct under the threat of entry: the case of the Brazilian cement industry*. London School of Economics, 2004 (Sticerd Working Paper, EI 38).

- \_\_\_\_\_. *Price competition, Mergers and structural estimation in oligopoly*. London School of Economics, 2005 (Ph.D. Dissertation).
- SCHMALENSSEE, R. Entry deterrence in the ready-to-eat breakfast cereal industry. *Bell Journal of Economics*, v. 9, p. 305-327, 1978.
- SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v. 6, p. 461-464, 1978.
- SHAPIRO, C. Mergers with differentiated products. *Antitrust*, v. 10, n. 2, p. 23-30, 1996.
- SLADE, M. E. Market power and joint dominance in UK brewing. *Journal of Industrial Economics*, v. LII, p. 133-163, 2004.
- SPENCE, M. Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies*, v. 43, p. 217-235, 1976.
- STONE, J. Linear expenditure systems and demand analysis: an application to the pattern of British demand. *Economic Journal*, v. 64, p. 511-527, 1954.
- SUSLOW, V. Estimating monopoly behavior with competitive recycling: an application to Alcoa. *Rand Journal of Economics*, v. 17, p. 389-403, 1986.
- SUTTON, J. *Technology and market structure: theory and history*. Cambridge, MA: The MIT Press, 1998.
- THEIL, H. The information approach to demand analysis. *Econometrica*, v. 6, p. 375-380, 1965.
- TRAJTNBERG, M. The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy*, v. 97, n. 2, p. 444-479, 1989.
- VERBOVEN, F. The nested logit model and representative consumer theory. *Economics Letters*, v. 50, n. 1, p. 57-63, 1996a.
- \_\_\_\_\_. International price discrimination in the European car market. *Rand Journal of Economics*, v. 27, n. 2, p. 240-268, 1996b.
- VILLAS-BOAS, J. M., ZHAO, Y. Retailer, manufacturers, and individual consumers: modeling the supply side in the ketchup marketplace. *Journal of Marketing Research*, v. 42, p. 83-95, 2005.
- VILLAS-BOAS, S. B. *Vertical contracts between manufacturers and retailers: an empirical analysis*. University of California at Berkeley, 2001, mimeo.
- VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica*, v. 57, p. 307-333, 1989.
- WERDEN, G. J. The history of antitrust market delineation. *Marquette Law Review*, v. 76, p. 123-215, 1992a.
- \_\_\_\_\_. Four suggestions on market delineation. *Antitrust Bulletin*, p. 107-121, 1992b.
- \_\_\_\_\_. A robust test for consumer welfare enhancing Mergers among sellers of differentiated products. *Journal of Industrial Economics*, v. 44, n. 4, p. 409-413, 1996.
- \_\_\_\_\_. Simulating the effects of differentiated products Mergers: a practitioner's guide. *Proceedings of the NE-165 Conference*, 1997.
- WERDEN, G. J., FROEB, L. M. The effects of Mergers in differentiated products industries: logit demand and Merger policy. *Journal of Law, Economics, & Organization*, v. 194, p. 407-426, 1994.

- WERDEN, G. J., FROEB, L. M., TARDIFF, T. J. The use of the logit model in applied industrial organization. *International Journal of the Economics of Business*, v. 3, n. 1, p. 83-105, 1996.
- WERDEN, G. J., FROEB, L. M., TSCHANTZ, S. The logit model for simulating unilateral competitive effects. In: HARKRIDER, J. (ed.). *Econometrics in antitrust*. ABA Section on Antitrust Law, 1994 (forthcoming).
- WILLIG, R. D. Merger analysis, industrial organization theory, and the Merger guidelines. *Brookings Papers on Economic Activity, Microeconomics*, p. 281-332, 1991.
- WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press, 2002.

## APÊNDICE I

### DEMANDA

#### I.1 Dedução da função de demanda AIDS — expressão (15)

Seguimos Deaton e Muellbauer (1980a) ao deduzirmos a função de demanda AIDS no formato de *share* de valor. Consideremos uma classe de preferências representada por uma função de custos do seguinte tipo:

$$\log c(u, p) = (1-u)\log\{a(p)\} + u \log\{b(p)\}$$

onde  $u$  denota utilidade,  $p$  denota preços, e  $a(p)$  e  $b(p)$  são funções homogêneas de primeiro grau em preços, definidas conforme segue.<sup>110</sup>  $\log\{a(p)\}$  é função quadrática dos logaritmos dos preços

$$\log\{a(p)\} = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j$$

e  $\log\{b(p)\} - \log\{a(p)\}$  é dado por

$$\log\{b(p)\} - \log\{a(p)\} = \beta_0 \prod_k p_k^{\beta_k}$$

onde  $\alpha_i$ ,  $\beta_i$  e  $\gamma_{ij}^*$  são parâmetros, tal que  $\sum_i \alpha_i = 1$  e  $\sum_j \gamma_{kj}^* = \sum_k \gamma_{kj}^* = \sum_j \beta_j = 0$ . (Essas restrições são requeridas para que a função de custos seja linearmente

110. Observe que  $c(u, p) = [a(p)]^{1-u} [b(p)]^u$ . Com algumas exceções (ver o Apêndice de Deaton e Muellbauer (1980)), o valor de  $u$  cairá entre 0 e 1 tal que  $a(p)$  e  $b(p)$  podem ser interpretados como os custos de subsistência ( $u=0$ ) e saciedade (ou bliss,  $u=1$ ), respectivamente.

homogênea em preços.) A função de custos (ou melhor, o respectivo log) pode ser escrita como

$$\log c(u, p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j + u \beta_0 \prod_k p_k^{\beta_k} \quad (39)$$

Pelo Lema de Shephard,  $\frac{\partial c(u, p)}{\partial p_i} = q_i(u, p)$ , a demanda compensada pelo bem  $i$ , e portanto:

$$\frac{\partial \log c(u, p)}{\partial \log p_i} = \frac{p_i}{c(u, p)} \frac{\partial c(u, p)}{\partial p_i} = \frac{p_i q_i(u, p)}{c(u, p)} = s_i(u, p) \quad (40)$$

onde  $s_i$  denota a participação orçamentária (*share* de valor) do bem  $i$ . Essa participação orçamentária é então obtida derivando (39):

$$s_i(u, p) = \frac{\partial \log c(u, p)}{\partial \log p_i} = \alpha_i + \sum_j \gamma_{ij}^* \log p_j + \beta_i u \beta_0 \prod_k p_k^{\beta_k} \quad (41)$$

onde  $\gamma_{ij}^*$  denota a média aritmética de  $\gamma_{ij}^*$  e  $\gamma_{ji}^*$

$$\gamma_{ij}^* = \frac{1}{2} (\gamma_{ij}^* + \gamma_{ji}^*) \quad (42)$$

e, portanto,  $\gamma_{ij}^* = \gamma_{ji}^*$  (simetria). A maximização de utilidade implica que  $c(u, p)$  seja igual ao dispêndio total  $Y$ . A função de custos definida por (39) pode então ser invertida, gerando:

$$\begin{aligned} u \beta_0 \prod_k p_k^{\beta_k} &= \log Y - \left( \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j \right) \\ &= \log Y - \left( \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j \right) \end{aligned} \quad (43)$$

onde se utiliza (42) no último passo.

Defina agora o índice de preços  $P$  dado pela expressão translog

$$\log P := \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j \quad (44)$$

A participação orçamentária do bem  $i$  dada por (41) pode então ser expressa, usando (43) e (44), como função de preços e do dispêndio:

$$s_i(u, p) = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log \left( \frac{Y}{P} \right) \quad (45)$$

As restrições aplicadas aos parâmetros de (39) e a restrição de simetria implicam

$$\sum_i \alpha_i = 1 \quad \sum_i \gamma_{ij} = \sum_i \beta_i = 0 \quad (\text{"somatórios"})$$

$$\sum_j \gamma_{ij} = 0 \quad (\text{homogeneidade})$$

$$\gamma_{ij} = \gamma_{ji} \quad (\text{simetria})$$

Contanto que essas restrições sejam observadas, (45) caracteriza um sistema de funções de demanda cujo somatório corresponde ao dispêndio total ( $\sum_i s_i = 1$ ), são homogêneos de grau 0 com relação a preços e dispêndio e satisfazem a propriedade de simetria de Slutsky. Variações nos preços relativos impactam demanda através dos parâmetros  $\gamma_{ij}$ : uma variação percentual unitária no preço do bem  $j$  tem um efeito  $\gamma_{ij}$  sobre a participação orçamentária do bem  $i$ , mantido constante o dispêndio real  $\frac{Y}{P}$ . Variações no dispêndio real impactam demanda através dos parâmetros  $\beta_i$ . Esses parâmetros somam 0 e são positivos para “bens de luxo” e negativos para “bens de necessidade”.

Na prática [ver a aplicação empírica de Deaton e Muellbauer (1980a), bem como a nota 18], o índice de preços  $P$  dado em (44) pode ser aproximado por um índice de preços de  $P^S$  [Stone (1954)]:

$$\log P^S := \sum_i s_i \log p_i \quad (46)$$

Convém lembrar que essa dedução refere-se à demanda do consumidor individual. A agregação das demandas individuais, gerando demanda agregada, baseia-se em extenso material teórico-econômico, ressaltando-se o trabalho de Gorman (1995).

### I.2 Dedução da elasticidade-preço referente ao sistema de escolha em dois estágios — expressão (18)

Lembrando que  $s_{int} = \frac{p_{int} q_{int}}{Y_{nt}}$ , temos que

$$\log q_{int} = \log Y_{nt} - \log p_{int} + \log s_{int}$$

onde  $s_{int}$  é dado pela equação de nível inferior (15). Assim, as elasticidades-preço de demanda podem ser calculadas como (suprimindo os subscritos  $n$  e  $t$ , para simplificar a exposição, e lembrando que a função indicadora  $1[i = j]$  retorna 1 se  $i = j$ , retornando 0 no caso contrário)

$$\eta_{ij} = \frac{\partial \log q_i}{\partial \log p_j} = \frac{\partial (\log Y - \log p_i + \log s_i)}{\partial \log p_j} = \frac{\partial \log Y}{\partial \log p_j} - 1[i = j] + \frac{\partial \log s_i}{\partial \log p_j} \quad (47)$$

onde, diante de (15),

$$\frac{\partial \log s_i}{\partial \log p_j} = \frac{1}{s_i} \frac{\partial s_i}{\partial \log p_j} = \frac{1}{s_i} \left( \beta_i \frac{\partial \log Y}{\partial \log p_j} - \beta_i \frac{\partial \log P}{\partial \log p_j} + \gamma_{ij} \right) \quad (48)$$

Usando (48) em (47), temos

$$\eta_{ij} = \frac{\partial \log Y}{\partial \log p_j} \left( 1 + \frac{\beta_i}{s_i} \right) - 1[i = j] + \frac{1}{s_i} \left( \gamma_{ij} - \beta_i \frac{\partial \log P}{\partial \log p_j} \right) \quad (49)$$

Recordando o índice de preços de Stone (16), temos que

$$\frac{\partial \log P}{\partial \log p_j} = \frac{\partial (\sum_i w_i \partial \log p_i)}{\partial \log p_j} = w_j + \sum_i \frac{\partial w_i}{\partial \log p_j} \log p_i \simeq w_j \quad (50)$$

assumindo que  $\frac{\partial w_i}{\partial \log p_j} \simeq 0$ .<sup>111</sup> Quanto a  $\frac{\partial \log Y}{\partial \log p_j}$ , ainda que  $Y = \sum_i p_i q_i$ , podemos aproximar<sup>112</sup>  $Y$  pelo produto do consumo total  $Q = \sum_i q_i$  e o índice de preços Stone  $P$ ; ou seja:

$$Y = \sum_i p_i q_i \simeq P \sum_i q_i = PQ$$

o que equivale a

$$\log Y = \log P + \log Q$$

onde  $\log Q$  é dado pela equação de nível superior (17). Dessa forma,

$$\frac{\partial \log Y}{\partial \log p_j} = \frac{\partial \log P}{\partial \log p_j} + \frac{\partial \log Q}{\partial \log p_j} = \frac{\partial \log P}{\partial \log p_j} \left( 1 + \frac{\partial \log Q}{\partial \log P} \right)$$

111. Essa aproximação segue, por exemplo, Hausman e Leonard (2002). Caso se utilize o índice de preços de Stone (46),  $\sum_i s_i \log p_i$  em vez da versão média do índice de preços de Stone (16),  $\sum_i w_i \log p_i$  temos que (suprimindo por conveniência o subscrito  $i$ )

$\frac{\partial \log P}{\partial \log p_j} = s_j + \sum_i \frac{\partial s_i}{\partial \log p_j} \log p_i$  onde  $\frac{\partial s_i}{\partial \log p_j} = \gamma_{ij} + \beta_i \frac{\partial \log Y}{\partial \log p_j} - \beta_i (\alpha_j + \sum_k \gamma_{kj} \log p_k)$ . Ver nota 18.

112. Podemos evitar esta aproximação ao definirmos a variável  $Q$  na regressão (17) como sendo dada pela divisão da receita total da indústria  $Y$  pelo índice de preços de Stone  $P$  (ou seja,  $Q = Y/P$ ), em vez de definirmos  $Q$  como o consumo total (ou seja,  $Q = \sum_i q_i$ ) como fizemos na apresentação do estágio superior. Na prática, as estimativas devem ser próximas.

que pode ser reescrita, diante de (50) e (17), como

$$\frac{\partial \log Y}{\partial \log p_j} = w_j (1 + \gamma)$$

Usando esta equação assim como novamente a (50) em (49), temos finalmente que:

$$\eta_{ij} = w_i (1 + \gamma) \left( 1 + \frac{\beta_i}{s_i} \right) - 1[i = j] + \frac{1}{s_i} (\gamma_{ij} - \beta_i w_j)$$

que é a expressão para as elasticidades dada no texto.<sup>113</sup>

## APÊNDICE II

### ECONOMETRIA

#### II.1 O método delta<sup>114</sup>

O método delta é uma forma de obter a distribuição de funções de um estimador. Seja  $\{\hat{\theta}_N\}$  uma seqüência de estimadores de um vetor  $p$ -dimensional  $\theta \in \Theta$ , e suponha que  $\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, V)$ , onde  $V$  é uma matriz positiva definida de dimensão  $p \times p$ .

Considere agora uma função  $f: \Theta \rightarrow \mathbb{R}^q$  continuamente diferenciável no espaço  $\Theta \subset \mathbb{R}^p$ , com  $q \leq p$ , e suponha que  $\theta$  seja um ponto interior de  $\Theta$ . Defina ainda  $f_\theta := \partial f / \partial \theta$  como o gradiente de  $f$ , com dimensão  $q \times p$ . Então:

$$\sqrt{N}(f(\hat{\theta}_N) - f(\theta_0)) \xrightarrow{d} N\left(0, f_\theta(\theta_0) \cdot V \cdot f_\theta(\theta_0)'\right)$$

Demonstração informal: como  $\theta_0$  é ponto interior de  $\Theta$  e  $\hat{\theta}_N$  é um estimador consistente de  $\theta_0$ , tem-se que  $\hat{\theta}_N$  é ponto interior de um subconjunto convexo

113. Essa equação é utilizada, por exemplo, por Hausman e Leonard (2002). Para uma discussão sobre diferentes expressões para as elasticidades utilizadas na literatura, ver Alston, Foster e Green (1994).

114. A discussão desta seção segue de perto Wooldridge (2002, cap. 12), cuja leitura recomendamos.

de  $\Theta$  contendo  $\theta_0$  com probabilidade tendendo a 1. Assim, com probabilidade tendendo a 1, pode-se usar a expansão  $f(\hat{\theta}_N) = f(\theta_0) + f_{\theta}(\theta^*) \cdot (\hat{\theta}_N - \theta_0)$  onde  $\theta^*$  é um ponto interior de um segmento ligando  $\hat{\theta}_N$  e  $\theta^*$ . Como  $\theta^*$  está no segmento (estando, inclusive, mais próximo de  $\theta_0$  que  $\hat{\theta}_N$ , ele também vai convergir em probabilidade para  $\theta_0$ ). Resulta daí que  $f_{\theta}(\theta^*) \rightarrow^p f_{\theta}(\theta_0)$  pelo Teorema de Slutsky. Pode-se então escrever

$$\begin{aligned}\sqrt{N}(f(\hat{\theta}_N) - f(\theta_0)) &= f_{\theta}(\theta^*) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) \\ &= f_{\theta}(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) + [f_{\theta}(\theta^*) - f_{\theta}(\theta_0)] \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) \\ &= f_{\theta}(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) + o_p(1) \cdot O_p(1) \\ &= f_{\theta}(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) + o_p(1)\end{aligned}$$

onde a segunda linha é obtida somando-se e subtraindo-se  $f_{\theta}(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0)$  ao lado direito da igualdade na primeira linha; a terceira linha é obtida da consistência de  $f_{\theta}(\theta^*)$  (pois  $f_{\theta}(\theta^*) \rightarrow^p f_{\theta}(\theta_0)$  equivale a  $f_{\theta}(\theta^*) - f_{\theta}(\theta_0) = o_p(1)$ ), e do fato de  $\sqrt{N} \cdot (\hat{\theta}_N - \theta_0)$  ser limitada em probabilidade ( $O_p(1)$ ); a última linha vem do fato de o produto entre uma variável aleatória que converge para 0 e uma limitada em probabilidade convergir também para 0.

Conclui-se então que, multiplicando-se  $f_{\theta}(\theta_0)$  a uma variável aleatória  $N(0, V)$  obtém-se  $N\left(0, f_{\theta}(\theta_0) \cdot V \cdot f_{\theta}(\theta_0)'\right)$ , e o resultado,  $\sqrt{N}(f(\hat{\theta}_N) - f(\theta_0)) \rightarrow^d N\left(0, f_{\theta}(\theta_0) \cdot V \cdot f_{\theta}(\theta_0)'\right)$ . A estimção consistente da matriz de covariância é feita plugando-se  $\hat{\theta}_N$  no lugar de  $\theta_0$  no gradiente  $f_{\theta}(\cdot)$ .

## II.2 O teste de Vuong

Focaremos agora critérios de seleção de modelos não-aninhados, ou seja, modelos em que um não pode ser obtido como caso particular do outro. Apesar de possuir algumas inconveniências, tais como desempenho insatisfatório em pequenas amostras, utilizaremos o teste de Vuong (1989), não só por ser bem estabelecido, como por não ter competidores bem estabelecidos [ver ainda Rivers e Vuong, (2002), para generalizações do Teste de Vuong].

O teste de Vuong é baseado no critério de informação de Kullback-Leibler [KLIC - Kullback e Leibler (1951)], uma medida de “distância”,<sup>115</sup> ou discrepância, entre dois modelos. O KLIC é definido por:

$$KLIC := E_0 \left[ \ln b_0(Y_i | X_i) \right] - E_0 \left[ \ln f(Y_i | X_i, \beta^*) \right]$$

onde:

$b_0(\cdot|\cdot)$  é a densidade condicional verdadeira de  $Y_i$  dado  $X_i$ , que é, na verdade, desconhecida;

$E_0$  é a esperança tomada de acordo com o modelo verdadeiro; e

$\beta^*$  são os valores pseudo-verdadeiros de  $\beta$ , ou seja, as estimativas de  $\beta$  quando  $f(Y_i|X_i)$  não é o modelo verdadeiro.

A partir do KLIC, define-se o “melhor” modelo como sendo aquele que minimiza a discrepância, pelo fato de estar tão próximo quanto possível do modelo verdadeiro. Tal modelo é obtido pela maximização de  $E_0 \left[ \ln f(Y_i | X_i, \beta^*) \right]$ , tal como no critério de quasi-máxima verossimilhança (QML). Tal observação permite concluir que escolher entre modelos consiste em comparar os valores esperados de suas log-verossimilhanças, e é exatamente nisso que se baseia o teste de Vuong.

Para fixar idéias, considere dois modelos,  $M_f = \{F_\beta, \beta \in B\}$  e  $M_g = \{G_\gamma, \gamma \in \Gamma\}$ , onde  $F$  e  $G$  são funções distribuição, caracterizando os processos geradores de dados dos dois modelos, com densidades  $f$  e  $g$ , respectivamente. Os modelos acima definidos são chamados (estritamente) não-aninhados quando  $M_f \cap M_g = \emptyset$ , que pode ser melhor entendido como refletindo o fato de não haver nenhuma densidade condicional  $(Y|X)$  que seja elemento de  $M_f$  e  $M_g$ , para quase todo  $x$ .

*Exemplo.* Considere modelos de regressão linear em que se tem interesse de testar a distribuição dos erros. Se  $M_f$  contempla o caso de erros com distribuição Normal e  $M_g$  o caso de erros com distribuição logística, os modelos são não-aninhados. ■

Uma dificuldade óbvia no que concerne a testes de hipóteses decidindo entre  $M_f$  e  $M_g$  é que se for imposta a nula de acordo com a qual o modelo  $M_f$  é verdadeiro, sua rejeição não implicará nenhuma informação quanto à correção do modelo

---

115. O termo distância é referido entre aspas pelo fato de, a rigor, o KLIC não ser uma distância, por não valer  $KLIC(f, g) = KLIC(g, f)$ , propriedade fundamental a que uma distância deve satisfazer.

$M_g$ . Naturalmente, então, vem a idéia de se considerar a nula como supondo a equivalência entre os modelos comparados, ou seja,

$$H_0 : E_0 \left[ \ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right] = 0$$

De acordo com a hipótese nula, ambos os modelos estão igualmente próximos da especificação correta.<sup>116</sup> É importante notar que agora existem duas hipóteses alternativas, dadas por:

$$H_f : E_0 \left[ \ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right] > 0$$

em que o modelo  $M_f$  é preferido ao  $M_g$ , e

$$H_g : E_0 \left[ \ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right] < 0$$

segundo a qual o modelo  $M_g$  é preferido ao  $M_f$ .

Apesar de a esperança acima ser desconhecida, Vuong (1989) demonstra que, sob condições de regularidade,

$$\frac{1}{N} LR_N(\hat{\beta}, \hat{\gamma}) \xrightarrow{a.s.} E_0 \left[ \ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right]$$

onde:

$$LR_N(\hat{\beta}, \hat{\gamma}) = \ln f(Y | X, \hat{\beta}) - \ln g(Y | Z, \hat{\gamma})$$

---

116. Note que não se fala em  $f$  ou  $g$  serem “os modelos verdadeiros” (ou estarem corretamente especificados) — os modelos são aqui tratados simetricamente.

é a estatística da razão de verossimilhança. Sob a hipótese nula, tem-se então a seguinte convergência assintótica:

$$\frac{LR_N(\hat{\beta}, \hat{\gamma})}{\sqrt{N} \cdot \hat{w}_N} \xrightarrow{d} N(0, 1)$$

onde:

$$\hat{w}_N^2 = \frac{1}{N} \sum_{i=1}^N \left[ \ln \frac{f(Y_i | X_i, \hat{\beta})}{g(Y_i | X_i, \hat{\gamma})} \right]^2 - \left[ \frac{1}{N} \sum_{i=1}^N \ln \frac{f(Y_i | X_i, \hat{\beta})}{g(Y_i | Z_i, \hat{\gamma})} \right]^2$$

Ou seja, a estatística de teste pode ser entendida como a razão de log-verossimilhança média, convenientemente normalizada.

Definindo-se as hipóteses alternativas  $H_f$  e  $H_g$  segundo as quais os modelos  $f$  e  $g$ , respectivamente, são preferidos, o teste pode ser interpretado da seguinte forma: não se pode rejeitar a hipótese nula caso a estatística de teste esteja suficientemente próxima de 0; não se pode rejeitar  $H_f(H_g)$  caso a estatística de teste seja significativamente maior (menor) que 0. Em termos formais, Vuong (1989) mostra que, sob  $H_f$  tem-se

$$\frac{LR_N(\hat{\beta}, \hat{\gamma})}{\sqrt{N} \cdot \hat{w}_N} \xrightarrow{as} +\infty$$

enquanto, sob  $H_g$ ,

$$\frac{LR_N(\hat{\beta}, \hat{\gamma})}{\sqrt{N} \cdot \hat{w}_N} \xrightarrow{as} -\infty$$

O resultado teórico sugere um procedimento simples ao se implementar o teste. Primeiro, escolhe-se um valor crítico  $c$  associado a um quantil de interesse da distribuição Normal-padrão; se a estatística de teste é maior que o valor crítico  $c$ , o modelo  $M_f$  é o preferido; caso a estatística de teste tome valor menos que  $-c$ , o modelo  $M_g$  é o preferido. Se o valor da estatística de teste estiver situado entre  $-c$  e  $c$ , não se pode discriminar entre os modelos.

Caso o número de coeficientes difira entre os modelos, é preciso fazer uma correção de graus de liberdade, e Vuong sugere o uso de critérios como os de Akaike (1973) ou Schwarz (1978). Como o critério BIC de Schwarz é sabidamente mais indicado que o de Akaike, sugerimos o seu uso na correção, o que gera a estatística de teste corrigida:

$$\widetilde{LR}_N(\hat{\beta}, \hat{\gamma}) := LR_N(\hat{\beta}, \hat{\gamma}) - \left[ \frac{p}{2} \ln(N) - \frac{q}{2} \ln(N) \right]$$

onde  $p$  e  $q$  são as dimensões dos vetores  $\beta$  e  $\gamma$ , respectivamente.

Ao usar o teste de Vuong, deve-se levar em conta que esse é um teste pouco potente — a convergência da estatística de teste para a distribuição Normal precisa de amostras de tamanho razoável para ocorrer.



# Capítulo 2

Como discutido no Capítulo 1, a identificação de custos e condutas no equilíbrio de mercado pode ser tratada pela estimativa simultânea de oferta e demanda. Em complemento à excelente resenha feita naquele texto, era importante estudar o lado da oferta em casos de atos de concentração, pois, na ausência dessa estimativa, não se poderiam simular variantes de equilíbrio pós-fusão para diferentes cenários de variação dos custos ao longo da curva de custos original ou em resposta a deslocamentos da mesma propiciados pelas eficiências (específicas e não-pecuniárias) do ato.

Pioneer e Pinheiro, no Capítulo 2, dedicaram-se a examinar como as eficiências alegadas pelas requerentes deveriam ser incorporadas na simulação do novo equilíbrio ou, invertendo a pergunta, de quanto deveriam ser as eficiências para que os preços e as quantidades de equilíbrio pré-fusão fossem mantidos. Além disso, investigaram qual o grau de incerteza — traduzido por margem de erro estatístico — incorrido nesses exercícios. Tentou-se ainda avaliar a probabilidade de entrada propiciada pelo aumento da concentração, mas isso exige modelos computacionalmente complexos e que ainda demandam bastante pesquisa para se firmarem como alternativas viáveis ao analista antitruste.

Para a estimativa da oferta do mercado, os autores apresentam duas alternativas: a estimativa direta da função custo e a estimativa da função de produção. A escolha entre uma e outra depende do conjunto de dados disponível. Os dados requeridos para a estimativa e a simulação em cada caso são explicitamente listados pelos autores, o que ajuda o analista a escolher o método e/ou modelo mais adequado. Um procedimento simples de *bootstrap* é sugerido como metodologia de cálculo das margens de erro dos novos equilíbrios. Rotinas de MatLab para a simulação são incluídas no Apêndice. Os

autores também discutem a quais eficiências e parâmetros o analista antitruste deve atentar, para cada medida de bem-estar cuja perda a autoridade antitruste queira minimizar.

## CAPÍTULO 2

# MARGENS DE ERRO E EFICIÊNCIAS EM FUSÕES

Heleno Martins Pioner

Da University of Chicago

Maurício Canêdo-Pinheiro

Do Instituto Brasileiro de Economia e da Escola de Pós-Graduação em Economia da FGV

### 2.1 INTRODUÇÃO

O objetivo deste capítulo é apresentar métodos quantitativos na análise de fusões horizontais, mais especificamente, mostrar formas de estimar intervalos de confiança e margens de erro para os resultados de modelos de simulação e para o cálculo de eficiências.

O que chamamos de modelo de simulação de fusões, como é mostrado na Seção 2.2, nada mais é do que estimar um modelo estrutural de indústria e usar esta estimativa para prever o novo equilíbrio de preços das firmas após a fusão. Como modelo estrutural queremos dizer que fazemos uma série de hipóteses sobre como os consumidores e firmas se comportam. Essas hipóteses estão representadas na especificação do sistema de demanda (lado dos consumidores) e na forma da interação estratégica entre as firmas (lado da oferta). Uma vez com o modelo em mãos, podemos recorrer aos dados de preços, quantidades e outros tipos de controles de forma a estimar os parâmetros desse modelo. É para isso que tanto discutimos a estimação da demanda quando falamos sobre simulação de fusões, pois da forma como o modelo é feito, os dados de demanda são suficientes para a sua identificação. O modelo estimado possui toda a informação necessária para que possamos construir o nosso contrafactual: se duas firmas nessa indústria se fundirem, qual será o novo equilíbrio de preços? Assim, usando-se os parâmetros estimados na primeira etapa pode-se inferir os preços nesse novo equilíbrio.

Como toda estimação de modelos estruturais, os resultados são, de certa forma, dependentes das hipóteses feitas e das formas paramétricas assumidas. Assim, nos modelos usados até hoje, não se costuma estimar a função custo ou a função

de produção para determinar o lado da oferta; as hipóteses sobre a forma de interação estratégica (Bertrand) e as hipóteses sobre a forma funcional do custo marginal (constante) permitem que o analista infira os custos marginais de cada firma usando apenas os dados de equilíbrio observados no mercado antes da fusão. Como mostramos na Seção 2.2, essas hipóteses são bastante fortes e podem viesar os resultados da simulação.

Devido à dependência das estimativas com a forma funcional escolhida para a demanda, vários autores buscam formas alternativas de testar o aumento de poder de mercado que sejam robustas à forma funcional. O método escolhido até então leva o nome de redução compensatória de custo marginal. A idéia seria encontrar qual deveria ser a redução no custo marginal das firmas fusionadas de forma a manter os preços pós-fusão iguais àqueles antes da fusão. Werden (1996) argumenta que, usando as mesmas hipóteses sobre interação estratégica, poderíamos estimar essa redução e esse valor seria independente da forma funcional da demanda. Infelizmente, esse resultado não é robusto: como mostramos na Seção 2.2, essas reduções são dependentes dos valores das elasticidades pré-fusão, as quais são estimadas mediante algum tipo de hipótese sobre a forma funcional da demanda. Logo, continuamos na busca de algum tipo de simulação que seja robusto à forma funcional da demanda.

Até aqui discutimos apenas a determinação desse novo preço de equilíbrio. Entretanto, estimativas de ponto como essas possuem pouco significado estatístico se não forem acompanhadas da correspondente estimativa da variância da mesma. A princípio, seria possível argumentar que não haveria problema em estimar essa variância: bastaria aplicar o método delta e estaríamos completos. Todavia, a realidade é bem distinta; devido à natureza não-linear do problema, é bastante complicado derivar a variância desses preços pós-fusão analiticamente. A alternativa encontrada por Capps Jr., Church e Love (2003) foi partir para o uso de métodos não-paramétricos para essa estimação. O problema no uso dessa técnica está nas características dos dados utilizados: em geral, os dados de demanda possuem algum tipo de correlação serial. Na eventualidade de não ser possível corrigir essa autocorrelação, os valores estimados por *bootstrap* serão seriamente viesados, o que invalidaria os intervalos de confiança estimados.

Uma vez explicado o modelo de simulação, passamos à análise das eficiências. Como é discutido mais profundamente na Seção 2.3, o primeiro passo para a análise de eficiências está na determinação da função de bem-estar a ser otimizada. Algumas legislações consideram apenas o bem-estar dos consumidores; outras incluem os ganhos das firmas no cálculo do bem-estar da economia. Apesar de a escolha da função não ser objeto de estudo aqui, podemos apresentar quais serão

as variáveis de interesse do regulador para cada uma das funções acima. Particularmente, um regulador que estiver interessado apenas no bem-estar dos consumidores deverá se preocupar apenas com reduções no custo marginal de produção, enquanto um regulador que almeja calcular a eficiência total da economia precisa olhar também para reduções no custo fixo das firmas.

Como os resultados de eficiência estão intrinsecamente ligados às propriedades da função custo/função de produção, o passo seguinte seria estudar como estimar essas duas funções e qual delas é a mais indicada para o problema de análise antitruste. Isso é feito no meio da Seção 2.3, em que apresentamos os métodos mais indicados para cada uma das funções. Mostramos que, com maior disponibilidade de dados desagregados (firmas e/ou plantas), a literatura tem seguido para a estimativa de funções de produção. Além disso, a dificuldade para lidar com erros de medida nos preços dos insumos prejudica a utilização da função custo.

Contradicoriatamente, encerramos a Seção 2.3 propondo a estimativa da função custo como sendo a melhor abordagem para um regulador interessado em medir os ganhos de eficiência de uma fusão. Isso porque dados necessários para a identificação da função de produção — em especial, sobre a entrada e saída de firmas da indústria —, são de acesso difícil para o regulador, tanto em termos de coleta como em termos de tempestividade da análise. Dessa forma, utilizar os dados das firmas fusionadas para estimar a função de produção pode gerar mais problemas que soluções. Não há garantias de que a estimativa da função custo não gere problemas; eles existem e são apontados na seção correspondente, mas, a nosso ver, com os dados disponíveis ao analista, esta é a melhor opção existente.

O trabalho está organizado da seguinte forma: na Seção 2.2 apresentamos os modelos de simulação, seus resultados e seus defeitos, além das reduções compensatórias de custo marginal e a estimativa de intervalos de confiança para o preço pós-fusão. Na Seção 2.3, definimos o que seriam ganhos de eficiência e, traduzimos em termos econômicos os diversos termos que a literatura antitruste considera ser eficiência. Nas Subseções 2.3.4 e 2.3.5 mostramos o que tem sido feito em termos de estimativa de função custo e função de produção, respectivamente. Por fim, encerramos a Seção 2.3 com uma nota sobre a implementação dos métodos descritos, isto é, quais variáveis devem ser coletadas e como estimar a função custo. A Seção 2.4 conclui o trabalho.

## **2.2 SIMULAÇÃO DE FUSÕES**

O uso de guias de análise (*guidelines*) para determinar os potenciais efeitos anticompetitivos de concentrações horizontais é relativamente recente. Até 1982,

o Departamento de Justiça e o Federal Trade Commission (FTC) — os dois principais órgãos responsáveis pela análise antitruste nos Estados Unidos — não possuíam um procedimento-padrão para analisar fusões. Na realidade, havia muita discricionariedade por parte dos juízes responsáveis pela apreciação dos casos. Esses juízes muitas vezes se valiam de informações como fatia do mercado compreendida pelas firmas como um indicador dos potenciais efeitos anticompetitivos.

Em 1982, esses dois órgãos apresentaram o primeiro guia de análise de fusões. Preparado por advogados e economistas, o *guideline* norte-americano aplica pela primeira vez os conceitos utilizados em organização industrial e teoria dos jogos como uma “receita” de como analisar uma concentração horizontal. Atualizado duas vezes (1984 e 1992), o guia norte-americano serviu de exemplo para grande parte das autoridades antitruste mundiais.

Apesar do grande avanço científico que representou a introdução do guia, o seu uso em casos práticos serviu para apontar muitas das deficiências nas hipóteses por trás de seu modelo. Isso já é observado na primeira etapa da análise, que consiste em determinar o mercado relevante. Em teoria, o mercado relevante é o menor conjunto de produtos substitutos entre si. Para a sua determinação, as autoridades se valiam do teste do monopolista hipotético: qual o menor conjunto de bens susbtitutos àqueles produzidos pelas firmas analisadas que, se fossem ofertados apenas por um monopolista, este acharia lucrativo realizar um aumento substancial e permanente em seus preços.

O teste do monopolista hipotético é bastante intuitivo e relativamente simples de ser implementado, principalmente se os produtos em análise forem homogêneos ou apresentarem um reduzido grau de substitutibilidade. Na prática, poucos são os produtos em que uma análise como esta pode ser feita sem prejuízo da generalidade. Quase todos os produtos apresentam diferenciação e relativa substitutibilidade, de forma que a delimitação de onde acaba o mercado relevante é muito difícil de ser feita. Além disso, a forma como é delimitado o mercado relevante muitas vezes viesa a análise dos efeitos da fusão, uma vez que, quanto mais restrito for o mercado, maior será o poder de mercado nas firmas pleiteantes à fusão e, por consequência, maiores os prejuízos decorrentes dessa fusão. Como é citado por Werden e Froeb (1996), a Suprema Corte norte-americana já observou que “a definição do mercado geralmente determina o resultado do caso”.

Ao observarem esse problema Werden e Froeb (1994) propuseram uma nova ferramenta de análise de fusões. Dado que o objetivo último da análise seria estimar qual o aumento dos preços na indústria após a fusão, por que não utilizar diretamente os valores de preços e quantidades disponíveis, colocá-los em um modelo

de oligopólio e, partindo do equilíbrio pré-fusão, recuperar os parâmetros estruturais e simular qual seria o novo equilíbrio em preços? Desde então, com certas modificações quanto à forma funcional da demanda [Hausman, Leonard e Zona (1994) e Epstein e Rubinfeld (2001)] modelos dessa natureza têm sido utilizados na análise de concentrações horizontais. Temos conhecimento do uso dessas técnicas na instrução de processos nos Estados Unidos, Canadá e na Comissão Européia e, mais recentemente, no Brasil, no caso Nestlé-Garoto. Contudo, não conseguimos informações sobre o número de casos onde essas técnicas foram efetivamente usadas nessas outras legislações, nem sobre qual peso é dado aos resultados dessa técnica na decisão do processo.

### 2.2.1 Modelos de simulação

Seja uma indústria constituída por  $N$  firmas, cada uma produzindo apenas um produto, ou uma marca.<sup>1</sup> Assuma-se que as firmas atuam estratégicamente e a variável de escolha seja preço, isto é, as firmas interagem conforme um jogo de Bertrand:

$$\max_{p_i} (p_i - c_i) \cdot q_i(p) \quad (1)$$

onde  $p_i$  é o preço escolhido pela firma  $i$ ,  $c_i$  é o custo marginal da firma  $i$ , e  $q_i(p)$  é a quantidade demandada do produto produzido pela firma  $i$ . A condição de primeira-ordem do problema da firma fica:

$$q_i(p) + (p_i - c_i) \frac{\partial}{\partial p_i} q_i(p) = 0 \quad (2)$$

Reescrevendo-se essa expressão, em termos das elasticidades, temos:

$$p_i + (p_i - c_i) \epsilon_{ii} = 0 \quad (3)$$

Observe que, se tivermos os valores da elasticidade da demanda pelo produto  $i$ , podemos obter implicitamente o valor do custo marginal da firma  $i$  a partir de (3). Esse ponto é crucial nos modelos de simulação de fusões e será explorado com

1. Essa hipótese pode ser relaxada para o caso de firmas multiproduto.

mais detalhes adiante. Após a fusão entre as firmas  $i$  e  $j$ , a firma resultante passa a resolver o seguinte problema:

$$\max_{p_i, p_j} (p_i - c_i) \cdot q_i(p) + (p_j - c_j) \cdot q_j(p) \quad (4)$$

cujas condições de primeira-ordem serão:

$$q_i(p) + (p_i - c_i) \frac{\partial}{\partial p_i} q_i(p) + (p_j - c_j) \frac{\partial}{\partial p_i} q_j(p) = 0 \quad (5)$$

$$q_j(p) + (p_j - c_j) \frac{\partial}{\partial p_j} q_j(p) + (p_i - c_i) \frac{\partial}{\partial p_j} q_i(p) = 0$$

sendo que as demais firmas continuam a resolver o mesmo problema (1). Dado o sistema de equações que caracteriza o equilíbrio de preços dessa indústria (2) e (5), os algoritmos utilizados na simulação de fusões [incluindo o Antitrust Logit Model (ALM) utilizado pelo Departamento de Justiça norte-americano] seguem as seguintes etapas:

- a) Escolha<sup>2</sup> do número de firmas que competem na indústria, bem como suas participações de mercado (em receita) e preços.
- b) Escolhe-se um sistema de demanda [em geral, utiliza-se *logit* ou Almont Ideal Demand System (AIDS)].
- c) Com os dados de preços e quantidades determina-se a matriz de elasticidades da indústria ou estima-se diretamente o sistema de demanda escolhido.
- d) A partir do equilíbrio pré-fusão (3), inferem-se os custos marginais de cada firma.
- e) Com os dados de custo marginal resolve-se, em função dos preços, o sistema que caracteriza o equilíbrio pós-fusão [(2) para as firmas fora da fusão e (5) para as participantes].

---

2. Aqui, dizemos escolha porque existe uma certa subjetividade na definição da extensão do mercado. Como discutimos adiante, apesar de o modelo permitir a existência de produtos substitutos, não é possível incluir todos os produtos feitos na economia e medir sua substitutibilidade em relação àqueles envolvidos na fusão. Dessa maneira, é necessário que sejam escolhidos os produtos substitutos mais significativos e, uma vez definidos, fazer o levantamento de quais as firmas que produzem esses bens.

Apesar de o procedimento visto anteriormente ser simples, a implementação computacional do mesmo pode ser complexa; afinal, o sistema de equações que caracteriza o equilíbrio pós-fusão é não-linear. De qualquer forma, com dados de preços e participação de mercado das firmas, bem como a matriz de elasticidades-preço dos produtos, implementamos o modelo de simulação de fusões da forma como Crooke, Froeb e Tschantz (2002) apresentam no site [www.antitrust.org](http://www.antitrust.org).<sup>3</sup> As rotinas encontram-se no Apêndice.

A título de ilustração, vamos tomar o mesmo exemplo citado por Crooke, Froeb e Tschantz (2002), que foi apresentado por Jerry Hausman em um caso analisado pelo Federal Communications Commission (FCC). Os autores consideram o mercado de telefonia à distância norte-americano. Segundo eles haveria apenas três grandes firmas: AT&T, MCI e Sprint, as quais detêm aproximadamente 85% do mercado de chamadas à distância. A Tabela 2.1 apresenta, detalhadamente os valores de participação de cada firma.

Para fins da simulação eles consideraram que todas as demais firmas participantes do mercado de telefonia à distância — na Tabela 2.1 foram agregadas em outras — constituem apenas uma firma. Além disso, consideram que o preço da produto — no caso, 1 minuto de ligação — é de 7 centavos de dólar. Por fim, Crooke, Froeb e Tschantz (2002) apresentam a seguinte matriz de elasticidades-preço para esse mercado:

**TABELA 2.1  
ESTRUTURA DO MERCADO UTILIZADO NA SIMULAÇÃO**

Firma	% do mercado
AT&T	63,4
MCI	16,4
Sprint	5,0
Outras	15,2

**TABELA 2.2  
MATRIZ DE ELASTICIDADES DO MERCADO UTILIZADO NA SIMULAÇÃO**

	AT&T	MCI	Sprint	Outras
AT&T	-1,12	0,09	0,03	0,16
MCI	0,50	-1,33	0,06	0,23
Sprint	0,61	0,22	-1,81	0,30
Outras	0,47	0,12	0,04	-1,33

3. Página acessada em dezembro de 2004.

Dados esses valores, foi simulado qual seria o equilíbrio de preços após uma fusão entre MCI e Sprint utilizando os sistemas de demanda Constant Elasticity of Substitution (CES), AIDS e *logit*. Os resultados, utilizando as rotinas presentes no Apêndice, são apresentados na Tabela 2.3.

Apesar da aparente facilidade do método (pelo menos conceitualmente), existem várias hipóteses restritivas no modelo e que interferem diretamente na robustez dos resultados obtidos. São elas:

*a)* A extensão da indústria é determinada pelo número de firmas/produtos escolhidos. De uma certa forma, é como se houvesse uma predeterminação do mercado relevante.

*b)* As firmas escolhem preços num jogo do tipo Bertrand, tanto antes como após a fusão.

TABELA 2.3  
RESULTADOS DA SIMULAÇÃO PARA AS DEMANDAS AIDS, CES E *LOGIT*

	Preço pós	MC pré	Margem pré	Margem pós
<b>AIDS</b>				
AT&T	0.07100	0.00750	0.89286	0.89437
Outras	0.07114	0.01737	0.75188	0.75584
MCI	0.07809	0.01737	0.75188	0.77758
Sprint	0.08348	0.03133	0.55249	0.63474
<b>CES</b>				
AT&T	0.07000	0.00750	0.89286	0.89286
Outras	0.07000	0.01737	0.75188	0.75188
MCI	0.07900	0.01737	0.75188	0.78014
Sprint	0.08976	0.03133	0.55249	0.65102
<i>Logit</i>				
AT&T	0.07012	0.00774	0.88949	0.88968
Outras	0.07003	0.02625	0.62501	0.62514
MCI	0.07117	0.02592	0.62967	0.63578
Sprint	0.07409	0.02884	0.58801	0.61075

c) A forma da demanda é estabelecida.

d) O custo marginal de cada firma é constante.

A sensibilidade dos resultados à especificação da demanda já tem sido discutida na literatura [Crooke, Froeb e Werden (1999)]. Vamos nos fixar principalmente nos pontos *b* e *d*, sobre a hipótese da forma de interação entre as firmas e sobre a escolha da função custo. Pouco há escrito a respeito de como essas duas hipóteses interferem nos resultados,<sup>4</sup> o que consideramos ser tão importante quanto a escolha do sistema de demanda a ser utilizado. A seguir, vamos detalhar como essas hipóteses viesam os resultados da simulação.

#### 2.2.1.1 Interação entre as firmas

O fato de ser usado um modelo de Bertrand na simulação é algo que viesa os resultados. É conhecido na literatura de organização industrial que indústrias que seguem o modelo de Bertrand seriam mais “competitivas” que as que seguem o modelo de Cournot, no sentido de o equilíbrio apresentar preços menores [Carlton e Perloff (2004)]. Em todos os artigos que apresentam modelos dessa natureza, não se discute a relevância do uso de um tipo ou outro na modelagem (a princípio, com os mesmos dados utilizados no ALM seria possível simular um jogo de Cournot, bastando apenas inverter o sistema de demanda).<sup>5</sup>

A fim de demonstrar esse resultado, implementamos a rotina de simulação para resolver o equilíbrio de Cournot para o exemplo já citado, tomando-se o sistema de demanda como sendo CES. Os resultados encontram-se na Tabela 2.4.

TABELA 2.4  
COMPARAÇÃO DOS RESULTADOS DA SIMULAÇÃO PARA OS MODELOS DE BERTRAND E COURNOT

	Bertrand		Cournot	
	MC pré	Preço pós	MC pré	Preço pós
AT&T	0.0075	0.0700	-0.0001	0.0700
MCI	0.0174	0.0790	0.0134	0.0768
Sprint	0.0313	0.0898	0.0301	0.1061
Outras	0.0174	0.0700	0.0122	0.0700

4. Uma excessão é Peters (2003), que discute a *performance* dos modelos de simulação e separa os principais componentes da diferença entre os valores simulados e os valores realizados.

5. Alguns autores mostram que, apesar de o uso da variável estratégica preço ser o mais intuitivo em termos de como se imagina que as firmas tomam suas decisões, modelos em que a variável estratégica seria quantidade fariam mais sentido se fosse considerado que as firmas fazem, em um primeiro estágio, a escolha de capacidade. Kreps e Scheinkman (1983) — para o caso de produtos homogêneos — e Maggi (1999) — para o caso de bens diferenciados — mostram que, se assumirmos que as firmas escolhem primeiro capacidade e depois competem em preços, então o equilíbrio desse jogo dinâmico seria equivalente ao de um jogo de Cournot estático.

Dois pontos devem ser notados na Tabela 2.4. Em primeiro lugar, observe que o custo marginal inferido para a AT&T é negativo. Apesar de, na prática, esse resultado ser inverossímil, o modelo de simulação permite que valores negativos para o custo marginal sejam inferidos, como é observado por Werden, Froeb e Scheffman (2004). Isso ocorre não apenas porque utilizamos o modelo de Cournot para representar a forma de interação estratégica; modelos de Bertrand podem também apresentar resultados semelhantes. Como já apontamos, tanto as hipóteses sobre a forma de concorrência, como a especificação da função custo podem levar a resultados absurdos como esses.

O segundo ponto a ser observado é a significativa diferença entre os preços de equilíbrio pós-fusão para os dois tipos de modelo. Repare que os preços previstos pelo modelo de Cournot são significativamente mais altos que os mesmos preços previstos pelo modelo de Bertrand, o que confirma o resultado teórico previamente mencionado.

A avaliação sobre o melhor modelo de comportamento para a indústria pode ser dividida em duas questões principais: tipo de variável de escolha — preço ou quantidade — e se a interação é dinâmica ou estática. Como dito, basicamente todos os modelos que são usados em simulação de fusões utilizam um modelo estático em que a variável de escolha é preço. Então seria razoável perguntar: a indústria *realmente* se comporta dessa forma? Existe evidência empírica sobre esse tipo de comportamento?

Desde os anos 1970, a literatura empírica de organização industrial tem buscado identificar qual a forma de interação estratégica em que as firmas se engajam a partir de dados de demanda. Sob o nome de Nova Organização Industrial Empírica [New empirical Industrial Organization (Neio)], esses artigos tipicamente partem da condição de primeira-ordem do oligopolista

$$P + \theta_i q_i \frac{\partial P}{\partial Q} = MC_i(q_i) \quad (6)$$

onde  $P$  é a demanda da indústria,  $Q$  é a demanda da indústria e  $\theta_i$  é um parâmetro de conduta ou, como gostam de se referir os autores,  $\theta_i$  seria uma medida das “conjecturas” da firma  $i$  a respeito da estratégia das demais competidoras [Bresnahan (1981)]. Bresnahan (1989) apresenta uma boa revisão sobre essa metodologia e os artigos que a adotam.

À primeira vista, (6) define um modelo estrutural e usando-se dados de demanda, custo e instrumentos — para lidar com o problema de endogeneidade —

o parâmetro  $\theta$  poderia ser identificado e representaria essa medida da estratégia. Entretanto, Lindh (1992), Corts (1999) e mais recentemente Reiss e Wolak (2005) mostram que não só as hipóteses para identificação de  $\theta$  são muito fortes como o próprio parâmetro  $\theta$  não tem o significado que é apresentado.

Dentre os modelos teóricos conhecidos em organização industrial, o parâmetro  $\theta$  assume apenas alguns valores específicos:  $\theta=0$ , se a indústria for perfeitamente competitiva;  $\theta=1/N$ , se a indústria se comporta de acordo com Cournot (monopólio, caso haja apenas uma firma). Entretanto, vários autores estimam  $\theta$  como sendo um parâmetro que assume valores no contínuo. Bresnahan (1981) e depois Robson (1983) apresentaram modelos onde  $\theta$  representaria as crenças — conjecturas — das firmas a respeito do comportamento das demais e assim teríamos um arcabouço teórico para justificar esse procedimento empírico.

Lindh, por sua vez, mostrou que, em equilíbrio, essas conjecturas — diferentes daquelas resultantes de Cournot, monopólio e Stackelberg — não conseguem ser obtidas como resultado de equilíbrio para qualquer especificação de custo e demanda. Na realidade, ele mostra que essas conjecturas, se existirem, não poderiam ser identificadas (como  $\theta$  é resultado de um sistema de equações diferenciais parciais, várias formas de interação estratégica entre as firmas poderiam gerar o mesmo valor de  $\theta$ ) e se puderem ser identificadas seria porque estariam impondo que o espaço de estratégias das firmas fosse dependente da forma funcional do custo e da demanda, uma hipótese que não é trivial na teoria da firma. Assim, qual seria o significado de  $\theta=1,5$ ? Reiss e Wolak e mesmo Lindh acreditam que  $\theta$  poderia assumir valores no contínuo a partir de modelos dinâmicos de formação de crenças. Entretanto, eles mesmos apontam que se for o caso,  $\theta$  só poderia ser interpretado como um parâmetro estrutural se fosse gerado a partir de um modelo estrutural. Da forma como é apresentado em (6),  $\theta$  é apenas uma medida da nossa ignorância a respeito da forma de interação estratégica das firmas.

Esse ponto nos leva a outra controvérsia sobre a validade do uso de modelos de simulação: pode um jogo estático simular, com relativa precisão, os resultados de uma fusão ou seria necessário modelar o processo como um jogo dinâmico? Berry e Pakes (1993) argumentam que não. Eles compararam os resultados de oligopólio onde a variável estratégica é quantidade usando dois tipos de modelos: um onde a interação entre as firmas é estática e outro em que as firmas competem em um jogo dinâmico. O primeiro caso é analisado em Salant, Switzer e Reynolds (1983) que mostram que, em um jogo estático, nenhuma firma teria incentivo à fusão, pois a redução em quantidade por parte da firma fusionada seria compensada com o aumento das rivais e os lucros seriam menores. Cheong e Judd (2001)

analisam o problema dinâmico. Eles mostram, que ao se tomar o mesmo modelo em um ambiente dinâmico com custos de ajustamento, os resultados podem ser opostos aos previstos no modelo estático, isto é, fusões podem ser lucrativas mesmo quando as firmas competem em quantidades. Nesse caso, a firma fusionada aumentaria o valor presente dos seus lucros na trajetória até o novo *steady state*. Resultados dessa natureza também surgiram em modelos como o de Pakes e McGuire (1994) ou Ericson e Pakes (1995) e mais recentemente Gowrisankaran (1999).

Apesar de trazer mais realismo à modelagem, o uso de jogos dinâmicos na simulação de fusões ainda não é uma realidade. Em primeiro lugar porque apenas recentemente têm surgido artigos que se propõem a estimar os parâmetros estruturais dessa classe de modelos [por exemplo, Bajari, Benkard e Levin (2004)], o que é essencial para realizar as simulações do novo equilíbrio da indústria. Em segundo — e relacionado ao fato de só agora estimações desses modelos começarem a existir —, esses modelos dinâmicos apresentam múltiplos equilíbrios, o que dificulta imensamente a identificação dos parâmetros de interesse. Isso sem contar que uma das hipóteses mais discutidas em simulação — a manutenção da mesma forma de interação estratégica após a fusão —, ainda não foi contemplada em modelos teóricos, portanto é algo que a teoria ainda não pode ajudar na prática.

### **2.2.1.2 Especificação da função custo**

Como já foi apontado, nos modelos de simulação atualmente utilizados, uma das principais hipóteses é a de custo marginal constante, isto é, a função custo é linear em quantidades.<sup>6</sup> Um dos principais fatores que dá força a essa hipótese é a crença — largamente difundida entre os autores da Neio — que o custo é não-observável ou que, mesmo se for observável, os erros na sua medida invalidariam os resultados obtidos com a inclusão desses valores. Realmente, como dados de demanda em geral são públicos e dados de oferta são de propriedade das firmas, a estimativa direta da função custo ou da função de produção torna-se uma tarefa muito difícil. Assim sendo, o melhor procedimento para obtenção do custo de produção seria inferi-lo implicitamente a partir de dados de equilíbrio da indústria, das hipóteses sobre a demanda e sobre a forma de competição entre as firmas. É aqui que entra a hipótese sobre custos marginais constantes: ao inferir indiretamente o custo marginal, pode-se mostrar que o modelo consegue identificar apenas o custo marginal total. Dessa forma, a hipótese de custo marginal constante permitiria

---

6. Essa hipótese pode fazer sentido ou não, dependendo da indústria em análise, como em qualquer estimação estrutural. O argumento mais utilizado para justificá-la seria que, no intervalo de valores observado, a função custo seria bem aproximada por uma reta e que, no curto prazo, a tecnologia não se alteraria o suficiente para que o custo marginal se alterasse. Isso pode ser válido para muitas indústrias, mas deve ser bem discutido pelo analista antes de aceitar a hipótese de custos marginais constantes.

extrapolar a curva de custo da firma para outros valores de quantidade produzida, que é interessante na criação de contrafactuals (que é o caso da simulação).

Genesove e Mullin (1998) apresentam um exercício interessante: compararam os valores de custo marginal, inferido pelos dados de equilíbrio da indústria de açúcar refinado norte-americana com os valores reportados de custo marginal.<sup>7</sup> Eles utilizam uma especificação linear para o custo marginal, do tipo

$$C = c_0 + kP_{\text{raw}} \quad (7)$$

onde  $P_{\text{raw}}$ , representa o preço do açúcar bruto e  $c_0$  consolidaria todos os outros custos envolvidos no refinamento do açúcar.

Eles então estimam o valor de  $c_0$  e  $k$  usando uma especificação linear para a demanda de açúcar refinado, (7) como sendo a forma funcional do custo marginal e usam também a condição de equilíbrio (6), comum na literatura Neio. O resultado é que o valor estimado para  $c_0$ , usando-se a técnica de inferir os custos marginais a partir dos dados de equilíbrio de mercado, é bem superior ao valor “verdadeiro” do mesmo, obtido a partir dos estudos da época. É verdade que  $c_0$  representa apenas 7,5% do custo marginal mas, de acordo com Corts (1999), esses problemas podem ser ainda maiores em indústrias mais concentradas, onde  $\theta$  for alto.<sup>8</sup>

A hipótese de custo marginal constante afeta não apenas a inferência do valor do custo marginal (que pode até assumir valores negativos, como mostrado no exemplo de Cournot), como também os resultados da simulação. Assuma-se uma especificação mais geral para a função custo, como por exemplo  $C(q)$ , onde  $q$  representa a quantidade produzida de cada um dos produtos. Essa é uma especificação trivial de funções custo de firmas multiproduto e, como será visto nas seções seguintes, é assumida por vários autores na estimativa de custos, sendo o caso mais popular a especificação translog.

Observe que o sistema de equações descrito por (3) e (5) passa a ser escrito como:

$$q_i(p) + \left( p_i - \frac{\partial}{\partial q_i} C_n(q) \right) \frac{\partial}{\partial p_i} q_i(p) = 0 \quad (8)$$

7. Os autores se baseiam nos depoimentos de produtores de açúcar no começo do século XX e estudos a respeito da indústria. Apesar de o valor do custo fixo ser controverso, eles revelam que a estimativa do custo marginal é precisa e que a tecnologia é comum a todas as firmas.

8. Se considerarmos o modelo de Cournot,  $\theta = \frac{1}{N}$ , logo, quanto menor o número de participantes no mercado, maior será o viés na estimativa.

$$q_i(p) + \left( p_i - \frac{\partial}{\partial q_i} C_f(q) \right) \frac{\partial}{\partial p_i} q_i(p) + \left( p_j - \frac{\partial}{\partial q_j} C_f(q) \right) \frac{\partial}{\partial p_i} q_i(p) \quad (9)$$

$$q_j(p) + \left( p_j + \frac{\partial}{\partial q_j} C_f(q) \right) \frac{\partial}{\partial p_j} q_j(p) + \left( p_i - \frac{\partial}{\partial q_i} C_f(q) \right) \frac{\partial}{\partial p_j} q_j(p) = 0$$

onde  $C_f$  representa o custo total da firma fusionada e  $C_n$  o custo total das demais firmas. Essa diferença nos índices é proporcional, uma vez que estamos permitindo que as firmas sejam multiproduto. Logo, os índices  $n$  e  $f$ , são indicativos de quem detém os direitos de propriedade, enquanto os índices  $i$  e  $j$ , referem-se aos produtos propriamente ditos.

É fácil perceber que (8) e (9) representam um sistema de equações diferenciais parciais não-linear, o qual não necessariamente vai apresentar a mesma solução que o sistema representado em (3) e (5). A diferença aqui está no fato de que ao permitir que o custo marginal varie com a quantidade produzida e com a cesta de produtos ofertada, as firmas disponham de um conjunto de possibilidades muito maior para se diferenciar após a fusão.

Todavia, o arcabouço do ALM ou qualquer outro modelo de simulação que se baseie exclusivamente em dados de demanda não permite a inferência de funções custo mais elaboradas. Dessa forma, sem o auxílio de funções custo estimadas econometricamente, não é possível obter resultados de simulação para outros tipos de especificação de custos.

É importante deixar claro que reconhecemos a dificuldade da estimação direta da função custo, seja por causa da disponibilidade de dados, seja pelo tipo de técnica necessária para a estimativa dos parâmetros. Sem dúvida, pode ser um procedimento impeditivo em termos de tempo de análise. Não há uma resposta certa se o benefício dessa análise mais completa seria maior que o seu custo. No caso americano os documentos do Departamento de Justiça comentam muito pouco sobre a estimativa direta da função custo, talvez porque os prazos de análise nos Estados Unidos sejam muito curtos e assim eles preferiram não estimá-las. O que podemos afirmar é que, uma vez conseguidos os dados necessários (os quais serão apresentados na seção correspondente à estimativa da função custo), a estimativa em si é simples e rápida, podendo ser utilizados pacotes estatísticos comerciais (como Stata). Entendemos que o problema pode estar na obtenção desses dados,

a qual não podemos dizer qual seria o custo aproximado (em termos de tempo) para obtê-los e assim afirmar se a estimativa da função custo é viável.

### 2.2.2 Reduções compensatórias de custo marginal

Modelos de simulação servem não apenas para dizer quais seriam os novos preços de equilíbrio após uma fusão; eles também podem ser utilizados para responder a seguinte pergunta: qual seria a redução no custo marginal das firmas fusionadas para que os preços da indústria não fossem alterados? Essa é a idéia por trás do cálculo das reduções de custo marginal compensatórias [Compensating Marginal Cost Reductions (CMCR)].

O uso do CMCR é defendido por muitos por apresentar a desejável propriedade de ser *independente da forma funcional da demanda*. Werden (1996) apresenta a prova desse resultado. Mostraremos aqui que esse resultado não está correto, isto é, o CMCR depende da forma funcional da demanda.

Vamos inicialmente considerar o caso onde o custo marginal é constante. Aqui, as equações que definem o equilíbrio são as mesmas que obtivemos na Subseção 2.2.1, isto é,

$$q_i + (p_i - c_i) \frac{\partial q_i}{\partial p_i} = 0 \quad (10)$$

para cada uma das firmas antes da fusão e

$$q_i + (p_i - c_i) \frac{\partial q_i}{\partial p_i} + (p_j - c_j) \frac{\partial q_j}{\partial p_i} = 0 \quad (11)$$

$$q_j + (p_j - c_j) \frac{\partial q_j}{\partial p_j} + (p_i - c_i) \frac{\partial q_i}{\partial p_j} = 0$$

para as firmas fusionadas [sendo que as firmas que não participam da fusão continuam obedecendo (10)]. Aqui, apesar de escrevermos apenas  $q_i$ , continuamos a considerar que a demanda de cada produto  $i$  depende do vetor de preços da indústria, ou seja,  $q_i = q_i(p)$ .

Pela definição do CMCR, os preços antes e após a fusão serão os mesmos. A única variável que temos, portanto, passa a ser o custo marginal de cada firma. Werden (1996) reescreve a condição de primeira-ordem das firmas antes da fusão como função da elasticidade própria

$$\frac{p_i - c_i}{p_i} \frac{\partial q_i}{\partial p_i} \frac{p_i}{q_i} = -1 \quad (12)$$

$$m_i^0 = -\frac{1}{\epsilon_{ii}^0}$$

onde  $m_i^0$  representa a margem da firma  $i$  antes da fusão. As condições de primeira-ordem para as firmas fusionadas também podem ser escritas como função das elasticidades próprias e cruzadas, da seguinte forma:

$$\frac{p_i - c_i}{p_i} \frac{\partial q_i}{\partial p_i} \frac{p_i}{q_i} + \frac{p_j - c_j}{p_j} \frac{\partial q_j}{\partial p_i} \frac{p_i}{q_j} \frac{p_j q_j}{p_i q_i} = -1 \quad (13)$$

$$\frac{p_j - c_j}{p_j} \frac{\partial q_j}{\partial p_j} \frac{p_j}{q_j} + \frac{p_i - c_i}{p_i} \frac{\partial q_i}{\partial p_j} \frac{p_j}{q_i} \frac{p_i q_i}{p_j q_j} = -1$$

$$m_i^1 \epsilon_{ii}^1 + m_j^1 \epsilon_{ji}^1 \frac{p_j q_j}{p_i q_i} = -1 \quad (14)$$

$$m_j^1 \epsilon_{jj}^1 + m_i^1 \epsilon_{ij}^1 \frac{p_i q_i}{p_j q_j} = -1$$

$$m_i^1 - m_j^1 d_{ji} \frac{p_j}{p_i} = -\frac{1}{\epsilon_{ii}^1} \quad (15)$$

$$m_j^1 - m_i^1 d_{ij} \frac{p_i}{p_j} = -\frac{1}{\epsilon_{jj}^1}$$

onde  $d_{ij} = \frac{\varepsilon_{ij}^1 q_i}{\varepsilon_{jj}^1 q_j}$ , também chamada de *diversion ratio*. Essa quantidade captura a fração de mercado que é perdida por um dos produtos da firma fusionada, que é capturada pelo outro produto feito pela firma fusionada quando o preço daquele primeiro produto aumenta. Resolvendo-se o sistema anterior em termos das margens pós-fusão, obtemos:

$$m_j^1 = \left[ -\frac{p_i}{p_j} \frac{d_{ij}}{\varepsilon_{ii}^1} - \frac{1}{\varepsilon_{jj}^1} \right] \frac{1}{1 - d_{ij} d_{ji}} \quad (16)$$

$$m_i^1 = \left[ -\frac{1}{\varepsilon_{ii}^1} - \frac{p_j}{p_i} \frac{d_{ji}}{\varepsilon_{jj}^1} \right] \frac{1}{1 - d_{ij} d_{ji}}$$

Pela definição do CMCR, os preços antes e depois da fusão são os mesmos. Isso implica dizer que o valor das elasticidades pós-fusão é o mesmo que calculamos antes da fusão, pois a elasticidade nada mais é do que o valor da inclinação da curva de demanda para cada preço (lembrar que a elasticidade é uma medida local). Como o ponto onde estamos avaliando essa derivada não mudou (o preço permaneceu o mesmo), assim também não mudou o valor da derivada naquele ponto. Portanto, podemos escrever  $\varepsilon_{ij}^0 = \varepsilon_{ij}^1 = \varepsilon_{ij}$ , para quaisquer produtos  $i$  e  $j$  da indústria. Usando-se (15) é fácil reescrever as margens pós-fusão como função das margens pré-fusão:

$$m_j^1 = \left[ m_i^0 d_{ij} \frac{p_i}{p_j} + m_j^0 \right] \frac{1}{1 - d_{ij} d_{ji}} \quad (17)$$

$$m_i^1 = \left[ m_i^0 + m_j^0 d_{ji} \frac{p_j}{p_i} \right] \frac{1}{1 - d_{ij} d_{ji}}$$

Finalmente, escrevemos o CMCR como função das margens antes e após a fusão da seguinte forma:

$$\begin{aligned}
 CMCR_i &= \frac{c_i^0 - c_i^1}{c_i^0} \\
 &= \frac{p_i - p_i + c_i^0 - c_i^1}{p_i - p_i + c_i^0} \\
 &= \frac{p_i - c_i^1 - (p_i - c_i^0)}{p_i} \frac{p_i}{p_i - (p_i - c_i^0)} \\
 &= \frac{m_i^1 - m_i^0}{1 - m_i^0}
 \end{aligned} \tag{18}$$

Substituindo-se a expressão da margem pós-fusão obtida em (16) temos:

$$\begin{aligned}
 CMCR_i &= \left[ m_i^0 + m_j^0 d_{ji} \frac{p_j}{p_i} \right] \frac{1}{(1 - m_i^0)(1 - d_{ij} d_{ji})} - \frac{m_i^0}{1 - m_i^0} \\
 &= \left[ m_i^0 d_{ij} + m_j^0 \frac{p_j}{p_i} \right] \frac{d_{ji}}{(1 - m_i^0)(1 - d_{ij} d_{ji})}
 \end{aligned} \tag{19}$$

Werden argumenta, então, que a expressão (19) pode ser calculada independentemente do sistema de demanda escolhido: as *diversion ratio* seriam estimadas a partir de dados sobre preço e quantidade vendidas no varejo (tais como os dados obtidos por institutos como Nielsen) e as margens pré-fusão poderiam ser inferidas por dados de balanço ou pelas condições de primeira ordem antes da fusão. Cabe aqui a pergunta: considerando-se que essas quantidades (*diversion ratio* e margem pré-fusão) não são diretamente observáveis, isto é, precisam ser estimadas, podemos inferi-las sem assumir alguma forma específica para a demanda? A resposta é não.

Comecemos pelas margens pré-fusão. Werden indica duas possíveis formas de inferi-lo: a partir de dados contábeis ou a partir das condições de primeira ordem antes da fusão. A primeira opção (dados contábeis) realmente não necessita de hipóteses sobre a forma da demanda. Como os preços pré-fusão são observáveis, bastaria calcular o custo marginal pré-fusão. Entretanto, calcular o custo marginal

não é uma tarefa simples. Com os dados contábeis é possível calcular o custo médio de produção, o qual pode ser bastante diferente do custo marginal de produção, que é aquele em que estamos interessados. Portanto, o uso de dados contábeis não nos isenta do erro de medida, que pode ser alto.

Além disso, como veremos na seção que trata da estimação da função custo, existem vários problemas relacionados à estimação dessa função, inclusive o problema de escolha de sua forma funcional. O uso de técnicas não-paramétricas requer uma grande disponibilidade de dados o que, em se tratando de dados sobre custo, não é factível. Dessa forma, no que diz respeito à estimação das margens, troca-se a dependência da forma funcional da demanda pela forma funcional do custo. Não há estudos que digam qual dos dois seria mais robusto.

Se o analista preferir inferir os custos a partir das condições de primeira-ordem pré-fusão então será necessário calcular as elasticidades próprias pré-fusão, como fica claro a partir de (12). Todavia, *a estimação das elasticidades pré-fusão requer a especificação de alguma forma paramétrica para a demanda e os valores estimados serão dependentes dessa forma funcional escolhida*. Tome, por exemplo, o caso da demanda linear

$$\ln q_i = \alpha_i + \sum_{j=1}^N \beta_{ij} \ln p_j + \varepsilon_i \quad (20)$$

Nesse caso, assumindo-se que temos instrumentos necessários para eliminar o problema de simultaneidade, a elasticidade própria seria dada por  $\varepsilon_{ii} = \beta_{ii}$ . Se, por outro lado, assumirmos uma especificação translog para a demanda

$$\ln q_i = \alpha_i + \sum_{j=1}^N \beta_{ij} \ln p_j + \sum_{k=1}^N \sum_{j=1}^N \gamma_{ijk} \ln p_k \ln p_j + \varepsilon_i \quad (21)$$

Dessa forma, a elasticidade própria seria dada por:

$$\varepsilon_{ii} = \beta_{ii} + \sum_{j=1}^N \gamma_{iij} \ln p_j \quad (22)$$

A menos que o analista restrinja  $\gamma_{iij} = 0$  para todo  $j$ , as duas elasticidades serão diferentes! Isso vale não só para esse exemplo, como para qualquer outro

sistema de demanda que for considerado. Mesmo se quisermos estimar não-parametricamente essa elasticidade, isto é, se quiséssemos estimar

$$q_i = f_i(p, \varepsilon_i) \quad (23)$$

onde  $p$  é o vetor de preços e  $f_i(\cdot)$  é a função que define a demanda pelo produto  $i$ ; a estimação não-paramétrica dessa função requereria quantidades muito grandes de dados. Basicamente, seria preciso observar as quantidades demandadas para muitos valores de preços, de forma a traçar a curva da demanda. Isso na maioria das aplicações não é factível. Além disso, mesmo se tivéssemos essa quantidade de dados, os resultados seriam dependentes do número de produtos incluídos como sendo substitutos, isto é, nada garante que a função demanda estimada para o vetor de preços  $p$  seja igual a outra função de demanda estimada para o vetor de preços  $\hat{p}$ , onde  $p \subseteq \hat{p}$ . Assim, no tocante à estimação das margens, o uso das condições de primeira-ordem não elimina o problema da dependência da forma funcional da demanda.

O mesmo problema acontece na determinação do *diversion ratio*. A forma mais comum de estimação dessa quantidade é via estimação paramétrica do sistema de demanda. No exemplo da demanda linear, o *diversion ratio* seria dado por:

$$d_{ij} = -\frac{\beta_{ij}}{\beta_{jj}} \frac{q_i}{q_j} \quad (24)$$

Novamente, diferentes formas paramétricas para o sistema de demanda poderão implicar diferentes valores estimados para o *diversion ratio*. Além disso, como a elasticidade cruzada não aparece nas condições de primeira-ordem pré-fusão, não é possível inferi-la usando a estimação do custo (como é o caso da elasticidade própria). Mesmo que se opte por métodos não-paramétricos, os valores estimados serão dependentes da extensão da indústria considerada, isto é, os valores estimados vão depender do número de produtos escolhidos como sendo possíveis substitutos. E isso é uma certa forma de restrição no sistema de demanda. Repare: mesmo que o analista consiga estimar as margens pré-fusão usando dados de custo, o *diversion ratio* continuará sendo dependente da especificação da demanda. Assim, o CMCR, bem como os resultados de simulação dos preços pós-fusão, não é robusto à especificação da demanda.

Um dos resultados das rotinas desenvolvidas é o cálculo do CMCR. Os resultados, para o exemplo do mercado de telefonia norte-americano, são apresentados na Tabela 2.5.

TABELA 2.5  
RESULTADO DO CMCR PARA O EXEMPLO CONSIDERADO NA SUBSEÇÃO 2.2.1

	MC pré	MC pós	CMCR
CES			
MCI	0.0174	0.0151	-0.1296
Sprint	0.0313	0.0254	-0.1905
<i>Logit</i>			
MCI	0.0259	0.0245	-0.0545
Sprint	0.0288	0.0245	-0.1501

Repare que os valores do CMCR para as especificações CES e *logit* são distintos. Isso decorre do fato de que enquanto a matriz de elasticidades pré-fusão para o sistema CES for igual àquela presente na Tabela 2.2, a matriz de elasticidades para o sistema *logit* é distinta: na realidade, para o exemplo em questão, ela é derivada a partir de um valor de elasticidade da indústria  $\varepsilon = 0.763$  e o parâmetro  $\beta = 25$ . Usando-se as expressões para as elasticidades presentes em Werden e Froeb (1996) para o sistema *logit*

$$\varepsilon_i = \frac{p_i}{\bar{p}} (\beta \bar{p} (1 - s_i) + \varepsilon s_i) \quad (25)$$

$$\varepsilon_{ij} = \frac{s_j p_j}{\bar{p}} (\beta \bar{p} - \varepsilon)$$

obtemos a seguinte matriz de elasticidades para o sistema *logit* a qual é bastante diferente daquela presente na Tabela 2.2.

TABELA 2.6  
MATRIZ DE ELASTICIDADES DO MERCADO PARA A DEMANDA *LOGIT*

	AT&T	Outras	MCI	Sprint
AT&T	-1,12	0,15	0,16	0,05
Outras	0,63	-1,60	0,16	0,05
MCI	0,63	0,15	-1,60	0,05
Sprint	0,63	0,15	0,16	-1,70

Uma vez que o CMCR padece do mesmo problema que a simulação de preços, cabe a pergunta: qual a vantagem no seu uso? Se o analista optar pela estimativa da função custo e depois estimar o *diversion ratio*, a análise passa a ser dependente de duas formas funcionais: a forma da função custo e a da demanda. Qual a direção do viés, usando essa abordagem, é uma questão em aberto. Todavia, existe mais uma questão relacionada ao uso do CMCR: como dizer que o valor calculado representa maior ou menor risco de exercício de poder de mercado? O CMCR nos diz qual seria a redução em custos marginais decorrentes da fusão necessária para manter os preços nos níveis anteriores à mesma. Entretanto, para saber se esse valor é factível ou não, é preciso estimar qual seria essa redução em custos, que é o propósito do cálculo das eficiências decorrentes da operação. Sem estimar essa eficiência, o CMCR passa a ser um número desprovido de utilidade para política antitruste: dizer que o CMCR é de 15% pode tanto ser um número muito alto (se o possível ganho de eficiência for muito baixo), como pode ser um número facilmente alcançável (se houver grande sinergia entre as fusionadas). Portanto, voltamos a depender de algum tipo de estimativa das eficiências para fazermos alguma comparação, o que nos faz retornar ao ponto de precisarmos estimar ou a função custo ou a função de produção.<sup>9</sup>

### **2.2.3 Intervalos de confiança**

Como pode ser observado nos artigos seminais de Werden e Froeb (1994) e depois nos artigos de Hausman, Leonard e Zona (1994), bem como Nevo (2000), a parte de simulação de preços do equilíbrio pós-fusão limita-se à determinação de estimadores de ponto. Apesar de ser uma informação relevante, não há como verificar a significância desse resultado sem uma medida de dispersão. Por outro lado, a natureza altamente não-linear dos estimadores de ponto impede o uso de técnicas como o método delta para a determinação do desvio-padrão dos estimadores.

Para resolver esse problema, Capps Jr., Church e Love (2003) lançam mão de meios não-paramétricos para determinar o intervalo de confiança. Mais especificamente, eles utilizam a técnica de *bootstrap* para esse fim e apresentam seus resultados para dois sistemas de demanda — Aids e Rotterdam.

---

9. Poderia ser argumentado o uso de formas tecnológicas para a medida de eficiência da operação. Por exemplo, usando-se pareceres técnicos de engenheiros sobre qual seria o ganho de eficiência da nova firma. Entretanto, como abordamos na seção correspondente às eficiências, pareceres técnicos permitiriam apenas calcular as eficiências tecnológicas, as quais são apenas uma parte da possível economia em custos devido à operação: após a fusão, a firma enfrenta um novo problema de minimização de custos, que pode ser devido tanto à mudança na tecnologia como nos preços dos insumos (preços não-lineares, poder de barganha etc.). Logo, pareceres técnicos possivelmente não teriam o poder de predizer todo o ganho de eficiência devido à operação.

Vamos apenas apresentar o sistema de demanda Rotterdam. Esse modelo de demanda não é derivado a partir de algum tipo de função utilidade ou despesa, apenas satisfaz condições de simetria e homogeneidade. Basicamente, podemos expressar a demanda como sendo:

$$w_i \ln\left(\frac{q_{it}}{q_{it-1}}\right) = a_i + b_i \sum_{j=1}^N w_j \ln\left(\frac{q_{jt}}{q_{jt-1}}\right) + \sum_{j=1}^N c_{ij} \ln\left(\frac{p_{jt}}{p_{jt-1}}\right) \quad (26)$$

As hipóteses de homogeneidade e simetria no sistema de demanda implicam as seguintes restrições nos parâmetros:

$$\sum_{i=1}^N a_i = 0, \quad \sum_{j=1}^N c_{ij} = 0, \quad \sum_{i=1}^N b_i = 1 \quad (27)$$

$$c_{ij} = c_{ji}$$

A partir de (26) é imediato obter as elasticidades marshalianas para esse sistema de demanda:

$$\epsilon_{ij} = \frac{c_{ij} - w_j b_i}{w_i} \quad (28)$$

Uma vez definido o sistema de demanda, a simulação acontece da mesma forma apresentada na Subseção 2.2.1: resolve-se o equilíbrio pré-fusão para obter implicitamente o custo marginal dos participantes do mercado (assumindo-se que esse custo marginal é constante) e depois resolve-se o sistema de condições de primeira-ordem que definem o equilíbrio pós-fusão em termos dos novos preços.

A inovação está no *bootstrap*. Resumidamente, *bootstrap* consiste na estimação da distribuição de uma estatística a partir do sorteio repetido da amostra original. Existe uma larga literatura sobre o tema.<sup>10</sup> Vamos inicialmente discutir como seria a implementação do *bootstrap* para o problema de simulação e depois discutir os possíveis problemas decorrentes de particularidades das séries de demanda.

---

10. Uma boa resenha sobre o assunto está em Horowitz (2001).

Como já apontamos na Subseção 2.2.1, a simulação começa pela estimação do sistema de demanda. Então, vamos supor um sistema de demanda muito simples, no caso, um sistema de demanda linear

$$\ln q_{jt} = \alpha_j + \sum_{i=1}^N \beta_{ji} \ln p_{it} + \varepsilon_{jt} \quad (29)$$

onde  $\varepsilon_{jt}$  pode apresentar correlação entre os produtos, mas não apresenta correlação temporal, isto é,  $\varepsilon\varepsilon'$  é bloco-diagonal.

Estimar a equação (29) não é simples, uma vez que preços e quantidades são determinados em equilíbrio. A forma mais adequada de resolver esse problema de simultaneidade está na utilização de variáveis instrumentais para contornar a endogeneidade dos preços. Entretanto, como apontam Capps Jr., Church e Love (2003) dados de varejo apresentam uma freqüência relativamente alta (mensal ou semanal) e é muito difícil encontrar instrumentos, como por exemplo dados de custo, com essa mesma freqüência. A forma que os autores encontraram para resolver esse problema foi assumir que os preços são fixados pelos varejistas antes que os consumidores decidam pela sua compra. Assim, eles estão assumindo que, no instante  $t$ , o preço observado é exógeno. Essa hipótese é bastante forte e pode gerar vieses tanto na determinação do intervalo de confiança, como também do estimador de ponto. Por outro lado, ela facilita muito a estimativa do sistema e a implementação do *bootstrap*. No momento, vamos assumir que a hipótese é válida para simplificar o entendimento da técnica.

Utilizando mínimos quadrados, o econometrista consegue estimar  $\hat{\alpha}$ ,  $\hat{\beta}$  e a série de resíduos  $\hat{\varepsilon} = \ln q_{jt} - \hat{\alpha}_j - \sum \hat{\beta}_{ji} \ln p_{it}$ . Se o modelo que estamos utilizando corresponde ao processo gerador de dados, então esses resíduos são estimadores não-viesados dos erros  $\varepsilon$ . Considerando-se que esses erros não apresentam correlação temporal, a série de resíduos que observamos corresponde a  $T$  realizações desta variável aleatória, cada uma delas com igual probabilidade de ter ocorrido. Por outro lado, a série original que observamos representa uma realização particular: o erro  $\varepsilon_1$  ter ocorrido na primeira posição, o erro  $\varepsilon_2$  ter ocorrido na segunda posição, e assim por diante. A idéia do *bootstrap* seria: como cada um desses erros tem probabilidade igual de ser sorteado em cada uma das posições, tomam-se novamente os erros estimados (resíduos) e sorteia-se novamente a ordem em que cada um vai aparecer na série. Agora temos uma nova série de resíduos, a série de

preços (suposta como exógena) e os valores estimados dos parâmetros. Assim, é imediato o cálculo das pseudoquantidades  $\tilde{q}_{ji}$ :

$$\ln \tilde{q}_{ji} = \hat{\alpha}_j + \sum_{i=1}^N \hat{\beta}_{ji} \ln p_{it} + \tilde{\varepsilon}_{ji} \quad (30)$$

onde  $\tilde{\varepsilon}$  representa a série de resíduos sorteada. Uma vez com essa pseudobase de dados em mãos, podemos novamente calcular  $\hat{\alpha}$  e  $\hat{\beta}$  para essa nova série, da mesma forma que fizemos para a base de dados original.

Esse processo de criação de pseudobases de dados e estimação dos parâmetros pode ser repetido inúmeras vezes, sempre utilizando a série original dos resíduos e os valores originais de  $\hat{\alpha}$  e  $\hat{\beta}$  calculados a partir dos dados originais. Isso vai permitir a construção de uma série de valores para  $\hat{\alpha}$  e  $\hat{\beta}$ . Uma vez com essa série de valores, a variância dos parâmetros será:

$$\sigma_{\alpha}^2 = \sum_{r=1}^R \frac{(\hat{\alpha}_j^r - \hat{\alpha}_j)^2}{R-1} \quad (31)$$

$$\sigma_{\beta}^2 = \sum_{r=1}^R \frac{(\hat{\beta}_{ji}^r - \hat{\beta}_{ji})^2}{R-1}$$

onde:

$$\hat{\alpha}_j = \sum_{r=1}^R \frac{\hat{\alpha}_j^r}{R} \quad (32)$$

$$\hat{\beta}_{ji} = \sum_{r=1}^R \frac{\hat{\beta}_{ji}^r}{R}$$

Se o interesse estivesse apenas na estimação do sistema de demanda, poderíamos encerrar por aqui. Entretanto, o nosso objetivo é estimar um intervalo de confiança para os preços simulados pós-fusão. Como vimos na Subseção 2.2.1, a simulação

depende dos valores dos parâmetros do sistema de demanda (nesse caso,  $\hat{\alpha}$  e  $\hat{\beta}$ ). Agora, como temos uma série de valores de parâmetros podemos realizar a simulação do preço de equilíbrio pós-fusão para cada um dos valores  $\hat{\alpha}'$  e  $\hat{\beta}'$ . Isso, logicamente, significa que agora teremos uma série com  $R$  valores dos preços pós-fusão. Com essa série podemos calcular a variância desses preços da mesma forma que calculamos a variância dos parâmetros do sistema de demanda, isto é, a variância do preço do produto  $j$  após a fusão entre duas firmas  $i$  e  $k$  pode ser escrita como:

$$\sigma_{\hat{p}_j}^2 = \sum_{r=1}^R \frac{(\hat{p}_j^r - \hat{p}_j)^2}{R-1} \quad (33)$$

onde:

$$\hat{p}_j = \sum_{r=1}^R \frac{\hat{p}_j^r}{R} \quad (34)$$

Uma vez entendida a implementação do método para o problema específico da simulação, cabe analisarmos se as hipóteses utilizadas no artigo de Capps Jr., Church e Love (2003) podem ser estendidas aos problemas de interesse. Ignorando-se o fato de os resultados serem sensíveis à forma funcional da demanda escolhida, podemos apontar duas hipóteses fortes utilizadas pelos autores:

- a) a correlação serial presente nos dados de demanda pode ser corrigida; e
- b) preços podem ser considerados exógenos, dado que são determinados pelos varejistas antes da realização da demanda.

Capps Jr., Church e Love (2003) argumentam que utilizam o procedimento presente em Freedman e Peters (1984) para lidar com a correlação serial presente em dados de demanda. A solução presente em Freedman e Peters, basicamente, consiste em estimar, em vez da equação (26), um modelo AR(1), de forma a considerar a possível correlação serial presente em dados de demanda. Capps Jr., Church e Love (2003) comentam que utilizaram uma modificação nesse procedimento, que seria estimar o sistema de demanda corrigindo para autocorrelação.

A existência de correlação serial na demanda é algo bastante razoável. Isso não só é observado empiricamente, como é simples de ser gerado pela teoria. Se imaginarmos que existe formação de hábito no consumo de certos produtos ou

mesmo que há aprendizado com o uso dos mesmos, então é automático observar que o estoque de consumo vai influenciar a quantidade consumida atualmente. Então, para certos produtos, essa dinâmica deve estar presente na estimação.

A presença de correlação serial nos erros de uma regressão gera sérios problemas para a implementação do *bootstrap*. Isso porque a idéia básica da técnica está em dizer que a distribuição dos resíduos que observamos na estimação original é um estimador consistente da distribuição verdadeira dos erros da regressão. Quando os erros são independentes podemos considerar que a ordem em que observamos cada um dos resíduos na série é fruto de outro sorteio, em que cada uma das realizações tem a mesma probabilidade de ocorrência para cada uma das posições na série. Essa é a lógica dos sorteios repetidos que nos permite gerar pseudobases de dados. Quando há dependência temporal na estrutura de erros, a ordem em que observamos os resíduos não é totalmente aleatória, isto é, a ordem em que os resíduos aparecem é também uma realização da variável aleatória. Agora, sortear novamente os resíduos ignorando essa correlação significa perder informação sobre a estrutura dos erros. Se as pseudobases de dados não corresponderem a realizações do verdadeiro processo gerador de dados, então todas as estimações provindas dessas novas bases de dados estarão viesadas e os resultados do *bootstrap* não terão significado.

Horowitz (2001) argumenta que, se for possível modelar a correlação serial a partir de um ARMA, então podemos implementar o *bootstrap* da mesma forma que para uma regressão comum. A idéia aqui é que, com um ARMA, é possível parametrizar a dependência existente nos dados e todo o componente aleatório restante seria ruído branco, o que nos permite continuar utilizando a técnica do sorteio repetido para a geração das pseudobases de dados. Essa é a idéia presente em Freedman e Peters. Ao estimar a demanda por energia como sendo um AR(1), os autores parametrizam a dependência temporal observada nos dados e assim o resíduo calculado é puramente ruído branco. A única ressalva que precisa ser feita refere-se à condição inicial. Lembre-se de que na geração das pseudobases de dados queremos calcular os novos valores da variável dependente, dado que sabemos os valores dos parâmetros (estimados), das variáveis explicativas (no nosso exemplo anterior, os preços) e os valores sorteados dos resíduos. No caso do AR, uma das variáveis explicativas é o valor defasado da variável dependente. Nesse caso, o procedimento mais comum utilizado na literatura é tomar o primeiro valor da variável dependente como sendo fixo e a partir daí calcular seus novos valores.

Alguns autores argumentam que, na impossibilidade de modelar o processo gerador de dados como sendo um ARMA, o *bootstrap* poderia ser realizado dividindo-se

a série de resíduos em blocos [por exemplo, bloco 1 seria  $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$ , bloco 2  $(\varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8)$  e assim por diante] e então sortear esses blocos como realizações independentes. A prática revela que esse procedimento não gera bons resultados, uma vez que ainda estamos “quebrando” a estrutura de correlação dos erros e, portanto, as pseudobases de dados geradas não corresponderão ao verdadeiro processo gerador de dados.

Entretanto, não podemos perder de vista que o procedimento indicado por Freedman e Peters é meramente estatístico, com o intuito de melhorar o poder explicativo da regressão. Modelar a demanda como um ARMA simplesmente para reduzir a correlação dos resíduos não tem significado econômico, e não é esse o objetivo quando procuramos estimar o sistema de demanda. As formas funcionais da demanda utilizadas neste estudo e aquelas apresentadas no Capítulo 1 deste livro são decorrentes de modelos econômicos, onde procuramos reproduzir a economia. Somente nesses casos um modelo de simulação passa a ser válido, pois estamos supondo que as firmas entendem que os consumidores se comportam dessa forma e assim tomam suas decisões. Incluir regressores ou variáveis defasadas que não tenham contrapartida no comportamento dos agentes não resolve o problema.

Enquanto a primeira hipótese diz respeito à viabilidade da implementação do *bootstrap*, a segunda hipótese impõe restrições muito fortes sobre a economia. Vamos supor um modelo um pouco mais geral, onde há correlação serial e os preços são determinados anteriormente à realização da demanda (isto é, preço  $p_t$  é determinado pelo varejista no instante  $t$  baseado na demanda observada em  $t-1$  e dados de custo observados em  $t$ ). As equações que definem essa economia seriam:

$$\ln q_{jt} = \alpha_j + \sum_{i=1}^N \beta_{ji} \ln p_{it} + \varepsilon_{jt} \quad (35)$$

$$\ln p_{jt} = \gamma_j + \rho_j \ln q_{jt-1} + \sum_{k=1}^N \delta_{kj} x_{kt} + \nu_{jt}$$

onde  $\varepsilon_{jt}$  apresenta correlação serial e  $\nu_{jt}$  seria independente. Simplificando-se ainda mais, vamos assumir por um instante que os preços são ditados apenas pela demanda anterior, isto é:

$$\ln p_{jt} = \gamma_j + \rho_j \ln q_{jt-1} + \nu_{jt} \quad (36)$$

Substituindo-se na equação de demanda, obtemos:

$$\ln q_{jt} = \alpha_j + \sum_{i=1}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} \quad (37)$$

$$\ln q_{jt} = \alpha_j + \beta_{jj} \gamma_j + \beta_{jj} \rho_j \ln q_{jt-1} + \sum_{i \neq j}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} + \beta_{jj} v_{jt}$$

$$\ln q_{jt} = \alpha_j + b_j \ln q_{jt-1} + \sum_{i \neq j}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} + \beta_{jj} v_{jt}$$

Isso significa que temos apenas duas relações:

$$a_j = \alpha_j + \beta_{jj} \gamma_j \quad (38)$$

$$b_j = \beta_{jj} \rho_j$$

e quatro incógnitas. Logo, não podemos identificar  $\alpha_j$  e  $\beta_{jj}$  estimando diretamente a partir da equação de demanda (37). Mas esse não é todo o problema. Observe que, se houver correlação serial na demanda, então o fato de assumir que os preços foram determinados anteriormente não resolve o problema da correlação entre os regressores (preços) e o erro, pois

$$\ln q_{jt-1} = \alpha_j + \sum_{i=1}^N \beta_{ji} \ln p_{it-1} + \varepsilon_{jt-1}$$

logo:

$$\begin{aligned} \ln q_{jt} &= a_j + b_j \left( \alpha_j + \sum_{i=1}^N \beta_{ji} \ln p_{it-1} + \varepsilon_{jt-1} \right) + \\ &+ \sum_{i \neq j}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} + \beta_{jj} v_{jt} \end{aligned} \quad (39)$$

e, por hipótese,  $\varepsilon_{jt-1}$  é correlacionado com  $\varepsilon_{jt}$ . Assim, as estimativas obtidas na regressão original são viesadas.

O que aconteceria se escrevêssemos a equação da demanda como sendo um AR, da mesma forma que Freedman e Peters? Nesse caso, teríamos:

$$\ln q_{jt} = \alpha_j + \phi_j \ln q_{jt-1} + \sum_{i=1}^N \beta_{ji} \ln p_{it} + \varepsilon_{jt} \quad (40)$$

e vamos dizer que agora  $\varepsilon_{jt}$  é independente ao longo do tempo. Substituindo-se a expressão do preço, temos:

$$\ln q_{jt} = \alpha_j + \phi_j \ln q_{jt-1} + \sum_{i=1}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} \quad (41)$$

$$\begin{aligned} \ln q_{jt} = & \alpha_j + \beta_{jj} \gamma_j + (\phi_j + \beta_{jj} \rho_j) \ln q_{jt-1} + \\ & + \sum_{i \neq j}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} + \beta_{jj} v_{jt} \end{aligned}$$

$$\ln q_{jt} = a_j + b_j \ln q_{jt-1} + \sum_{i \neq j}^N \beta_{ji} (\gamma_i + \rho_i \ln q_{it-1} + v_{it}) + \varepsilon_{jt} + \beta_{jj} v_{jt}$$

onde:

$$a_j = \alpha_j + \beta_{jj} \gamma_j \quad (42)$$

$$b_j = \phi_j + \beta_{jj} \rho_j$$

Agora não há, mais correlação entre os erros e os regressores. Logo, as estimativas da ultima expressão são não-viesadas e consistentes. No entanto, geramos outro problema:

a) Se a demanda corresponde realmente ao AR(1) descrito anteriormente, então essa forma funcional deve ser empregada na solução do problema da firma. Isso não é trivial, uma vez que agora, devido à característica dinâmica da demanda, o jogo entre as firmas deveria também ser dinâmico — afinal, o preço que as

firms escolhem hoje vai interferir na demanda de amanhã, impactando nos seus lucros. Como discutiremos na seção seguinte, o uso de jogos dinâmicos ainda é muito incipiente e suas previsões ainda não são robustas.

b) Se a demanda corresponde ao modelo estático original então, apesar de termos resolvido o problema da correlação entre os regressores e o erro, não conseguimos identificar os parâmetros de interesse (no caso,  $\alpha_j$  e  $\beta_{jj}$ ), estimando-se apenas a equação da demanda. Seria necessário estimar, inicialmente, a equação (35) para garantir a identificação. Mas Capps Jr., Church e Love (2003) não fazem isso.

Infelizmente, não há outras referências sobre o tema na literatura. A melhor alternativa seria estimar o sistema de demanda utilizando-se variáveis instrumentais, de forma a eliminar o problema de simultaneidade. No entanto, é possível que ainda assim os resíduos apresentem correlação serial. Tentar corrigir essa correlação modelando-se a demanda como um ARMA pode gerar mais problemas, tanto em termos teóricos como em termos de identificação. A outra opção, que seria realizar o *bootstrap* separando-se o resíduo em blocos, também tem seus percalços, uma vez que não estaremos reproduzindo o verdadeiro processo gerador de dados. Como apontam Hosken *et alii* (2002), o uso do *bootstrap* para a determinação de intervalos de confiança na simulação é uma idéia promissora, mas os problemas decorrentes da sua implementação ainda precisam ser resolvidos para que a técnica possa ser utilizada na prática antitruste.

#### **2.2.4 Possibilidade de entrantes**

A literatura de organização industrial devota grande parte de seu esforço em examinar quais as condições que levam à entrada de firmas em uma dada indústria. E isso está presente também nas preocupações do guia de análise de fusões.

Como observado anteriormente, alguns modelos dinâmicos têm incorporado os conceitos de custos afundados (*sunk cost*) e mínima escala viável e simulado as possíveis dinâmicas das indústrias. Modelos como Pakes e McGuire (1994), Ericson e Pakes (1995) e Gowrisankaran (1999) são bons exemplos dessa literatura. Infelizmente, como dissemos antes, são poucos os trabalhos empíricos que conseguem estimar os parâmetros desses modelos e assim utilizá-los na simulação da dinâmica da indústria. Além disso, ainda existe bastante controvérsia sobre esses estimadores, uma vez que as restrições que eles impõem ao jogo e à dinâmica não parecem ser muito realistas, bem como há dúvidas sobre as estratégias de identificação utilizadas.

Epstein e Rubinfeld (2001) apresentam uma idéia próxima daquela contida no cálculo do CMCR para verificar a viabilidade da entrada numa indústria. Eles

calculam qual seria o lucro de uma firma entrante na indústria após a fusão ao introduzir mais um produto e resolvem o sistema em (2) e (5) em função dos novos preços de equilíbrio. Como custo marginal da firma entrante eles consideram a média dos custos marginais apresentados pelas firmas participantes da indústria e, como estimativa das elasticidades, o arcabouço do Proportionally Calibrated Almont Ideal Demand System (PCAIDS) é bastante conveniente, uma vez que as elasticidades de cada produto são proporcionais à elasticidade da indústria (dado do problema) e às participações de mercado de cada firma, que é justamente um dos resultados da simulação.

Existem dois problemas com essa abordagem. Em primeiro lugar, com exceção do sistema de demanda PCAIDS, todos os outros sistemas são dependentes da estimativa econômétrica prévia da matriz de elasticidades. No caso da simulação de um entrante, seria preciso estimar quais as possíveis elasticidades próprias e cruzadas desse novo produto. Ainda há bastante controvérsia sobre como estimar a demanda de um novo produto, sendo que atualmente a literatura tem focado no uso de modelos hedônicos para a realização dessa tarefa. Imaginamos que o Capítulo 1, por tratar da estimação da demanda, possa apresentar com maiores detalhes o que tem sido feito nesse sentido.

O problema maior está em termos de inconsistência dinâmica. Ao introduzir um produto na indústria, as participações de mercado de cada uma das firmas são reduzidas e não é difícil obter resultados onde o lucro da firma pós-fusão seria inferior àquele obtido antes da fusão, caso não haja redução no custo marginal devido à fusão. Além disso, se levarmos em consideração valores de escala mínima viável, pode ser que não apenas a firma entrante não consiga alcançar essa escala, como as firmas participantes passem a operar abaixo da escala viável, forçando-as a sair do mercado. Esse resultado é demonstrado em Werden e Froeb (1998), utilizando-se indústrias competindo em um jogo de Bertrand.

Dessa forma, a utilização de modelos quantitativos para avaliar a possibilidade de entrada ainda é muito restrita. Imaginamos que, com o avanço na estimação de jogos dinâmicos seja possível estudar qual será a probabilidade de uma nova entrada na indústria após uma fusão. Entretanto, no atual estado da literatura, cálculos dessa natureza não são possíveis ou sofrem dos problemas já apontados, portanto, não têm aplicabilidade em termos de ferramenta de política.

### 2.3 EFICIÊNCIAS

A razão pela qual as autoridades antitruste se preocupam em saber qual será o novo preço de equilíbrio no mercado, após a fusão, é saber qual será o dano, em termos

de bem-estar, dos consumidores. Entretanto, não há consenso entre as diversas legislações antitruste sobre qual deveria ser a função de bem-estar a ser analisada. Como pode ser visto na Tabela 2.6 presente no Apêndice [extraída do ICN (2004)], das dez legislações analisadas, sete consideram apenas o bem-estar dos consumidores, enquanto as outras três levam em consideração o bem-estar total da economia.

Então, qual dos padrões é o mais correto? Essa questão não cabe neste estudo. A escolha da função de bem-estar é um problema de economia política. O legislador, ou, em última instância, a sociedade, pode considerar que uma redução de custos das firmas que não seja transferida para os consumidores não representa uma mudança benéfica para a sociedade; ou também pode ser o caso de o legislador considerar que qualquer aumento no preço final do bem represente uma perda de bem-estar para a sociedade, a qual não pode ser compensada por nenhuma redução de custos por parte dos produtores. Nesse caso, a função de bem-estar apenas dos consumidores é o mais natural para essa sociedade. Outro legislador pode considerar que as firmas são controladas, em última instância, pelos consumidores e que, portanto, aumento nos seus lucros representa aumento na renda dos consumidores. Se esse for o caso, é natural que ganhos das firmas entrem compensando as perdas dos consumidores, o que justificaria o uso de uma função de bem-estar total da economia (excedente do consumidor mais lucros das firmas).

Se, por um lado, a escolha dos padrões não cabe no escopo deste estudo, por outro, cabe aqui apresentar quais os tipos de incentivos que cada um dos padrões representa e quais os parâmetros da função custo (ou do problema dual, o da função de produção) são de interesse do regulador em cada uma das análises. Para isso vamos escrever uma economia composta por um agente representativo e um oligopólio. O agente possui utilidade  $u(x)$ , onde  $x$  representa o vetor de produtos consumidos. Nesse vetor, estamos incluindo não apenas os produtos que são objeto de análise na fusão, mas também todos os demais produtos que o agente consome. Essa função utilidade obedece às hipóteses tradicionais da teoria dos preços, ou seja, é crescente em cada um dos seus argumentos (consumir mais de qualquer um dos bens provê mais utilidade ao agente) e também é côncava (agentes preferem consumir “um pouco” de cada um dos produtos do que consumir apenas um dos bens). Considerando-se  $p$  como sendo o vetor de preços da economia e  $I$  a riqueza do agente, o problema que o agente resolve é:

$$\max u(x) \quad (43)$$

$$s.t. \quad px = I$$

A solução desse problema é caracterizada pelas condições de primeira-ordem

$$u_i(x) = \lambda p_i \quad (44)$$

onde  $\lambda$  é o preço-sombra da renda e  $u_i$  representa a utilidade marginal do bem  $i$  avaliada no ponto onde a cesta de consumo é  $x$ . Aqui, vale fazer a ressalva: caracterizar a solução do problema do consumidor utilizando-se as condições de primeira-ordem requer não apenas que a função utilidade seja diferenciável, mas também que a solução do problema seja interior, isto é, que as quantidades consumidas de cada um dos bens sejam positivas. Dado que o nosso objeto de estudo é o mercado de produtos diferenciados, essa hipótese pode ser entendida como sendo muito forte. Afinal, uma das características desse mercado seria que, dentre todos os produtos que lhe são próprios, os agentes escolhem apenas uma marca. Entretanto, se enxergarmos o agente representativo como sendo a média dos consumidores, então os valores de consumo que obtemos para cada um dos produtos seria a parcela de mercado que cada uma dessas marcas possui, se considerarmos o mercado como um todo. Nesse caso, assumir que as quantidades consumidas serão positivas não é uma hipótese tão forte assim.

Uma vez caracterizada a solução do problema do consumidor, podemos utilizar as condições de primeira-ordem e a restrição orçamentária para escrever as quantidades consumidas de cada bem como sendo função dos preços e da renda dos agentes, isto é:

$$x_i = f_i(p, I) \quad (45)$$

Substituindo-se em (43), temos:

$$u(x(p, I)) = u(f_1(p, I), f_2(p, I), \dots, f_N(p, I)) \quad (46)$$

A representação mostrada anteriormente da função utilidade também é chamada de função dispêndio, isto é, qual o valor em termos monetários da cesta de consumo que o agente obteve em seu problema de otimização.

Repare que a partir da equação (46) obtemos uma relação direta entre os preços que os agentes observam e a utilidade que eles obtêm a partir do consumo dos bens. Logo, no modelo anterior, qualquer variação por parte das firmas que não tenha impacto nos preços não terá impacto na utilidade dos agentes. Entretanto,

isso não quer dizer que o aumento do preço do bem  $i$  vai necessariamente reduzir a utilidade do agente. Derivando-se a expressão (46) com respeito ao preço do bem  $i$ , temos:

$$\frac{\partial}{\partial p_i} u(x(p, I)) = \sum_{j=1}^N u_j(x) \frac{\partial}{\partial p_i} f_j(p, I) \quad (47)$$

Apesar de a utilidade marginal ser sempre um valor positivo, o termo  $\frac{\partial}{\partial p_i} f_i(p, I)$  não necessariamente terá o mesmo valor para todos os bens considerados. É razoável imaginar que, para o bem considerado, temos  $\frac{\partial}{\partial p_i} f_i(p, I) < 0$ , isto é, a quantidade demandada do bem  $i$  é reduzida quando aumentamos o preço do mesmo. Por outro lado, as demais quantidades  $\frac{\partial}{\partial p_i} f_j(p, I)$  podem muito bem ser positivas, principalmente se os bens  $i$  e  $j$  forem substitutos. Nesse caso, o efeito total na utilidade de um aumento no preço do bem  $i$  é incerto e passa a ser objeto de estudo empírico.

Do lado das firmas, vamos considerar o mesmo modelo já tratado na Subseção 2.2.1: as firmas competem num jogo estático, escolhendo preços como sua variável estratégica. A única diferença que acrescentamos aqui é que o custo seja composto de uma parte variável  $c$  e de uma parte fixa  $F$ . Dessa forma, o problema de cada firma passa a ser

$$\max_{p_i} (p_i - c_i) q_i - F_i \quad (48)$$

Assim como mostramos anteriormente, a solução do problema citado é dada pela condição de primeira-ordem

$$p_i + (p_i - c_i) \frac{\partial q_i}{\partial p_i} = 0 \quad (49)$$

Cada uma das firmas vai escolher preços de forma a obedecer à condição citada, assumindo que as demais firmas no mercado também farão o mesmo. Isso significa que *apenas o custo marginal será relevante na determinação do preço observado*. A presença de custos fixos terá outra importância: a determinação do número de

firms na indústria. Isso é fácil de perceber se retornarmos à expressão do lucro da firma,  $\pi_i = (p_i - c_i)q_i - F_i$ . Para que a firma opere, é preciso que esse lucro não seja negativo. Então, se escrevermos a demanda pelo bem  $i$  como sendo a diferença entre a demanda total pelos bens da indústria  $Q$  e a demanda pelos bens de todos os outros competidores  $q_{-i}$ , o lucro da firma  $i$  passa a ser:

$$\pi_i = (p_i - c_i)(Q - q_{-i}) - F_i \quad (50)$$

Logo, imagine que existam  $N$  firmas nessa indústria.  $N$  deve ser tal que se mais uma firma, digamos  $N + 1$ , decidir entrar no mercado, então os lucros dessa entrante serão negativos, isto é:

$$\pi_{N+1} = (p_{N+1} - c_{N+1})(Q - q_{-N+1}) - F_{N+1} < 0 \quad (51)$$

onde  $q_{-N+1} = \sum_{i=1}^N q_i$ .

Da discussão anterior podemos ver que o custo fixo não interfere na determinação do preço do bem, apenas no nível dos lucros de cada firma; por outro lado, os custos marginais afetam não apenas o nível dos lucros, mas também o preço de equilíbrio. Essa distinção é extremamente importante, pois é aqui que separamos os tipos de economias em custo que serão de interesse do regulador. Se a função de bem-estar considerada pela autoridade antitruste, levar em consideração apenas o bem-estar do consumidor, então cabe ao analista investigar apenas as possíveis economias no custo marginal que possam ser decorrentes da fusão. Se a função de bem-estar contemplar tanto o excedente dos consumidores como o lucro das firmas, então economias em custo fixo também devem ser investigadas no momento de determinar as eficiências decorrentes da fusão.

É importante lembrar que esses resultados são dependentes desse modelo utilizado em simulação de fusões. As simplificações aqui presentes, necessárias para a tratabilidade do problema, podem estar escondendo efeitos que deveriam ser considerados. Por exemplo, da forma que modelamos o problema do consumidor, não incluímos a possibilidade de a fusão introduzir novos produtos. Nesse caso, um produto novo poderia ser visto como um bem que, antes da fusão, tinha seu consumo restrito a 0, e após a fusão pode ser consumido fartamente. Assumindo-se as mesmas hipóteses sobre a função utilidade, a introdução desse produto

potencialmente aumenta o bem-estar do nosso consumidor.<sup>11</sup> Se, por outro lado, a introdução desse novo bem estiver condicionada a algum tipo de custo fixo (por exemplo, a construção de uma nova planta ou algum investimento em pesquisa), então a variação nos custos fixos passa a ter importância no bem-estar do consumidor. O problema aqui seria como medir o valor que o consumidor dá a esse novo bem antes de ser lançado (pois, como a fusão será analisada antes do lançamento desse novo bem, não há um preço que possamos medir e usar como referência). Todavia, algumas jurisdições que tomam como padrão apenas o bem-estar do consumidor levam em consideração essas economias de custo fixo que permitem a introdução de um novo produto, ou investimentos em tecnologia/qualidade que aprimorem os produtos já existentes, mesmo sem poder quantificar o benefício que essas mudanças representam para os consumidores.

O uso de modelos estáticos para a competição também colabora para a eliminação da dependência entre preços e custos fixos. Se utilizássemos um modelo dinâmico de indústria, como é o caso de Ericson e Pakes (1995), poderíamos gerar essa dependência. Em linhas gerais, em um modelo dinâmico, as firmas que já participam da indústria podem achar lucrativo definir seus preços abaixo dos valores ditados pela condição (49) se souberem que, ao fazer isso, os custos fixos das possíveis entrantes impedirão as mesmas de entrar no mercado. Se esse for o caso, mesmo se a função de bem-estar considerada fosse apenas a do consumidor, seria importante verificar as possíveis economias de custo fixo geradas pela fusão.

Modelos mais gerais que contemplem esses casos mencionados passam a ser muito complicados, tanto em termos de estimação, como em termos de predição. Isso é um grave problema em termos de ferramenta de política, uma vez que decisões precisam ser tomadas com rapidez e entendidas com facilidade, tanto pelos analistas como pelas partes interessadas no problema. Dessa forma, o modelo apresentado, apesar de sua simplicidade, é bastante utilizado como padrão nas análises antitruste ao redor do mundo.

Uma vez entendido em quais tipos de economias de custo estamos interessados, o estágio seguinte passa a ser identificar os fatores que interferem na determinação de custos marginais e custos fixos, para que possamos investigar as possíveis economias nesses dois custos. McFetridge (1999) aponta que ganhos de eficiência podem ser entendidos como reduções no custo médio de produção, controlado pela qualidade. Observe que, ao considerar o custo médio, o autor se aproxima do

---

11. Dizemos potencialmente porque, se ao mesmo tempo que esse bem for introduzido os preços dos demais produtos substitutos forem aumentados, então pode ser o caso de após a fusão a utilidade ser menor que anteriormente.

padrão de bem-estar total da economia pois, como já apontamos, um regulador interessado apenas no bem-estar dos consumidores poderia se restringir a analisar fatores que interferissem somente no custo marginal. De qualquer forma, seguindo a definição de McFetridge, devemos identificar quais os canais para se alterar o custo médio de um bem. Basicamente temos:

*a) Economias de escala ou escopo:* relaciona-se à curvatura da função de produção.

*b) Ganhos de produtividade:* relaciona-se ao nível da função de produção.

É importante ter em mente esses dois pontos. Grande parte da literatura sobre cálculo de eficiências perde-se em conceitos e nomenclaturas que são meras variações desses dois conceitos econômicos. Retornos de escala ou escopo e produtividade são elementos presentes na especificação da função custo/função de produção e existe uma literatura especializada na sua estimativa. Conceitos como economias de racionalização, economias de utilização de capacidade, sinergias, não estão presentes em modelos estruturais e, portanto, não são factíveis de serem testados ou medidos. A abordagem que seguimos é traduzir esses conceitos em termos dos dois pontos citados, e assim procurar fazer inferências sobre as possíveis eficiências resultantes da fusão.

McFetridge (1999) e ICN (2004) apresentam, com bastante detalhe, os principais fatores considerados pelas autoridades antitruste ao analisar a presença de eficiências em uma fusão. A seguir, enumeramos esses pontos.

*a) Economias em custo fixo.*

*b) Eficiências produtivas:* envolvem economias de escala, escopo e sinergias.

*c) Economias em propaganda e marketing:* na medida em que os contratos não forem feitos dependentes da quantidade produzida, pode ser considerada uma redução de custo fixo.

*d) Eficiências dinâmicas:* desenvolvimento de novos produtos ou introdução de novas tecnologias produtivas. Esse caso se enquadraria justamente em ganhos de produtividade. Como não sabemos como investimento em tecnologia altera a forma da função de produção, não há modelo que possa simular esse tipo de eficiência ou criar um intervalo de confiança a respeito dos possíveis ganhos.

*e) Eficiências transacionais:* envolvem todo tipo de redução nos custos de contratos com terceiros, como, por exemplo, contratos de distribuição. Esse tipo de economia pode reduzir tanto o custo fixo (se se referir a custos intrinsecamente

ligados à decisão de a firma funcionar ou não) ou o custo marginal (se se referir ao custo unitário de produção). Entretanto, seguindo a discussão levantada na seção anterior a respeito de economias de escala, não é claro quanto dessa economia está relacionada ao aumento do poder de barganha da firma fusionada ou quanto é decorrente de não-linearidades específicas dos preços desses contratos. De qualquer forma, eficiências transacionais devem ser analisadas sob a ótica de economias de escala/escopo. Como se trata da repartição do excedente entre os diversos níveis da cadeia produtiva, deixamos para o Subcapítulo 5A deste livro a análise desse tipo de eficiências.

*f)* Efeitos de rede pelo lado da demanda: acontecem quando o valor do bem percebido pelo consumidor aumenta com o número de usuários do mesmo. Envolve as propriedades da função de demanda, as quais não são objeto deste estudo.

*g)* Redução de custos administrativos: é toda redução em custos decorrente da eliminação de níveis hierárquicos e ou substituição de pessoal.

*h)* Redução de custos de capital: envolve todo tipo de não-linearidade no preço do serviço do capital devido a imperfeições no mercado de crédito. Pode-se apresentar na forma de taxas de juros mais baixas para a firma fusionada, acesso a novos mercados de crédito, redução do colateral. Esse seria um dos casos mais próximos onde a não-linearidade no preço do insumo não seria decorrente do aumento de poder de barganha da firma fusionada, mas sim de propriedades do mercado desse insumo e, consequentemente, seria mais condizente com a definição de economias de escala.

Como pode ser observado, com exceção do item *(b)*, os demais itens considerados eficiências podem ser divididos em redução de custos fixos — itens *(a)*, *(c)*, *(e)* e *(g)* — ou redução de preços de insumos — itens *(e)* e *(h)* — a qual relaciona-se à redução nos custos marginais. Para podermos discutir como esses conceitos se relacionam a economias de escala ou escopo, vamos primeiro relembrar esses conceitos. Na subseção seguinte, apresentamos como a literatura empírica tem tratado o problema de estimação dessas relações.

### 2.3.1 Economias de escala

Panzar (1989) apresenta duas definições para economias (ou retornos) de escala, as quais nem sempre se equivalem. Considere-se uma tecnologia  $T$  e um vetor de quantidades  $(x, y)$  tal que  $y$  é um vetor de produtos (com  $J$  componentes) e  $x$  é um vetor de insumos (com  $I$  componentes) e  $y = f(x)$ .

Definição 1 (economias de escala tecnológicas). Defina a medida de economia de escala como sendo

$$\tilde{S}(x, y) = \frac{\sum_{j=1}^J \sum_{i=1}^I x_i f_{ji}(x)}{\sum_{j=1}^J y_j} \quad (52)$$

onde  $f_i(x)$  é a produtividade marginal do insumo  $x_i$  na produção do bem  $j$ . Dizemos que o retorno de escala é crescente, constante ou decrescente se  $\tilde{S}$  for maior, igual ou menor que 1, respectivamente.

Definição 2 (economias de escala em custo). Defina a medida de economia de escala como sendo

$$S(y, w) = \frac{C(y, w)}{\sum_{i=1}^J y_i C_i(y, w)} \quad (53)$$

Dizemos que o retorno de escala é crescente, constante ou decrescente se  $S$  for maior, igual ou menor que 1, respectivamente.

A primeira definição é certamente a mais intuitiva: se aumentarmos em  $k$  vezes a quantidade demandada de todos os insumos, a tecnologia de produção apresentará retornos crescentes de escala se  $y$ , a quantidade produzida, aumentar em mais de  $k$  vezes. A segunda definição tem uma interpretação bastante simples no caso de firmas que produzem um único bem. Reescrivendo-se a expressão (53), temos  $S = \frac{C}{y} \cdot \frac{1}{C'}$ . Derivando-se com respeito a  $y$ , temos:

$$\begin{aligned} \frac{d}{dy} \left( \frac{C}{y} \right) &= \frac{C'}{y} \left( 1 - \frac{C}{y} \cdot \frac{1}{C'} \right) \\ &= \frac{C'}{y} (1 - S) \end{aligned} \quad (54)$$

Como o custo marginal é positivo, o valor da derivada citada vai depender de  $S$  ser maior ou menor que 1, ou seja, se a função custo apresentar retornos crescentes de escala, então o custo médio de produção será decrescente.

Seriam equivalentes as duas definições? A princípio não. Enquanto na definição 1 a existência ou não de economias de escala está relacionada exclusivamente às propriedades da função de produção  $f(\cdot)$ , na definição 2 retornos de escala seriam o resultado não apenas da tecnologia utilizada, mas também do processo de minimização de custos da firma. Com isso queremos dizer que uma firma que deseja aumentar em  $k$  vezes a sua produção não necessariamente vai aumentar em  $k$  vezes a sua demanda por cada insumo. Pode ser o caso em que a solução do problema de minimização de custos implique a escolha de um outro vetor de insumos  $x'$  e que esse vetor resulte em um custo final inferior a  $kC$ , onde  $C$  representa o custo do vetor inicial de produção.

A proposição 2 de Panzar mostra que, se considerarmos as firmas como sendo tomadoras de preço — preços lineares nos insumos e a função de produção pelo menos fracamente monotônica e continuamente diferenciável em  $x$  — então as duas definições de retornos de escala serão equivalentes. Podemos ser indiferentes às duas definições ou essas hipóteses são muito fortes?

Quando se analisa a fusão entre duas ou mais firmas, as duas primeiras hipóteses podem não ser satisfeitas. A existência de preços não-lineares entre firmas em diferentes pontos da cadeia de produção é algo comum e muitas vezes reportado pelas próprias firmas fusionadas e, nesse caso, mesmo simples esquemas de tarifas, como por exemplo desconto por quantidade, podem ruir os resultados previstos pela teoria da dualidade. O seguinte exemplo ilustra bem o fato.

Suponha-se que esses preços obedeçam a um esquema não-linear, do tipo  $w_i = \bar{w}_i - \beta x_i$ , isto é, a firma obtém descontos ao comprar em grandes quantidades. Vamos assumir também uma função de produção linear bem simples, do tipo

$$y = f(x_1, x_2) = \alpha x_1 + (1 - \alpha)x_2 \quad (55)$$

onde:

$$\alpha \in [0, 1]$$

Claramente a função de produção (55) apresenta retornos constantes de escala, segundo a definição 1. O problema que a firma resolve é:

$$\underset{x_1, x_2}{\text{Min}} w_1 x_1 + w_2 x_2 \quad (56)$$

$$s.t. y \leq f(x_1, x_2)$$

onde  $w_1 = \bar{w}_1 - \beta x_1$  e  $w_2 = \bar{w}_2$ . As condições de primeira-ordem do problema ficam

$$[x_1] : -(\bar{w}_1 - \beta x_1) + \beta x_1 + \lambda \alpha = 0 \quad (57)$$

$$[x_2] : -\bar{w}_2 + \lambda(1 - \alpha) = 0$$

Simplificando-se as duas expressões e substituindo-se em (56), obtemos a função custo

$$C(y, w) = \lambda y + \beta x_1 (w_1, w_2, \alpha, \beta) = \lambda y + F \quad (58)$$

Isto é, a função custo apresenta um termo de custo fixo, positivo. Usando-se a definição 2 é imediato observar que a firma apresenta retornos crescentes de escala.

Preços não-lineares podem surgir como decorrência de o mercado de insumos ser monopolístico. Nesse caso, uma fusão poderia resultar em economias de escala específicas, pois somente uma firma com maior escala seria capaz de gozar de descontos nos preços dos insumos. Também podem ser decorrentes de características do insumo propriamente dito. Muitas vezes, as firmas fusionadas alegam, por exemplo, economias com firmas de consultoria, propaganda, auditoria, os quais têm seus preços não-atrelados à quantidade produzida.

Como foi observado, se o mercado de insumos não for perfeitamente competitivo, os resultados de dualidade também passam a não ser válidos. Esse seria o caso onde, à medida que os produtores aumentam a sua demanda por insumos, eles passam a ter vantagens na negociação dos preços dos mesmos. Esse caso é muitas vezes alegado pelas firmas fusionadas como argumento pró-eficiência, erroneamente. Efeitos resultantes de variação no poder de barganha são meramente distributivos (teorema de Coase) e, portanto, não alteram o bem-estar social

como um todo. A discussão mais aprofundada sobre como acontece essa repartição do excedente e como ela deve interferir no bem-estar total da economia será melhor discutida no Subcapítulo 5A deste livro.

Nos casos vistos existe uma literatura teórica relativamente vasta sobre como aconteceriam essas relações entre os diversos pontos da cadeia produtiva — uma referência seria Katz (1989). Infelizmente, muito pouco existe de empírico.<sup>12</sup> Uma possibilidade a ser explorada seria na linha de Villas-Boas (2004) ou Hastings (2004); no entanto, seria preciso não apenas obter os preços que as firmas observam para tais insumos, como também detalhes do contrato, para verificar a presença de não-linearidades nos preços.

De qualquer forma, como será visto nas seções sobre estimação da função custo e da função de produção, a literatura tem apresentado mais avanço na estimativa da função de produção. Por um lado, os resultados de economia de escala obtidos a partir da função de produção são legitimamente ganhos de eficiência, dado que estão relacionados à tecnologia. Por outro, deixam a cargo do pesquisador inferir a função custo e a partir daí simular as reduções em custo proporcionadas pelas variações nos preços dos insumos. Não tivemos ciência de artigos que usem essa abordagem na estimativa da função custo pós-fusão, mas imaginamos ser possível utilizar uma abordagem como esta para auxiliar no cálculo das eficiências (desde que haja dados suficientes para a estimativa da função de produção). Entretanto, é bom frisar que resta o problema de saber se a função de produção das firmas fusionadas será igual àquela de antes da fusão. A literatura ainda não apresentou, até o momento, bons resultados empíricos ou teóricos sobre o assunto, que possam nos auxiliar.

### **2.3.2 Economias de escopo**

A definição de economias de escopo, ao contrário da definição de retornos de escala, é exclusivamente baseada nas propriedades da função custo. Novamente, utilizamos a definição de Panzar (1989) para guiar o nosso estudo.

Definição 3 (economias de escopo). Seja  $\widehat{T}$  um subconjunto do espaço de produtos ( $\widehat{T} \subset T$ ), no ponto onde o vetor de quantidades é  $y$ . O grau de economias de escopo é definido como

$$SC_T(y) = \frac{C(y_T) + C(y_{\widehat{T}}) - C(y)}{C(y)} \quad (59)$$

12. Existe alguma coisa sobre custos de transação, que é um dos pontos levantados como possíveis eficiências. Para isso, teríamos Mortimer (2002) e Sieg (2000). Além desses, existem os trabalhos de Bresnahan e Reiss (1985) e Corts (2001).

onde  $y_T$  é a quantidade produzida do subconjunto de produtos  $T$  e  $y_{\hat{T}}$  é a quantidade produzida do subconjunto de produtos  $\hat{T}$ . A função custo apresenta economias de escopo se  $SC$  for maior que 1.

Intuitivamente, a presença de economias de escopo significa que o custo da produção conjunta de dois bens é menor que o custo total de produzi-los em plantas e ou firmas separadas. Como é apresentado na seção de estimação, isso é testado a partir da propriedade de subaditividade da função custo.

Apesar de a definição de economias de escopo não ter uma contrapartida exclusivamente tecnológica, Panzar aponta que as razões da presença de economias de escopo são de natureza tecnológica. Em suma, ele apresenta duas possíveis razões para isso:

- a) A tecnologia de produção do conjunto  $\hat{T}$  utiliza insumos “públicos”, isto é, uma vez adquirido o insumo  $x_i$  para a produção do bem  $y_j$ , este mesmo insumo pode produzir o bem  $y_l$  sem acréscimo de custo.
- b) A produção de um bem gera externalidades positivas na produção de outros. Isso poderia ser o caso de tecnologias que apresentem *learning by doing*, apesar de que seria difícil identificar o que seriam economias de escala e o que seriam economias de escopo.

Como será visto na subseção seguinte, pode ser encontrado muito pouco na literatura sobre como testar subaditividade da função custo. Grande parte do problema está na falta de disponibilidade de dados mas, mesmo na presença destes, fatos como a especificação da forma funcional e problemas de identificação gerados por preços não-lineares também complicam a estimativa. Uma possível alternativa seria Beresteanu (2004), que parte para testes não-paramétricos da subaditividade da função custo. Entretanto, o uso de métodos não-paramétricos, como já foi apontado na seção correspondente ao cálculo do CMCR é inviável, devido à enorme quantidade de dados necessários para a sua implementação. Abordamos aqui apenas o procedimento proposto por Evans e Heckman (1983).

### **2.3.3 Produtividade**

Ganhos de produtividade são outra forma de uma fusão poder alterar a curva de custos das firmas. Basicamente, ganho de produtividade seria uma alteração na função de produção que permitiria aumentar a quantidade produzida utilizando-se a mesma quantidade de insumos. Uma boa resenha a respeito de como produtividade tem sido medida e sobre seus impactos em organização industrial pode ser encontrado em Bartelsman e Doms (2000).

A literatura sobre organização industrial usa o mesmo procedimento empregado pela literatura sobre crescimento no cálculo da produtividade total dos fatores [Total Factor Productivity (TFP)]: assume uma forma paramétrica para a tecnologia de produção, estima os seus coeficientes e utiliza o resíduo da regressão como sendo o indicador para produtividade. Essa é a abordagem utilizada, por exemplo, por Olley e Pakes (1996) e Levinsohn e Petrin (2003). Aparentemente, essa não seria uma tarefa muito complexa, principalmente se assumirmos uma especificação paramétrica para a função de produção. A realidade é um pouco diferente: devido a diferenças no conjunto de informação das firmas e do econometrista, é preciso tomar cuidado na escolha das variáveis e no método de estimativa para que as estimativas não sejam viesadas. Tratamos disso na seção correspondente à estimativa da função de produção.

É bastante razoável imaginar que alguns dos ganhos explorados pelas firmas fusionadas estejam em implementar mudanças na tecnologia de produção, que podem ser realizadas apenas se a firma possuir uma escala de produção maior do que detém antes da fusão. É, por exemplo, o caso de investimentos em pesquisa e desenvolvimento (P&D) que resultem em melhores tecnologias de produção.

Da forma como tem sido medida, o termo de produtividade pode acrescentar outras “eficiências” que não necessariamente seriam devidas à mudança tecnológica. Suponha-se que a fusão sirva para que a produção seja transferida de uma planta menos produtiva para outra mais produtiva. Essa eficiência alocativa é medida como ganho de produtividade, apesar de não ter envolvido a mudança na fronteira tecnológica. Outro caso seria a eliminação de alguns tipos de restrição na produção de alguma firma, por exemplo, restrições de crédito.

De qualquer forma, todas essas possibilidades constituem eficiências legítimas da operação. O grande problema, do ponto de vista do regulador que precisa estimar *ex ante* quais seriam esses ganhos, está em imaginar como a fronteira tecnológica pode ser alterada com uma fusão. Isso requereria modelos dinâmicos que ainda não possuem uma contrapartida empírica que possa validá-los. Verificação de restrições de crédito ou estimativa das produtividades atuais da indústria é um exercício factível e nós cremos que esta deva ser a abordagem de um regulador interessado em estimar os possíveis ganhos de eficiência de uma fusão.

Uma vez explicado por que é importante para o regulador estimar a função custo ou a função de produção vamos apresentar o que tem sido feito na literatura para estimar essas funções. Feito isso, discutimos quais as variáveis necessárias para a estimativa dos modelos e quais desses modelos seriam os mais indicados para o problema do regulador.

### 2.3.4 Estimação da função custo

Existem duas formas de caracterizar a tecnologia das firmas em análise: podemos estimar diretamente a função de produção ou utilizar o problema dual, que seria estimar a função custo. A relação entre essas duas funções é bastante simples, se observarmos o problema da firma. Independentemente de qual seja a sua função objetivo (pode ser maximizar lucro ou maximizar receita ou, às vezes, maximizar prestígio, como se supõe ser a função objetivo das universidades), sabemos que as firmas procuram minimizar custos. Então o problema da firma fica:

$$\min_x px \quad (60)$$

$$s.t. y \leq f(x)$$

onde  $x$  representa o vetor de insumos,  $p$  é o vetor de preços para esses insumos e  $f(x)$  é a função de produção. A solução do problema anterior é caracterizada pelas condições de primeira-ordem

$$-p_i + f_i(x) = 0 \quad (61)$$

isto é, a produtividade marginal de cada insumo deve ser igual ao preço do mesmo. Podemos usar essas condições para escrever a demanda por insumos como função dos seus preços. Dessa forma, o custo de produção de  $y$  unidades do bem final passa a ser:

$$C = px(p) = C(p) \quad (62)$$

que é a tradicional função custo. Como a demanda por cada insumo é derivada utilizando a função de produção, a função custo anterior possui uma relação direta com a função de produção. Por isso pudemos caracterizar, na subseção anterior, o conceito de economias de escala segundo as duas funções.

Considerando-se que o modelo que apresentamos na Subseção 2.2.1 usa o custo marginal da firma como dado e, além disso, os fatores relacionados à eficiência dizem respeito a algum tipo de economia em custos (marginal ou fixo, dependendo

da função de bem-estar), o caminho natural para o analista seria estimar a função custo de cada firma e, de posse dos parâmetros estimados, fazer inferências sobre os custos marginais e fixos. Infelizmente, a literatura não seguiu esse rumo.

Até meados dos anos 1980, a literatura sobre organização industrial tinha como foco a estimação de funções custo. Para isso, utilizava diferentes formas funcionais, como são mostradas a seguir nos trabalhos de Spady e Friedlaender (1978), Evans e Heckman (1983), entre outros. Essa prática entrou em desuso, principalmente por problemas de mensuração nos preços.

Mensuração de preços dos insumos é um problema crítico. Imagine o problema de estimar a função custo da firma  $i$ . Para isso, dispomos de uma série temporal dos custos da mesma, bem como dos preços dos insumos utilizados por ela. Vamos utilizar uma especificação linear para facilitar o nosso trabalho. A função a ser estimada será:

$$\ln C_{it} = \alpha_i + \gamma_i \ln p_{it} + \sum_{j=1}^J \beta_{ij} \ln w_{jt} + \varepsilon_{it} \quad (63)$$

Vamos por um instante abstrair quaisquer problemas de simultaneidade que possam estar correlacionando o termo não-observável com os preços dos fatores, isto é, suponha-se que os preços dos fatores e do bem  $i$  são realmente independentes do termo  $\varepsilon_{it}$ . Nesse caso, a identificação dos parâmetros  $\beta_{ij}$  vai depender exclusivamente da variação observada nos preços. É justamente essa variação que não temos quando utilizamos séries temporais dos preços dos insumos para cada firma ou para firmas dentro da mesma indústria. Essas séries apresentam longos períodos sem flutuação alguma, seja por causa de alguma rigidez nos preços (talvez o desenho dos contratos entre os diferentes níveis da cadeia produtiva mantenha os preços dos insumos fixos por algum período), seja porque os preços que observamos não conseguem refletir o custo de utilização dos insumos que, em última instância, são as variáveis que desejamos empregar na regressão.

Em qualquer uma das duas hipóteses (falta de variabilidade devido à característica da série ou devido a erro de medida), a estimação da função citada se tornou inviável quando a análise passou a ser com dados desagregados. É por essa razão que os trabalhos empíricos mais recentes em teoria da produção são para estimar funções de produção e não funções custo.

De qualquer maneira, apresentamos a seguir os trabalhos mais relevantes sobre estimação de função custo, no caso Spady e Friedlaender (1978), Evans e

Heckman (1983) e Berry, Kortum e Pakes (1996). Os detalhes sobre a estimação de funções de produção (bem como seus possíveis problemas), são tratados na seção seguinte.

#### 2.3.4.1 Spady e Friedlaender

Esses autores apresentam um modelo econométrico que leva em consideração diferenças em atributos ou qualidade dos produtos. Eles argumentam que funções de custo hedônicas seriam um caminho natural na especificação de custos em indústrias onde os produtos apresentem variação em qualidade, uma vez que não apenas corrigiriam pela diferenciação dos produtos como também evitariam o problema de dimensionalidade de parâmetros quando se tenta discretizar o espaço de produtos por esses atributos.

Eles definem a função custo como sendo

$$C(\psi(y, q), w) \quad (64)$$

onde  $\psi(y, q)$  representa um vetor de funções que mede produção efetiva (isto é, controlada pelos atributos) e  $w$  representa o vetor preços dos fatores. Essa função, como os autores os observam, apresenta separabilidade na qualidade. Com isso, eles impõem que as variações em custo, devido a mudanças na qualidade dos bens produzidos, são independentes dos preços dos insumos. Essa hipótese não é pouco restritiva. Tome o caso analisado por eles, cargas rodoviárias. Separabilidade nos atributos implica que aumentos no preço do combustível não alteram o tipo de caminhão que seria utilizado para transportar o mesmo peso, pela mesma distância (os autores consideram como sendo um dos atributos da regressão hedônica a relação peso/distância). Talvez na estimação da função custo de curto prazo essa hipótese possa não ser muito restritiva, se imaginarmos que as transportadoras não serão capazes de adequar a frota (que basicamente determina a relação peso/distância) ao novo preço de combustível. Entretanto, no longo prazo, é muito provável que ajustes ocorram na frota (portanto, no atributo  $q$ ) de forma a se adequar ao novo preço do combustível.

Além da hipótese de separabilidade em qualidade, os autores assumem que a função  $\psi(.)$  seja homogênea de primeiro grau em quantidades físicas,  $y$ , isto é,

$$\psi^i(y_i, q^i) = y_i \phi(q_1^i, \dots, q_r^i) \quad (65)$$

Spady e Friedlaender utilizam uma especificação translog para estimar a função custo, da forma:

$$\begin{aligned} \ln C(\psi, w) = & \alpha_0 + \sum_{i=1}^N \alpha_i (\ln \psi_i - \ln \bar{\psi}_i) + \sum_{s=1}^M \beta_s (\ln w_s - \ln \bar{w}_s) + \\ & + \frac{1}{2} \left[ \sum_{i=1}^N \sum_{j=1}^N A_{ij} (\ln \psi_i - \ln \bar{\psi}_i) (\ln \psi_j - \ln \bar{\psi}_j) + \right. \\ & \left. + \sum_{s=1}^M \sum_{t=1}^M B_{st} (\ln w_s - \ln \bar{w}_s) (\ln w_t - \ln \bar{w}_t) \right] + \\ & + \sum_{i=1}^N \sum_{s=1}^M C_{is} (\ln \psi_i - \ln \bar{\psi}_i) (\ln w_s - \ln \bar{w}_s) \end{aligned} \quad (66)$$

Os autores consideram que os preços dos insumos são exógenos, o que lhes permite usar o lema de Shepard para determinar o sistema de equações que caracteriza a demanda de fatores. Essas equações podem ser estimadas junto com a equação de custo de forma a aumentar a eficiência dos estimadores:

$$\frac{w_s x_s}{C} = \beta_s + \sum_{s=1}^M B_{st} (\ln w_t - \ln \bar{w}_t) + \sum_{i=1}^N C_{is} (\ln \psi_i - \ln \bar{\psi}_i) \quad (67)$$

$$s = 1, \dots, m-1$$

Devido à homogeneidade da função hedônica podemos escrever

$$\ln \psi^i = \ln y_i + \ln \phi^i (q_1^i, \dots, q_r^i) \quad (68)$$

Usando-se uma aproximação translog à equação anterior, temos:

$$\begin{aligned} \ln \psi^i = & \ln y_i + \sum_{b=1}^H \alpha_b^i (\ln q_b^i - \ln \bar{q}_b^i) + \\ & + \frac{1}{2} \sum_{b=1}^H \sum_{l=1}^H b_{bl}^i (\ln q_b^i - \ln \bar{q}_b^i) (\ln q_l^i - \ln \bar{q}_l^i) \end{aligned} \quad (69)$$

Por fim, Spady e Friedlaender impõem homogeneidade nos preços dos fatores na função custo, bem como simetria

$$\sum_{s=1}^M \beta_s = 1; \quad (70)$$

$$\sum_{s=1}^M B_{st} = 0, \quad t = 1, \dots, m;$$

$$\sum_{s=1}^M C_{is} = 0, \quad i = 1, \dots, n;$$

$$B_{st} = B_{ts}; \quad A_{ij} = A_{ji}.$$

O sistema de equações dado por (66) e (67) é estimado conjuntamente, por mínimos quadrados generalizados. Também é simples de se obter a expressão para o custo marginal de cada produto. Ele é dado por:

$$MC_i = \frac{C}{y_i} \frac{\partial \ln C}{\partial \ln y_i} = \frac{C}{y_i} \left\{ \alpha_i + \sum_{i=1}^N A_{im} (\ln y_i - \ln \bar{y}_i) + \sum_{j=1}^M C_{ij} (w_j - \bar{w}_j) \right\} \quad (71)$$

logo, a especificação dada permite testar a presença ou não de complementaridades em custo:<sup>13</sup> se  $A_{ij} < 0$  então podemos dizer que há complementaridades em custo. A presença de economias de escala para essa função custo pode ser testada calculando-se

$$S_m = \frac{C(Y)}{\sum_{i=1}^N Y_i \frac{\partial C}{\partial Y_i}} \quad (72)$$

---

13. Complementaridade em custo pode ser definida como  $\frac{\partial^2 C}{\partial y_i \partial y_j} \leq 0$ , isto é, um aumento na quantidade produzida do bem  $i$  não aumenta o custo marginal de produzir o bem  $j$ .

onde o teste seria contra a hipótese de que  $S_m$  seria menor ou igual a 1. A distribuição de  $S_m$  poderia ser calculada utilizando-se o método delta, uma vez que a estimativa da função custo já proveria a variância dos parâmetros presentes em  $\frac{\partial C}{\partial Y_i}$ .

O artigo de Spady e Friedlaender é interessante pois é um dos primeiros a utilizar funções translog como aproximação de segunda ordem da função custo. Além disso, eles também são um dos primeiros a estimar uma função custo hedônica, que leve em consideração os atributos dos produtos. Entretanto, alguns pontos devem ser observados no tocante ao uso dessa abordagem em estimação de funções custo:

- Como pode ser visto, a função translog não é definida no 0. Então, dependendo do nível de desagregação da base de dados, o seu uso pode ser inviável. Tome-se, por exemplo, um painel de dados de plantas. É possível que algumas plantas sejam multiproduto e outras não, ou mesmo que algumas plantas que produziam mais de um produto, especializaram-se em apenas um subconjunto dos mesmos. Nesse caso, o uso da translog não é possível. Este é também o problema para a aplicação do teste de economias de escopo apresentado por Friedlaender, Winston e Wang (1983). Basicamente, o teste consiste em verificar a subaditividade da função custo, da forma:

$$S_e = \frac{C(Y_T) + C(Y_{N-T}) - C(Y_N)}{C(Y_N)} \quad (73)$$

Como  $C(Y_T)$  e  $C(Y_{N-T})$  implicam calcular a função custo em pontos onde alguns dos produtos apresentam quantidade nula, essas duas funções custo, na especificação translog, são indeterminadas, o que inviabiliza o teste.

- Alguns outros problemas também já foram apontados no uso da especificação translog. Guilkey e Lovell (1980) observam que a forma translog tenderia a superestimar os retornos de escala da função custo e pequenas variações na especificação apresentariam diferenças significativas na estimativa desses retornos de escala.

- Uma forma de contornar os problemas de indeterminação da função translog no 0 seria utilizar a especificação quadrática, como é feito por Friedlaender, Winston e Wang (1983). No entanto, a especificação quadrática também tem seus reveses: os erros de aproximação da especificação quadrática em pontos afastados do original são maiores que os erros de aproximação obtidos com a translog.

- Como foi apontado anteriormente, se os preços dos insumos forem não-lineares, a derivação da demanda dos fatores apresentará outros termos que impossibilitarão a identificação dos parâmetros citados, usando-se apenas dados de qualidade, custos, demanda de fatores e preços dos insumos.

- Ainda sobre preços dos insumos: em dados de painel, para a mesma indústria, muitas vezes não se observa variabilidade suficiente nesses preços de forma a garantir a identificação do sistema citado. Uma forma de contornar esse problema seria tomar dados de firmas afastadas regionalmente ou de indústrias diferentes mas que utilizam o mesmo insumo. Entretanto, caberia a pergunta: por que as firmas, então, não se fixariam no lugar onde os insumos são mais baratos? Isto é, não haveria correlação entre esses preços e outras variáveis omitidas na estimação?

### **Resumo**

Objetivo: Estimar a função custo hedônica de um produto.

Motivação: Em vez de estimar a função custo para cada produto como um bem distinto (no caso de produtos diferenciados), considera-se cada produto como sendo um conjunto de características comuns a todos. Isso eleva a eficiência da estimação, pois aumenta o tamanho da base de dados considerada na estimação e reduz o número de parâmetros a serem estimados.

Variáveis: No modelo utilizado por eles, considera-se uma *cross-section* de firmas para um dado ano, tomando-se as seguintes variáveis:

- custo total de produção;
- características do produto (por exemplo, dimensões, potência, vida útil esperada, consumo de energia, entre outras);
- preço dos insumos (salários, materiais, entre outros); e
- fração do custo correspondente à demanda de cada insumo.

Metodologia: Estimação conjunta da função custo (assumida translog) e das demandas por fatores por mínimos quadrados generalizados (GLS). Para a construção do índice hedônico também assume-se uma especificação translog [ver equações (66) até (69)].

Aplicabilidade: Mercados de produtos diferenciados, onde a quantidade de produtos pode ser muito grande mas o número de características é limitado.

Vantagens:

- Reduz o número de parâmetros a serem estimados (se comparado com a estimação de uma função custo para cada produto distinto).

• É um método paramétrico, então pode ser facilmente implementado com qualquer pacote estatístico. Além disso, os testes sobre economias de escala e escopo passam a ser sobre a média dos parâmetros.

- Função translog é uma aproximação de segunda-ordem mais robusta que a função quadrática.

- A princípio, as variáveis necessárias para a estimação são de fácil coleta.

Desvantagens:

- Função translog não é definida no 0 (isso pode ser problemático se alguns produtos não apresentarem algumas características ou se, em algum período, a produção for 0).

- Função translog parece superestimar retornos de escala.

- Uso das equações de demandas por fatores: pode ser que preços sejam não-lineares ou haja algum poder de monopsônio, o que viesaria os resultados.

- Preços dos insumos, em geral, apresentam pouca variação, o que não permite a identificação do problema.

#### 2.3.4.2 Evans e Heckman

Esses autores apresentam um método para testar subaditividade da função custo no contexto de funções multiproduto. Apesar de o objetivo do artigo ser verificar a existência ou não de monopólio natural na indústria de telefonia americana, a metodologia se aplica a qualquer indústria ao avaliar economias de escala ou escopo. O artigo é bastante instrutivo, principalmente por apontar problemas nas especificações que a literatura até então assumia. Eles começam assumindo que a função custo da AT&T — até então monopolista do mercado norte-americano — é dada por:

$$C = f(L, T, r, m, w, t) \quad (74)$$

onde  $L$  é a quantidade de chamadas locais produzidas,  $T$  é a quantidade de chamadas a distância produzidas,  $r$  é o preço do serviço do capital,  $m$  é o preço dos materiais,  $w$  é o salário médio e  $t$  é um índice de mudança tecnológica. Ao estimar a função custo em vez da função de produção, os autores assumem a validade da teoria da dualidade bem como desejam manter-se fiéis à literatura de monopólio natural, a qual se baseia nas propriedades da função custo. Eles também explicam que desagregaram os produtos entre chamadas locais e a distância porque esses

seriam os dois maiores serviços providos pela firma. Inicialmente, apresentam as estimativas da função custo para uma especificação translog:

$$\begin{aligned} \ln C = & \alpha_0 + \sum_{i=1}^M \alpha_i \ln p_i + \sum_{k=1}^N \beta_k \ln q_k + \mu \ln t + \\ & + \frac{1}{2} \left[ \sum_{i=1}^M \sum_{j=1}^M \gamma_{ij} \ln p_i \ln p_j + \sum_{k=1}^N \sum_{l=1}^N \delta_{kl} \ln q_k \ln q_l \right] + \\ & + \sum_{i=1}^M \sum_{k=1}^N \rho_{ik} \ln p_i \ln q_k + \sum_{i=1}^M \lambda_i \ln p_i \ln t + \sum_{k=1}^N \theta_k \ln q_k \ln t + \tau (\ln t)^2 \quad (75) \end{aligned}$$

onde  $p$  representa o vetor de preços dos insumos ( $r, w, m$ ) e  $q$  representa o vetor de produtos ( $L, T$ ). Como, por hipótese, vale a teoria da dualidade, podemos aplicar o lema de Shepard:

$$x_i = \frac{\partial C}{\partial p_i} \quad (76)$$

onde  $x_i$  é a quantidade demandada do fator de produção  $i$  e assim obtemos as demais equações de demanda dos fatores

$$\frac{p_i x_i}{C} = \alpha_i + \sum_{j=1}^M \gamma_{ij} \ln p_j + \sum_{k=1}^N \rho_{ik} \ln q_k + \lambda_i \ln t \quad (77)$$

Se a firma (no caso, AT&T) opera eficientemente, isto é, na fronteira do conjunto de possibilidades de produção, então a teoria de produção implica certas restrições nos coeficientes do sistema de equações acima, que seriam:

a) Homogênea de primeiro grau no preço dos insumos

$$\sum_{j=1}^M \alpha_j = 1, \sum_{j=1}^M \gamma_{ij} = 0, \sum_{k=1}^N \rho_{ik} = 0, \sum_{i=1}^N \lambda_i = 0 \quad (78)$$

b) A matriz hessiana da função custo com respeito aos preços dos insumos é simétrica

$$\gamma_{ij} = \gamma_{ji} \quad (79)$$

c) Os  $\alpha$ ,  $\rho$ ,  $\gamma$  e  $\lambda$  são idênticos entre as equações.

Evans e Heckman, então, adicionam mais duas hipóteses ao modelo: separabilidade no espaço de produtos e aditividade em custos. A primeira implica dizer que podemos agregar os diversos produtos em um único índice, isto é,

$$C(L, T, r, m, w, t) = C(A(L, T), r, m, w, t) = C(Q, r, m, w, t) \quad (80)$$

Em termos da especificação translog, a função custo será separável se:

$$\rho_{il} \beta_k = \rho_{ik} \beta_l, \quad i=1, 2, 3, \quad k \neq 1 \quad (81)$$

Se essas restrições forem válidas, então é possível agregar todos os produtos como sendo apenas uma quantidade total, estimar a função custo desse único produto e testar para a presença de economias de escala de forma a verificar se a indústria pode ser caracterizada como monopólio natural. É importante observar que o teste mostrado serve apenas para rejeitar ou não a hipótese de separabilidade, mas não serve para propor formas específicas de agregação do produto.

Aditividade em custos implica que o custo de produzir vários produtos na mesma planta ou firma é o mesmo que a soma dos custos de produção desses mesmos bens em diferentes plantas, ou seja,

$$C(L, T, r, m, w, t) = C_L(L, r, m, w, t) + C_T(T, r, m, w, t) \quad (82)$$

Novamente, para o caso da especificação translog, *nonjointness* implica a seguinte restrição nos parâmetros:

$$\delta_{kl} = -\beta_k \beta_l \quad (83)$$

Se essa restrição for aceita, então a função custo não apresenta nem economias nem deseconomias de escopo.

Além de utilizar a especificação mostrada, os autores estimam a função custo da AT&T usando também a especificação translog modificada e Box-Tidwell, a qual nada mais é do que uma transformação Box-Cox nas variáveis independentes. Se chamarmos de  $g(\cdot)$  a função translog presente em (75), então:

- Translog

$$\ln C = g(\ln L, \ln T, \ln r, \ln w, \ln m, \ln t) \quad (84)$$

- Translog modificada

$$\ln C = g(L^*, T^*, \ln r, \ln w, \ln m, \ln t) \quad (85)$$

- Box-Tidwell

$$\ln C = g(L^*, T^*, r^*, w^*, m^*, t^*) \quad (86)$$

$$x^* = \frac{x^\eta - 1}{\eta}$$

onde a variável estrela é dada pela transformação Box-Cox.

As três especificações da função custo são estimadas com dados anuais do sistema Bell, de 1947 a 1977. Basicamente, para a especificação translog, os autores rejeitam a hipótese de simetria e homogeneidade em preços dos insumos — o que eles alegam que pode ser um sinal tanto de que a especificação translog não é apropriada ou que a função custo apresenta algum outro tipo de má especificação, como por exemplo, variáveis omitidas. Entretanto, considerando-se o modelo restrito, onde essas restrições valeriam, eles também rejeitam a hipótese de separabilidade.

O artigo termina apresentando o teste para subaditividade da função custo. Aqui eles desejam testar o caso mais simples, ou seja, qual cenário apresentaria o menor custo: uma firma produzindo dois bens ou duas firmas produzindo cada uma um dos bens. Então, toma-se  $Q_t^* = (Q_{1t}^*, Q_{2t}^*)$  como sendo o vetor de bens

produzidos num dado ano  $t$ . Considere-se também o vetor  $\bar{Q}_M = (Q_{1M}, Q_{2M}) = (\min Q_{1t}^*, \min Q_{2t}^*)$  que representa o menor vetor de produção presente na amostra. A firma  $A$  produz

$$\bar{Q}_{At} = (\phi Q_{1t} + Q_{1M}, \omega Q_{2t} + Q_{2M}) \quad (87)$$

enquanto a firma  $B$  produz

$$\bar{Q}_{Bt} = ((1-\phi)Q_{1t} + Q_{1M}, (1-\omega)Q_{2t} + Q_{2M}) \quad (88)$$

Os parâmetros  $\phi$  e  $\omega$  pertencem ao intervalo  $[0, 1]$ . Além disso, para evitar extrapolar a função custo em intervalos não-observados nos dados, eles impõem que a produção seja restrita aos valores observados, isto é,

$$R_L \leq \frac{\phi Q_{1t} + Q_{1M}}{\omega Q_{2t} + Q_{2M}} \leq R_U \quad (89)$$

$$R_L \leq \frac{(1-\phi)Q_{1t} + Q_{1M}}{(1-\omega)Q_{2t} + Q_{2M}} \leq R_U$$

onde  $R_L = \min(Q_{1t}/Q_{2t})$  e  $R_U = \max(Q_{1t}/Q_{2t})$ , sendo que o máximo e o mínimo se referem ao tempo. Juntas, as firmas  $A$  e  $B$  produzem

$$Q_{1t} + 2Q_{1M} = Q_{1t}^* \quad (90)$$

$$Q_{2t} + 2Q_{2M} = Q_{2t}^*$$

logo:

$$Q_{1t} = Q_{1t}^* - 2Q_{1M} \quad (91)$$

$$Q_{2t} = Q_{2t}^* - 2Q_{2M}$$

Essa alocação é possível somente se  $Q_i^* > 2Q_{iM}$ . Eles restringem o teste apenas em níveis de produção que respeitem essa restrição. Por fim, tome-se

$$C_{At}(\phi, \omega) = C(Q_{At}) \quad (92)$$

$$C_{Bt}(\phi, \omega) = C(Q_{Bt})$$

$$C_t^* = C(Q_{At} + Q_{Bt}) = C(Q_t^*)$$

O grau de subaditividade da função custo é dado por:

$$Sub_t(\phi, \omega) = \frac{C_t^* - C_{At}(\phi, \omega) - C_{Bt}(\phi, \omega)}{C_t^*} \quad (93)$$

Se  $Sub_t(\cdot)$  for menor que 0, a indústria apresenta subaditividade de custos. O teste funciona da seguinte forma: calcula-se  $Sub_t(\phi, \omega)$ . Se  $\max Sub_t(\phi, \omega)$  for negativo e estatisticamente significativo diferente de 0, então podemos aceitar localmente a hipótese de subaditividade de custos. Se, por outro lado, o valor de  $\max Sub_t(\phi, \omega)$  não for significativamente diferente de 0, então rejeita-se a hipótese de subaditividade de custos — mas não se pode rejeitar a hipótese de que a função custo seja localmente aditiva.

### Resumo

Objetivo: Testar a existência de monopólio natural.

Motivação: Idealmente, um teste de monopólio natural deveria comparar os custos de se produzir o mesmo bem em duas firmas distintas com o custo de consolidar a produção na mesma firma. Entretanto, no caso da telefonia norte-americana até a década de 1980, os dados disponíveis eram apenas dos custos da produção no monopólio. Os autores então propõem uma forma de gerar o contrafactual, isto é, qual seria o custo de produção em duas firmas e usam esses valores estimados para testar a existência de monopólio natural.

Variáveis: No modelo utilizado por eles, considera-se uma série temporal de uma única firma, tomando-se as seguintes variáveis:

- custo total de produção;
- quantidade produzida de cada produto (em termos de receita e deflacionada pelo preço dos mesmos);
- preço dos insumos (salários, materiais, entre outros); e
- fração do custo correspondente à demanda de cada insumo.

Metodologia: Estimação conjunta da função custo e das demandas por fatores por GLS segundo as especificações translog, translog modificada e Box-Tidwell. As firmas são consideradas multiproduto, mas não se usa a abordagem hedônica. Para o teste de monopólio natural, constrói-se uma série de valores para o custo de cada firma separada, conforme descrito a partir da equação (87).

Aplicabilidade: Mercados de produtos homogêneos, firmas multiproduto e teste de economias de escopo.

Vantagens:

- É um método paramétrico, então pode ser facilmente implementado com qualquer pacote estatístico.
- Propõe uma forma alternativa de testar economias de escopo.
- Função translog é uma aproximação de segunda ordem mais robusta que a função quadrática.
- A princípio, as variáveis necessárias para a estimação são de fácil coleta.

Desvantagens:

- Uso das equações de demandas por fatores: pode ser que preços sejam não-lineares ou haja algum poder de monopsônio, o que viesaria os resultados.
- Preços dos insumos, em geral, apresentam pouca variação, o que não permite a identificação do problema.

#### **2.3.4.3 Berry, Kortum e Pakes**

O objetivo deste artigo está em relacionar como mudanças no ambiente regulatório afetaram custos de produção e características de produtos na indústria automobilística. De certa forma, é um artigo relacionado com Berry, Levinsohn e Pakes (1995), mas do lado da oferta. Para isso, eles estimam uma função custo hedônica, a qual relaciona o custo de produção de cada bem com suas características. A motivação

deles é apontar que, ao não controlar pela diferenciação nos produtos, a estimativa da função custo pode ser viesada, principalmente porque variações em custos devido a mudanças nas características dos produtos podem ser classificadas como mudanças em produtividade.

Os autores utilizam dados de custo de cada planta, bem como os produtos feitos em cada uma delas e suas características. Entretanto, como o número de observações por produto é limitado, não é possível estimar uma função custo para cada produto, mas sim para cada planta. Eles definem a função custo como dependendo de capital, mão-de-obra e materiais e focam na estimativa da demanda de materiais. Isso porque a maior parte dos custos de produção está relacionada com materiais e, dos três insumos, este é o mais provável de estar sendo escolhido de forma ótima no curto prazo. Mão-de-obra e capital estão sujeitos a custos de ajustamento e, portanto, utilizar as hipóteses de minimização de custo estática poderia viesar a estimativa.

Dessa forma, eles consideram que a escolha de materiais é condicionada em um índice de mão-de-obra e capital. Esse índice pode variar para cada produto, dependendo de suas características e ao longo do tempo, e será definido como  $G(L, K, x, t)$ . Assim, assumindo-se que a produção é linear na quantidade de materiais usada, a demanda por materiais,  $M$ , também será linear no produto. O coeficiente da equação da demanda de materiais,  $c(x_j, \varepsilon_{pt}, \beta)$ , é função das características do produto,  $x_j$ , da produtividade da planta,  $\varepsilon_{pt}$  e dos parâmetros a serem estimados. Berry, Kortum e Pakes (1996), consideraram uma relação linear

$$c = x_j \beta + \varepsilon_p \quad (94)$$

Por fim, os autores assumem um choque de produtividade específico na dimensão temporal,  $\delta_t$ . Esse termo serviria para captar mudanças na tecnologia. A função de produção, então, será dada por

$$Q_{jpt} = \min \left\{ \frac{M}{\delta_t c(x_j, \varepsilon_{pt}, \beta)}, G(L, K, x, t) \right\} \quad (95)$$

Assim, a demanda por materiais que sai do custo variável de produzir o bem  $j$  na planta  $p$  no instante  $t$  é dada por:

$$M_{jpt} = \delta_t c(x_j, \varepsilon_{pt}, \beta) Q_{jpt} \quad (96)$$

Apesar de os custos variáveis serem constantes, existe a possibilidade de retornos crescentes de escala via componente fixo de custo. Esse componente fixo é chamado de  $\mu$ . Além disso, pode haver algum custo fixo de produzir mais de um produto na planta  $p$ , o qual eles representam por  $\Delta$ . Seja  $J(p)$  o conjunto de modelos produzidos na planta  $p$  e  $J_p$  como sendo o número de modelos produzidos nela. Então o uso total de fatores na planta  $p$  no instante  $t$  é dado por:

$$M_{pt} = \mu + J_{pt}\Delta + \sum_{i=1}^J M_{ipt} \quad (97)$$

Dividindo-se (97) pela produção por planta, obtemos a equação a ser estimada

$$\frac{M_{pt}}{Q_{pt}} = \frac{\mu}{Q_{pt}} + \frac{J_{pt}}{Q_{pt}}\Delta + \delta_t (\bar{x}_{pt} \beta + \epsilon_{pt}) \quad (98)$$

onde:

$$\bar{x}_{pt} = \sum_{i=1}^J x_{ipt} \frac{Q_{ipt}}{Q_{pt}} \quad (99)$$

Com exceção das *dummies* temporais  $\delta$ , a equação anterior pode ser estimada por Mínimos Quadrados Ordinários (MQO) [Ordinary Least Square (OLS)]. Os autores alertam para o fato de o modelo citado ignorar uma série de fatos econômicos e econometrícios como, por exemplo, a produção de cada planta ser tomada como exógena e servir de ponderação para o lado direito da regressão. Essa hipótese pode ser muito restritiva se considerarmos que a produtividade pode ser correlacionada no tempo e, portanto, a firma teria algum tipo de informação sobre quais modelos alocar para cada planta. Além disso, nos cálculos, eles ignoraram heterocedasticidade e correlação no termo  $\epsilon_{pt}$  entre as plantas, e a especificação dos custos, apesar de permitir a existência de custos fixos, não admite outras formas de retorno de escala. Por fim, não incorrem em maiores especificações sobre como acontece a substituição entre materiais e capital e labor.

### Resumo

Objetivo: Estimar a produtividade de diversas plantas utilizando-se uma função custo hedônica.

**Motivação:** Os autores argumentam que, ao corrigir para as características dos bens, evita-se que variações na qualidade/características sejam tomadas como variações na produtividade das plantas. Além disso, eles estimam usando apenas a função de custo variável, na forma de demanda por insumos, pois estes são os mais prováveis de estarem sendo escolhidos de acordo com a minimização de custos de curto prazo (mão-de-obra e capital possuem grandes custos de ajustamento).

**Variáveis:** No modelo utilizado, considera-se um painel de plantas, tomando-se as seguintes variáveis (anuais):

- custo total de produção;
- quantidade produzida de cada produto em cada planta;
- características de cada produto (no caso, por ser mercado automotivo, tomam como sendo características: peso, potência, presença de ar-condicionado, consumo, nível de emissões, entre outros); e
- preço dos insumos (no caso, custo do aço, motores, transmissão, e energia).

**Metodologia:** Estimação das demandas por fatores de cada planta por MQO. O índice hedônico é linear nas características e é dado pelas equações (94) e (96) e a equação a ser estimada por MQO é (98).

**Aplicabilidade:** Mercados de produtos diferenciados e que o custo variável seja responsável por uma grande parcela do custo total de produção.

**Vantagens:**

- Reduz o número de parâmetros a serem estimados (se comparado com a estimativa de uma função custo para cada produto distinto).
- É um método paramétrico, então pode ser facilmente implementado com qualquer pacote estatístico.
- Consegue estimar o custo marginal de produção de cada planta.

**Desvantagens:**

- Mede apenas o custo devido ao uso de insumos; se esse custo não for representativo comparado com o custo total de produção, então esse método não tem aplicabilidade.
- A produção de cada planta é tomada como sendo exógena (pode haver viés de seleção).
- Retorno de escala existe apenas por conta do componente de custo fixo na utilização de cada insumo; não permite outras formas de retornos de escala.

- Não permite calcular a existência de economias de escopo.
- Utiliza dados de plantas, o que pode ser difícil de ser obtido de todas as participantes do mercado.

### 2.3.5 Estimação da função de produção

A prática da estimação de funções de produção, como apontam Griliches e Mairesse (1995), começou em 1928, com o trabalho seminal de Cobb e Douglas. Inicialmente, o seu uso estava limitado à macroeconomia, mas com o aumento da disponibilidade de dados sobre firmas e plantas, a estimação dessas funções passou a ser uma ferramenta importante para a medida de produtividade e tecnologia das firmas (retornos de escala e escopo).

Todavia, desde o início, sempre houve muita discussão sobre a metodologia e a validade das estimativas obtidas por essas funções. Griliches e Mairesse enumeram várias das críticas, tais como problemas quanto à forma funcional assumida, erros de mensuração, diferenças de qualidade/atributos tanto nos insumos como nos produtos, entre outros. Aqui, nesta introdução ao assunto, seguiremos a exposição desses autores, até mesmo porque os artigos mais recentes sobre o tema procuram contornar os problemas levantados por eles.

Griliches e Mairesse concentram-se em entender se as hipóteses utilizadas a respeito da estrutura de mercado e comportamento das firmas, nos artigos publicados até então, faziam sentido e poderiam ser testadas. O primeiro dos problemas apontados pelos autores se refere à simultaneidade, o que remete ao artigo de Marschak e Andrews (1944). Basicamente, o procedimento utilizado até então, estimar via MQO uma função de produção utilizando dados de produto e insumos, gerava estimativas inconsistentes, dado que a escolha dos insumos estava relacionada a variáveis não-observadas pelo econometrista — produtividade, por exemplo.

O problema de simultaneidade pode ser melhor entendido escrevendo o termo de erro da regressão em três componentes:

$$u_{it} = \alpha_{it} + e_{it} + \varepsilon_{it} \quad (100)$$

onde  $\alpha$  representa choques, ou variáveis omitidas, observáveis pela firma no momento da escolha dos fatores;  $e$  representa outros choques que não são observados pela firma *ex ante* a sua escolha de fatores, mas que interferem na produção realizada em  $t$ . Portanto,  $\alpha$  e  $e$  são componentes observados pela firma, mas não pelo econometrista. Por outro lado,  $\varepsilon_{it}$  inclui todos os erros devido à mensuração.

A grande vantagem de separar os termos do erro dessa forma está em saber que, enquanto os dois primeiros termos estão correlacionados com a escolha de insumos, o último não sofre desse problema. Logo, a fim de identificar todos os parâmetros da função de produção, cabe ao econometrista encontrar formas de controlar por esses erros.

A primeira forma encontrada para resolver os problemas de endogeneidade apresentados foi o uso de dados em painel. Assumindo-se que os erros de especificação são, em sua maioria, fixos ao longo do tempo, a função de produção poderia ser escrita na forma:

$$\gamma_{it} = \alpha z_{it} + \beta x_{it} + \gamma_i + \lambda_t + u_{it} \quad (101)$$

onde  $\gamma$  e  $\lambda$  representam efeitos fixos com respeito à firma e ao tempo, respectivamente. Dessa forma, subtraindo as médias das variáveis com respeito à firma e ao tempo e dispondo de variáveis que fossem não-correlacionadas ao termo  $u_p$ , o sistema poderia ser estimado por Método dos Momentos Generalizados [Generalised Method of Moments (GMM)], da forma:

$$\gamma_{it} - \bar{\gamma}_i = \alpha(z_{it} - \bar{z}_i) + \beta(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i \quad (102)$$

onde a notação  $x_i$  representa a média da variável  $x$  com respeito ao tempo, para cada firma. Essa é a essência do método de Chamberlain (1982) e Arellano e Bond (1991), o qual será discutido com mais cuidado ao apresentarmos o procedimento de Blundell e Bond (2000). Imaginava-se que, ao escrever o problema dessa forma, se resolveria o problema de simultaneidade. Infelizmente, as estimativas com dados de painel apresentaram muito longe do que a profissão considera razoável: os coeficientes para o capital se mostravam baixos e insignificantes, enquanto as estimativas de retorno de escala estavam em níveis bem inferiores ao que se imaginava.

O problema que Griliches e Mairesse apontam é que a maior parte dos instrumentos disponíveis para o uso dessas técnicas possui pouco poder. Até então, a maior parte dos instrumentos utilizados, quando possuíam poder, não apresentavam uma boa justificativa comportamental para o seu uso. Os autores então propunham o uso do método de diferenças em diferenças para tentar contornar os problemas gerados agora pelo termo de erro  $u_p$ , o qual é a média dos erros passados e possui, assim como em (100), fatores observados pelas firmas e não-observados pelo econometrista. O resultado foram estimativas ainda mais baixas

para o coeficiente do capital, bem como rejeição da hipótese de retornos constantes de escala.

Os autores então apontam para a idéia de Olley e Pakes (1996). Estes últimos argumentam que, ao utilizar dados de painel para tentar corrigir os problemas de simultaneidade, os autores acabaram por introduzir um problema de seleção: ao utilizarem painéis balanceados, acabavam por estimar os coeficientes apenas das firmas que conseguiam sobreviver na indústria. Utilizando-se então um procedimento semiparamétrico para corrigir esse viés de seleção, as estimativas dos coeficientes voltam a ser significativas e mais próximas ao que diz a intuição econômica.

Griliches e Mairesse também discutem a respeito do uso de valores de vendas deflacionados como indicador para quantidades. Muitas vezes os preços individuais não são observados e assume-se algum tipo de índice, comum a todas as firmas das indústrias, para deflacionar os valores das vendas e assim obter-se os valores de quantidades.

Klette e Griliches (1994) mostram que, se as firmas apresentarem algum tipo de poder de mercado sobre o bem produzido (que é justamente o caso quando tratamos de antitruste), as estimativas feitas usando-se essa forma de deflacionamento resultarão em retornos decrescentes de escala, mesmo quando a tecnologia for retornos crescentes de escala.

Motivados pelas idéias de Griliches e Mairesse, apresentamos agora um resumo dos artigos mais recentes que procuram incorporar essas idéias. Com exceção de Blundell e Bond (2000), tanto Olley e Pakes (1996) e Levinsohn e Petrin (2003) tentam incorporar os problemas de simultaneidade e seleção na estimação da função de produção, para isso valendo-se de diferentes instrumentos e ou abordagens. Alguns outros artigos, como por exemplo Ackerberg, Caves e Frazer (2006) e Aguirregabiria (2004), propõem outras formas de lidar com o problema, talvez até mais apropriadas para casos onde não há muita disponibilidade de dados, e preços não podem ser usados como instrumentos. No entanto, são artigos ainda muito preliminares e seus resultados ainda não foram testados. Preferimos nos fixar nesses artigos que já são de maior conhecimento dentro da profissão.

### **2.3.5.1 Blundell e Bond**

O objetivo de Blundell e Bond (2000) é entender o porquê dos problemas levantados por Griliches e Mairesse a respeito das estimativas dos coeficientes da função de produção ao utilizar GMM: os coeficientes do capital apresentam-se em geral baixos e não-significativos, ao mesmo tempo que as estimativas de retorno de escala estão muito aquém do que a teoria considera razoável.

Eles então estendem o conhecido modelo de Arellano e Bond (1991), apontando que os problemas gerados pela estimação via GMM, utilizando primeiras diferenças, estariam no fato de haver correlação entre as taxas de crescimento de quantidades produzidas e demanda de insumos e os valores passados dessas variáveis.

A função de produção a ser estimada é uma Cobb-Douglas.

$$\gamma_{it} = \beta_l l_{it} + \beta_k k_{it} + \gamma_t + (\eta_i + v_{it} + m_{it}) \quad (103)$$

$$v_{it} = \rho v_{i,t-1} + e_{it} \quad |\rho| < 1$$

onde  $\beta_l$  e  $\beta_k$  são os coeficientes da mão-de-obra e capital, respectivamente;  $\gamma_t$  reflete um choque tecnológico comum às firmas no ano  $t$ ;  $\eta_i$  é um efeito específico da firma;  $v_{it}$  representa um choque de produtividade (possivelmente autoregressivo); e  $m_{it}$  simboliza todo tipo de erro de mensuração. Tanto  $m_{it}$  como  $e_{it}$  são assumidos independentes ao longo do tempo. É fácil perceber a analogia entre a estrutura de erros proposta por Griliches e Mairesse que está presente em (103).

O modelo anterior pode ser reescrito na forma dinâmica

$$\begin{aligned} \gamma_{it} = & \beta_l l_{it} - \rho \beta_l l_{i,t-1} + \beta_k k_{it} + \rho \beta_k k_{i,t-1} + \rho y_{i,t-1} + (\gamma_t - \rho \gamma_{i,t-1}) + \\ & + (\eta_i (1-\rho) + e_{it} + m_{it} - \rho m_{i,t-1}) \end{aligned} \quad (104)$$

ou:

$$\gamma_{it} = \pi_1 l_{it} - \pi_2 l_{i,t-1} + \pi_3 k_{it} + \pi_4 k_{i,t-1} + \pi_5 y_{i,t-1} + \gamma^* + (\eta_i^* + \omega_{it}) \quad (105)$$

sujeito às seguintes restrições:  $\pi_2 = -\pi_1 \pi_5$  e  $\pi_4 = -\pi_3 \pi_5$ . Além disso, é importante notar que  $\omega_{it} = e_{it} + m_{it} - \rho m_{i,t-1}$ , logo é um MA(1).

O procedimento de estimação de (104) proposto por Arellano e Bond consiste em assumir as seguintes condições de momento:

$$E[x_{i,t-s} \Delta \omega_{it}] = 0 \quad (106)$$

onde  $x_{it} = (l_{it}, k_{it}, y_{it})$  e  $s \geq 3$ . É importante lembrar aqui a validade dos valores passados de quantidades e insumos como instrumentos na estimação. Basicamente, assume-se que as escolhas feitas  $s$  períodos antes de  $t$  não têm impacto nas escolhas em  $t$ . Claramente, para dados de mais alta freqüência essa hipótese pode não ser verdadeira, se considerarmos a existência de contratos de trabalho por período determinado ou mesmo que a variação no capital instalado é reduzida, senão nula. Como apontam os autores, essa é uma das razões para os resultados controversos das estimativas por GMM apresentadas até então.

Blundell e Bond propõem o uso de mais uma condição de momento:

$$E\left[\Delta x_{i,t-s} (\eta_i^* + \omega_{it})\right] = 0 \quad (107)$$

para  $s = 2$ . Eles então demonstram a validade dessa condição de momento e apresentam os resultados da estimação para um painel de firmas norte-americanas, entre 1982 e 1989. Diferentemente dos resultados obtidos pelo método de Arellano e Bond, os autores não conseguem rejeitar a hipótese de retornos constantes de escala; além disso, o coeficiente do capital aumenta significativamente, aproximando-se dos valores anteriormente obtidos via MQO.

Note-se que esta última condição de momento não é compatível com o caso de aprendizado (*learning by doing*), uma vez que o termo  $v_{it}$  passaria a ser dependente do histórico de produção. Em termos do modelo apresentado, seria quebrar a hipótese de independência temporal do termo de erro  $e_{it}$ , o qual está presente em  $\omega_{it}$ , invalidando-se a última condição de momento. Caberia ao pesquisador avaliar quão razoável seria assumir essa hipótese. Além disso, como foi mostrado anteriormente, os autores não mencionam possíveis problemas de seleção que possam estar acontecendo devido a entrada e saída de firmas do painel. Esses problemas ficarão mais claros a seguir, ao discutirmos o artigo de Olley e Pakes.

### Resumo

**Objetivo:** Estimar a função de produção de um bem homogêneo usando-se dados de painel.

**Motivação:** Os autores procuram corrigir os resultados pouco intuitivos quando se estima a função de produção usando-se dados de painel (para cada planta): coeficiente do capital muito baixo (e não-significativo), inexistência de retornos de escala. Para isso, propõe-se mais uma condição de momento a ser utilizada na estimação por GMM.

Variáveis: No modelo utilizado pelos autores, considera-se um painel de firmas, dados anuais:

- quantidade produzida (em termos de receita e deflacionada pelo preço do bem em cada ano);
- estoque de capital [construído a partir da série de investimentos e descontada a depreciação — aplica-se a equação (107)]; e
- mão-de-obra é medida deflacionando-se o custo com folha de pagamentos pelo salário médio.

Metodologia: Estimação da função de produção por GMM. Eles assumem que a função de produção é Cobb-Douglas e os instrumentos que utilizam são as quantidades defasadas dos insumos e a variação em sua quantidade. Essas condições estão representadas pelas equações (105) até (107).

Aplicabilidade: Mercados de produtos homogêneos onde dados de painel para cada planta estejam disponíveis e não se observem entradas ou saídas de firmas no período.

#### Vantagens:

É relativamente simples de ser implementado. A maioria dos pacotes estatísticos já possui um módulo que permite a estimativa por GMM. Além disso, a especificação funcional é linear, o que facilita a implementação.

#### Desvantagens:

- Se a amostra apresentar bastante variabilidade em termos de entrada e saída de firmas, as estimativas apresentarão viés de seleção (apenas as firmas mais produtivas sobrevivem no período).
- Ignora aprendizado (pode haver correlação serial no erro, o que viesaria também as estimativas).
- Dificuldades podem surgir na medida do estoque de capital.
- Não controlam por qualidade da mão-de-obra.

#### **2.3.5.2 Olley e Pakes**

O objetivo desses autores é estimar os parâmetros da função de produção e, a partir dessas estimativas, analisar a evolução da produtividade das plantas na indústria de equipamentos de telecomunicação após a desregulamentação do mercado de telefonia norte-americano. Eles observam que, durante os anos seguintes à desregulamentação, houve um grande número de entradas na indústria de

equipamentos, bem como várias firmas e plantas foram fechadas. Como a decisão de entrar ou permanecer em uma indústria está relacionada à produtividade da firma e esta é uma variável não-observada, isso gera problemas de simultaneidade e seleção na determinação das estimativas da função de produção.

O primeiro problema (simultaneidade) ocorre porque as firmas observam a sua produtividade no momento da sua demanda por insumos. O segundo problema decorre de atritos na base de dados. Até então, na medida em que a disponibilidade de dados de painel aumentou, os estudos que se baseavam nesses painéis consideraram apenas as firmas que permaneciam na amostra durante todo o período coberto. Como as firmas decidem sua permanência ou não baseadas na sua produtividade, então o uso de um painel balanceado criaria um viés de seleção (as estimativas da função de produção serão das firmas com maior produtividade).

Para analisar esse problema, os autores baseiam-se nos modelos dinâmicos de Ericson e Pakes (1995) e Hopenhayn e Rogerson (1993), os quais apresentam um modelo de entrada e saída. Basicamente eles assumem que os lucros de uma firma dependem de algumas variáveis de estado da firma e de suas concorrentes, bem como do vetor de preços dos insumos. Esse vetor de estados consiste em três variáveis:  $a_t$ , idade da firma no instante  $t$ ;  $k_t$ , estoque de capital da firma; e  $\omega_t$ , um índice de eficiência da mesma. Os preços dos fatores são supostos iguais entre as firmas e evoluem exogenamente de acordo com um processo de Markov.

No começo do período, uma firma participante do mercado pode tomar três decisões: primeiro, continuar ou sair da indústria. Se sair, ela recebe um valor  $\phi$  e não aparece novamente. Se continuar, ela escolhe investimento e fatores variáveis — basicamente mão-de-obra — e investimento mais estoque de capital presente determinam o estoque de capital do período seguinte. Logo, a dinâmica de capital é a usual

$$k_{t+1} = (1 - \delta)k_t + i_t \quad (108)$$

Dado esse ambiente, a equação de Bellman que a firma resolve é:

$$V(\omega, a, k) = \max \left\{ \phi, \sup_{i \geq 0} \pi(\omega, a, k) - c(i) + \beta E[V(\omega', a', k')] \right\} \quad (109)$$

onde  $\pi(\cdot)$  representa a função lucro e  $c(\cdot)$  o custo total do investimento  $i$ . A solução desse problema gera uma regra de saída e uma equação para a demanda de investimento,  $i_t = i_t(\omega_t, a_t, k_t)$ .

Dada essa demanda por investimento, podemos falar sobre a estimação da função de produção propriamente dita. Olley e Pakes assumem que a indústria produz um bem homogêneo com tecnologia Cobb-Douglas e com produtividade fator neutro. Assim, a função de produção a ser estimada tem a forma

$$y_{it} = \beta_0 + \beta_a a_{it} + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \eta_{it} \quad (110)$$

onde  $y_{it}$  é o logaritmo do produto da planta  $i$  no ano  $t$ ;  $a_{it}$  é a idade da planta;  $k_{it}$  é o logartimo do estoque de capital;  $l_{it}$  o logaritmo da mão-de-obra;  $\omega_{it}$  a produtividade; e  $\eta_{it}$  pode ser interpretado tanto como erro de medida como um choque de produtividade não antecipado pela firma.

A inconsistência do MQO é clara se  $l$  é correlacionado com  $\omega$ , o que não é observado. Esse ponto já foi analisado tanto por Griliches como pelo artigo clássico de Marschak e Andrews (1944). O problema de seleção é um pouco mais sutil. Imagine, por simplicidade, que não há fatores variáveis. A esperança condicional de  $y_{it}$  (condicional nos fatores, na sobrevivência da firma até o instante  $t$  e na informação até  $t - 1$ ) inclui o termo:

$$E[\omega_t | a_t, k_t, \omega_{t-1}, \chi_t = 1] \quad (111)$$

Como  $\chi_t = 1$  se e somente se  $\omega_t = \underline{\omega}_t(a_t, k_t)$ . Além disso, se a função lucro for crescente em  $k$ , a função valor também o será e, consequentemente,  $\underline{\omega}_t(.)$  será decrescente em  $k$ . Isso quer dizer que firmas maiores devem permanecer no mercado mesmo para valores mais baixos de produtividade, o que implica que a esperança (111) deve ser decrescente em  $k$  e trazer um viés negativo no coeficiente do capital.

O algoritmo de Olley e Pakes consiste no seguinte: para valores positivos do investimento, Pakes (1994) mostra que a demanda por investimento é estritamente crescente em  $\omega$ . Assim é possível inverter a função (111) e escrever:

$$\omega_t = h_t(i_t, a_t, k_t) \quad (112)$$

Logo, podemos escrever a variável não-observada  $\omega$  como função das observáveis. Substituindo-se (112) na equação da produção, temos:

$$y_{it} = \beta_l l_{it} + \phi_t(i_{it}, a_{it}, k_{it}) + \eta_{it} \quad (113)$$

onde:

$$\phi_t(i_{it}, a_{it}, k_{it}) = \beta_0 + \beta_a a_{it} + \beta_k k_{it} + h_t(i_t, a_t, k_t) \quad (114)$$

A equação (114) é uma regressão semiparamétrica a qual consegue identificar  $\beta_l$ , mas não identifica  $\beta_a$  ou  $\beta_k$ . Para a identificação desses dois últimos termos, Olley e Pakes utilizam as probabilidades de sobrevivência, as quais são dadas por:

$$\begin{aligned} & \Pr\{\chi_{t+1} = 1 | \underline{\omega}_{t+1}(a_{t+1}, k_{t+1}), J_t\} = \\ &= \Pr\{\omega_{t+1} \geq \underline{\omega}_{t+1}(a_{t+1}, k_{t+1}) | \underline{\omega}_{t+1}(a_{t+1}, k_{t+1}), \omega_t\} = \\ &= \varphi_t\{\underline{\omega}_{t+1}(a_{t+1}, k_{t+1}), \omega_t\} \\ &= \varphi_t(i_t, a_t, k_t) \equiv P_t \end{aligned} \quad (115)$$

A esperança de  $y_{t+1} - \beta_l l_{t+1}$  condicional na informação em  $t$  e sobrevivência é dada por:

$$\begin{aligned} & E[y_{t+1} - \beta_l l_{t+1} | a_{t+1}, k_{t+1}, \chi_{t+1} = 1] = \\ &= \beta_0 + \beta_a a_{t+1} + \beta_k k_{t+1} + E[\omega_{t+1} | \omega_t, \chi_{t+1} = 1] = \\ &= \beta_a a_{t+1} + \beta_k k_{t+1} + g(\underline{\omega}_{t+1}, \omega_t) \end{aligned} \quad (116)$$

onde:

$$g(\underline{\omega}_{t+1}, \omega_t) = \beta_0 + \int_{\underline{\omega}_{t+1}}^{\omega_{t+1}} \frac{F(d\omega_{t+1} | \omega_t)}{\int_{\underline{\omega}_{t+1}}^{\omega_{t+1}} F(d\omega_{t+1} | \omega_t)} \quad (117)$$

Substituindo-se  $P_t$  e  $\phi_t$  em  $g(\cdot)$  e escrevendo  $\xi_{t+1}$  como sendo a inovação em  $\omega_{t+1}$ , temos:

$$y_{t+1} - \beta_l l_{t+1} = \beta_a a_{t+1} + \beta_k k_{t+1} + g(P_t, \phi_t - \beta_a a_t + \beta_k k_t) + \xi_{t+1} + \eta_{t+1} \quad (118)$$

Então estima-se o sistema composto por (114), (115) e (117). E estima-se  $\phi$  por um polinômio de quarta ordem em  $i$  e  $k$  e  $P$ , por um *probit*.

O algoritmo de Olley e Pakes foi a primeira tentativa de contornar os problemas de seleção presentes quando se estima a função de produção a partir de um painel de firmas ou plantas. Todavia, como Griliches e Mairesse apontam, ele se baseia em uma hipótese forte: existe apenas um componente não-observável pelo econometrista,  $\omega_{it}$ , o qual segue um processo de Markov de primeira ordem e é totalmente repassado à demanda do investimento. Qualquer choque macro é incorporado na variação temporal da função  $\phi$ . No entanto, é razoável imaginar que a demanda por investimento seja dependente das expectativas sobre a taxa de juros e sobre a demanda futura. Incorporar essas variáveis poderia invalidar a hipótese de monotonicidade da função  $h$ , e assim o modelo deixaria de ser identificado.

### **Resumo**

**Objetivo:** Estimar a função de produção de um bem homogêneo usando-se dados de painel e corrigindo-se para sobrevivência.

**Motivação:** Os autores procuraram corrigir os resultados pouco intuitivos quando se estima a função de produção usando-se dados de painel (para cada planta): coeficiente do capital muito baixo (e não significativo), inexistência de retornos de escala. Para isso, eles propõem utilizar dados de investimento e o cálculo da probabilidade de sobrevivência de cada firma de forma a corrigir para o problema de seleção: as firmas que sobrevivem são aquelas que possuem maior produtividade e firmas com maior produtividade tendem a apresentar maiores estoques de capital.

**Variáveis:** No modelo utilizado, considera-se um painel de plantas, dados anuais:

- quantidade produzida (em termos de valor adicionado — receita menos custo de materiais. A receita é deflacionada pelo índice de preços da indústria e o custo dos materiais é deflacionado por um índice de preço médio dos mesmos);
- estoque de capital [construído a partir da série de investimentos e descontada a depreciação — aplica-se a equação (108)]. Note-se que é necessário atribuir um valor inicial para o capital. Se observada a entrada da planta, toma-se o valor de mercado da planta na entrada. Caso contrário, toma-se o primeiro valor observado do valor de mercado da firma como a condição inicial;
- mão-de-obra é medida deflacionando-se o custo com folha de pagamentos pelo salário médio;

- investimento é construído também deflacionando-se o custo total do investimento pelo preço médio do investimento na indústria; e
- anos de existência de cada planta.

**Metodologia:** Estimação da função de produção via um método semiparamétrico de dois estágios. Num primeiro estágio os autores calculam a probabilidade de sobrevivência das firmas usando um *probit*, sendo as variáveis explicativas o tempo de existência da firma, o investimento e o estoque de capital. Ao mesmo tempo, eles estimam uma função de produção linear cujo único regressor seria a mão-de-obra. Usam então o resíduo dessa regressão como a variável dependente da regressão do capital, onde as variáveis explicativas seriam capital, tempo de existência da firma e uma função da probabilidade de sobrevivência da firma. Essas condições estão representadas pelas equações (114), (115) e (117).

**Aplicabilidade:** Mercados de produtos homogêneos onde dados de painel para cada planta estejam disponíveis se forem observadas entradas ou saídas de firmas no período. Além disso, é preciso observar valores positivos de investimento para as firmas ao longo do período, bem como deve-se ter disponível o instante em que cada firma entrou ou saiu do mercado.

#### Vantagens:

- Consegue corrigir para o viés de seleção que surge quando se estimam funções de produção utilizando-se painéis balanceados.
- Não depende da validade do lema de Shepard.
- Permite identificar quais as firmas/plantas com maior produtividade, o que pode ser útil como indicador da planta que será utilizada após a fusão (e assim determinar o ganho de eficiência devido à fusão).

#### Desvantagens:

- Requer uma base de dados rica em informações sobre investimento, bem como sobre quais firmas estão presentes em cada instante.
- Ainda sobre investimento: ao trabalhar com dados de plantas, pode-se observar várias delas com investimento nulo, pois as firmas tendem a concentrar seus investimentos em algumas plantas de cada vez. Como visto, a identificação do modelo depende de os valores de investimento serem não-nulos.
- Ignora choques agregados que podem interferir na produtividade e no investimento.
- Processo de Markov pode ser uma hipótese forte.

- Dificuldades podem surgir na medida do estoque de capital.
- Não controlam por qualidade da mão-de-obra.

### 2.3.5.3 Levinsohn e Petrin

Esses autores partem do mesmo arcabouço teórico que Olley e Pakes, mas utilizam insumos, ao contrário de investimento, para controlar as não-observáveis. A principal vantagem apontada pelos autores é que a abordagem anterior só é válida se as firmas reportarem valores positivos de investimento (como os próprios Olley e Pakes já tinham observado em seu modelo estrutural). Como, na maior parte das indústrias, as firmas costumam reportar 0 para investimento — e uma das razões para se supor isso está relacionada aos custos de ajustamento desse investimento —, truncar essas firmas pode causar outros vieses. Por outro lado, as firmas quase sempre reportam valores de outros insumos, o que evita esse problema de truncamento.

Os autores começam assumindo que a função de produção pode ser escrita como:

$$y_{it} = f(x_{it}, \varepsilon_{it}, \beta) \quad (119)$$

onde  $x_{it}$  representa o vetor de insumos;  $\beta$  o conjunto de parâmetros; e  $\varepsilon_{it}$  é um choque de produtividade, suposto Hicks neutro. Eles então escrevem a função de produção como sendo

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \beta_m m_t + \omega_t + \eta_t \quad (120)$$

A demanda por insumos intermediários pode ser escrita como:

$$m_t = m_t(\omega_t, k_t) \quad (121)$$

e deve ser monotônica em  $\omega_t$  para se qualificar como um bom instrumento. Os autores também assumem que os preços dos insumos são idênticos entre as firmas e que não há erro na equação que defina a demanda por insumos intermediários. Assumindo-se que vale monotonicidade, podemos inverter (121) e assim controlar por  $\omega_t$  e obter

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + h(m_t, k_t) + \eta_t \quad (122)$$

Da mesma forma que Olley e Pakes, (122) pode ser estimada por procedimentos triviais, tais como MQO e assim obter-se as estimativas para o coeficiente da mão-de-obra. O segundo estágio consiste na estimação do coeficiente do capital. Para tal, vamos escrever o valor da produção para o período seguinte:

$$\gamma_{t+1} = \beta_0 + \beta_l l_{t+1} + \beta_k k_{t+1} + E[\omega_{t+1} | I_t] + \xi_{t+1} + \eta_{t+1} \quad (123)$$

onde o termo  $E[\omega_{t+1} | I_t]$  representa o valor esperado da produtividade em  $t+1$  condicionado na informação disponível em  $t$ . A equação acima gera a seguinte condição de momento

$$E[\xi_{t+1} + \eta_{t+1} | k_{t+1}] = 0 \quad (124)$$

uma vez que o estoque de capital em  $t+1$  não é correlacionado, por definição, aos choques  $\xi_{t+1}$  e  $\eta_{t+1}$  (lembra que  $\xi_{t+1} + \eta_{t+1}$  representam toda a variação em produtividade e erro de mensuração que não podem ser explicados pelo estoque de capital, por isso constitui uma restrição de momento válida para a identificação de  $\beta_k$ ). A estimativa não-paramétrica de  $h(m_t, k)$  é idêntica àquela apresentada em Olley e Pakes e, portanto, não será repetida aqui.

Como visto, o procedimento de Levinsohn e Petrin é bastante similar ao empregado em Olley e Pakes, com a vantagem de que dados de demanda de insumos intermediários são mais prováveis de serem não-nulos, o que ajudaria na estratégia de identificação do coeficiente do capital.

Claramente, alguns problemas presentes na crítica de Griliches e Mairesse ainda estão presentes nesses dois últimos artigos. Um dos mais sérios diz respeito à mensuração das quantidades. Se os valores físicos das quantidades produzidas e dos insumos demandados não forem observados, e não houver dados sobre os preços individuais de cada uma dessas quantidades, então problemas como aqueles apontado por Klette e Griliches podem viesar bastante os resultados.

Como dissemos anteriormente, alguns artigos recentes têm procurado identificar problemas nos algoritmos de Olley e Pakes e Levinsohn e Petrin. Ackerberg e Caves (2004), por exemplo, argumentam que, se levarmos em consideração a cronologia da escolha dos diferentes insumos e da produção, a equação de momento (124) presente em Levinsohn e Petrin e Olley e Pakes não seria válida. Basicamente, os autores argumentam que ao condicionar a escolha de investimento, ou insumos intermediários, apenas por um escalar (no caso, o choque de produtividade), os

dois modelos requereriam uma estrutura temporal de escolha dos insumos mais complexa (por exemplo, mão-de-obra antes de intermediários). Entretanto, se considerarmos esses tipos de estruturas, a condição de momento citada não seria válida, a menos que a escolha de intermediários acontecesse simultaneamente com a produção.

### **Resumo**

**Objetivo:** Estimar a função de produção de um bem homogêneo usando-se dados de painel e corrigindo-se para sobrevivência.

**Motivação:** Os autores procuram corrigir os resultados pouco intuitivos quando se estima a função de produção usando-se dados de painel (para cada planta): coeficiente do capital muito baixo (e não-significativo), inexistência de retornos de escala. Para isso, eles propõem utilizar dados de demanda por insumos e o cálculo da probabilidade de sobrevivência de cada firma de forma a corrigir para o problema de seleção: as firmas que sobrevivem são aquelas que possuem maior produtividade e firmas com maior produtividade tendem a apresentar maiores estoques de capital.

**Variáveis:** No modelo utilizado por eles, considera-se um painel de plantas, dados anuais:

- Quantidade produzida: a receita é deflacionada pelo índice de preços da indústria.
- Estoque de capital [construído a partir da série de investimentos e descontada a depreciação — aplica-se a equação (108)]. Note-se que é necessário atribuir um valor inicial para o capital. Se observada a entrada da planta, toma-se o valor de mercado da planta na entrada. Caso contrário, toma-se o primeiro valor observado do valor de mercado da firma como condição inicial.
- Mão-de-obra é medida deflacionando-se o custo com folha de pagamentos pelo salário médio.
- Custo de materiais deflacionado pelo preço médio de materiais na indústria.
- Anos de existência de cada planta.

**Metodologia:** Estimação da função de produção via método semiparamétrico de dois estágios. Num primeiro estágio eles calculam a probabilidade de sobrevivência das firmas usando-se um *probit*, sendo as variáveis explicativas o tempo de existência da firma, a demanda por insumos e o estoque de capital. Ao mesmo tempo, estimam uma função de produção linear cujo único regressor seria a mão-de-obra. Eles então usam o resíduo dessa regressão como a variável dependente da

regressão do capital, onde as variáveis explicativas seriam capital, tempo de existência da firma e uma função da probabilidade de sobrevivência da firma. Essas condições estão representadas pelas equações (122) e (124).

**Aplicabilidade:** Mercados de produtos homogêneos onde dados de painel para cada planta estejam disponíveis se forem observadas entradas ou saídas de firmas no período. Além disso, é preciso observar valores positivos de demanda por insumos para as firmas ao longo do período, bem como deve-se ter disponível o instante em que cada firma entrou ou saiu do mercado.

Vantagens:

- Consegue corrigir para o viés de seleção que surge quando se estimam funções de produção utilizando-se painéis balanceados.
- Não depende da hipótese de investimento positivo para garantir identificação. Como a maior parte dos processos produtivos requer algum tipo de insumo diferente de mão-de-obra e capital, o método é quase sempre identificado.
- Não depende da validade do lema de Shepard.
- Permite identificar quais as firmas/plantas com maior produtividade, o que pode ser útil como indicador da planta que será utilizada após a fusão (e assim determinar o ganho de eficiência devido à fusão).

Desvantagens:

- Requer uma base de dados rica em informações sobre quais as firmas que estão presentes em cada instante.
- Ignora choques agregados que podem interferir na produtividade e no investimento.
- Processo de Markov pode ser uma hipótese forte.
- Dificuldades podem surgir na medida do estoque de capital.
- Não controlam por qualidade da mão-de-obra.

### 2.3.6 Implementação

Até aqui, apresentamos formas de estimar tanto funções custo como funções de produção, apontando as vantagens e os problemas de cada método. Três perguntas cabem no final desta subseção:

- Qual o melhor método indicado para analisar a função custo das firmas fusionadas?

- Quais variáveis são necessárias para a estimação?
- Como implementar a estimação?

Não há uma resposta única para a primeira pergunta. Como acontece na maioria dos problemas econômicos, os detalhes do ambiente estudado e a disponibilidade de dados vão ditar a técnica a ser utilizada para a estimação.

Imaginamos que algumas variáveis já são (ou podem ser) obtidas das firmas fusionadas. São elas:

- receita, por produto;
- preço, por produto;
- custo total de produção;
- valor do estoque de capital físico instalado;
- gasto total com folha de pagamentos;
- número de empregados; e
- quantidade demandada de eletricidade ou outra forma de energia.

No caso de uma firma manufatureira, podemos ainda pedir:

- quantidade demandada de materiais — por exemplo, alumínio, aço, plástico etc.; e

- preço dos materiais ou valor gasto nos mesmos.

A razão de incluirmos a demanda por energia está no fato de que a melhor medida para a função de produção seria o capital utilizado no instante  $t$  para a produção de  $y_t$ . O estoque de capital nem sempre apresenta-se como a melhor medida, pois não informa qual a intensidade no uso do capital instalado. A demanda por eletricidade pode dar uma noção muito melhor sobre esse número, uma vez que a grande maioria das firmas necessita de energia para operar. No entanto, é razoável imaginar que essa demanda de eletricidade por unidade de capital utilizado seja crescente na idade das máquinas, isto é, equipamentos mais antigos tendem a ser menos eficientes, o que poderia viesar o coeficiente do capital na função de produção. Então, a melhor idéia seria construir um índice de capital físico utilizado que levasse em conta a idade dos equipamentos, mas dificilmente o analista terá acesso a uma informação como essa. De qualquer forma, no caso de estimação da função de produção, seria conveniente usar a demanda por energia como sendo um instrumento para o capital.

A implementação do algoritmo de Olley e Pakes ou Levinsohn e Petrin para o problema específico de análise antitruste é bastante comprometida. Como pode

ser visto na seção correspondente ao método, a estimativa do coeficiente do capital é baseada no cálculo da probabilidade de sobrevivência da firma na indústria. Se os dados de custo e produção disponíveis forem apenas aqueles das firmas fusionadas, então não teremos como calcular essa probabilidade de sobrevivência. Lembramos novamente os resultados de Griliches e Mairesse: estimar a função de produção usando-se apenas as séries temporais das firmas sobreviventes na indústria (que seria o caso se utilizarmos apenas dados das firmas fusionadas) resultará em estimativas viesadas no coeficiente do capital. O que a literatura mostra é que esse coeficiente será baixo ou até mesmo não-significativo e as estimativas de retorno de escala indicarão quase sempre retornos decrescentes de escala. Dessa forma, a estimativa da função de produção usando-se apenas os dados das firmas fusionadas vai indicar, quase certamente, que a operação não apresenta eficiências, o que não necessariamente será verdadeiro.

Devido à escassez de dados, vamos refazer a primeira pergunta: o que é factível de ser estimado com as variáveis mostradas e apenas para as firmas fusionadas? Acreditamos que seja estimar diretamente a função custo. Vamos aqui novamente indicar as etapas do método.

1) O primeiro passo seria determinar os preços dos insumos a serem utilizados na estimativa. Como já mencionamos no início da seção referente a custos, as séries temporais de preços de insumos para as firmas (ou mesmo para a indústria) apresentam muito pouca variabilidade. Isso impediria a identificação do modelo. Uma possível solução para esse problema seria escolher instrumentos que permitissem medir esses preços com mais acuidade. Como instrumentos queremos dizer variáveis que estejam correlacionadas aos preços dos insumos mas não estejam correlacionadas com a produção do bem final. Uma escolha natural seria observar a demanda por esses mesmos insumos por parte de outras indústrias que não a que estamos considerando. A idéia aqui é que o preço final do insumo é resultado da demanda por parte de todas as indústrias que necessitam daquele insumo, mas a demanda das outras indústrias pelo insumo está correlacionada à demanda pelo produto final dessas outras indústrias, a qual é independente da demanda pelos produtos das firmas fusionadas.

Os dados de demanda dos insumos materiais podem ser obtidos a partir dos fornecedores das firmas fusionadas. Basicamente, é necessário saber dos fornecedores:

- quantidade total demandada dos insumos; ou
- receita total correspondente àquele insumo, bem como o preço médio de venda.

Assim, o preço de cada um dos insumos pode ser obtido como:

$$q_{kt} = \alpha_k + \alpha_Q Q_{kt} + \omega_{kt} \quad (125)$$

onde  $Q_{kt}$  é a quantidade total demandada do insumo  $k$  no instante  $t$ . A partir dos valores estimados, constrói-se a série de preços, da forma:

$$\hat{q}_{kt} = \hat{\alpha}_k + \hat{\alpha}_Q Q_{kt} \quad (126)$$

onde  $(\alpha_k, \alpha_Q)$  representa os parâmetros estimados.

Entretanto, resta saber quais os preços a serem utilizados com respeito a mão-de-obra e capital. O salário médio, obtido pela razão gasto total com mão-de-obra/número de funcionários, além de ser possivelmente uma série com pouca variabilidade (se tomarmos dados mensais) esconde a qualidade da mão-de-obra. Alguns autores, como Hausman, argumentam que se pode utilizar como instrumento o salário médio de firmas semelhantes em outras regiões do país. Então, por exemplo, se pudermos dividir os gastos com mão-de-obra entre funcionários na produção e na administração, poderíamos pesquisar o salário médio desses dois tipos de funcionários nas outras firmas que fazem parte da indústria, mas que se localizem em outras regiões. Não é claro se esses dados seriam possíveis de serem obtidos. Além disso, a hipótese que está por trás é de que os custos de transporte para esses funcionários é grande o suficiente para que as firmas localizadas nessas duas regiões não estejam competindo pelo mesmo trabalhador. Não é claro também que essa hipótese seja sempre válida; além disso, muitas firmas podem estar localizadas na mesma região, o que impediria a técnica de ser utilizada. Como qualquer tentativa empírica, estamos dependentes da qualidade dos dados.

O preço do capital pode ser obtido a partir de:

- valores locais do aluguel do m<sup>2</sup> e área instalada;
- quantidade mensal dos estoques e preço do bem final; e
- taxas de juros de mercado para firmas do porte das fusionadas.

Como sabemos a quantidade total de capital instalado, podemos calcular as parcelas de cada um dos três itens citados na composição do capital fixo (instalações, estoques e capital instalado) e assim montar um índice do preço do capital. Vamos lembrar mais uma vez que não há garantias que mesmo construindo-se índices dessa forma teremos a variabilidade necessária para identificar o modelo; vai

depender seriamente dos dados disponíveis. Outro ponto que estamos assumindo é que salários e preço do capital são exógenos às escolhas da firma. Isso pode não ser verdade, se por exemplo o capital instalado for específico de cada firma ou se a mão-de-obra também for específica naquela indústria. Entretanto, com a disponibilidade de dados que temos, é o melhor que podemos fazer.

2) Com os dados citados, estimamos a seguinte especificação quadrática:

$$\begin{aligned} C_{it} = & \alpha_i + \sum_{j=1}^N \alpha_j p_{ij} + \sum_{k=1}^M \beta_k q_{ik} + \frac{1}{2} \left[ \sum_{j=1}^N \sum_{r=1}^N \gamma_{jr} p_{ij} p_{ir} + \sum_{k=1}^M \sum_{l=1}^M \delta_{kl} q_{ik} q_{il} \right] + \\ & + \sum_{j=1}^N \sum_{k=1}^M \rho_{jk} p_{ij} q_{ik} \end{aligned} \quad (127)$$

onde  $i$  representa cada firma;  $j$  representa cada um dos insumos utilizados pela firma  $i$ ;  $k$  representa cada um dos produtos feitos pela firma  $i$ ; e  $q_{ik}$  é a quantidade produzida de cada um dos produtos. Se a firma analisada produzir apenas um produto, então todos os somatórios em  $k$  presentes na equação (127) desaparecem e simplifica-se bastante a estimativa.

Por que usar a especificação quadrática? A razão é aquela apontada na descrição do artigo de Spady e Friedlaender (1978): como temos produtos diferenciados, cada uma das firmas não estará produzindo quantidades positivas de todos os bens. Como o logaritmo de 0 não é definido, o uso da especificação translog é inviabilizado. Obviamente, abrir mão da especificação translog tem seu custo: a especificação quadrática é uma aproximação de segunda ordem de pior qualidade que a translog. Dizemos isso porque os erros de aproximação, para pontos afastados daqueles onde estamos medindo, são maiores na especificação quadrática.

Cabe então a pergunta: por que não estimar uma especificação translog para cada firma (pois, nesse caso, ambas possuiriam quantidades positivas)? Aqui estamos supondo que os coeficientes são iguais em ambas as firmas, ou seja, estamos assumindo que existe apenas uma tecnologia e qualquer economia de custos seria exclusivamente por meio de economias de escala ou escopo. Isto é uma simplificação, mas que tem como objetivo reduzir o número de parâmetros a serem estimados (o que ajuda em termos de eficiência do estimador) e também enriquece em termos de intuição na construção do contrafactual: se temos apenas uma tecnologia, então a firma resultante continuará com essa mesma tecnologia, mas operando em diferente escala. Essa diferença é que será responsável pelos ganhos de eficiência da fusão. A estimação de duas translogs requereria uma quantidade maior de dados,

bem como complicaria na montagem do contrafactual, pois qual das duas funções custo seria a função custo da firma resultante após a fusão?

Não vamos incluir as equações de demanda de fatores na nossa estimação. Isso porque a inclusão das mesmas implica dizer que as hipóteses do lema de Shepard estão valendo, isto é, os preços dos insumos são lineares e as firmas não têm poder de monopsônio. Como já apontamos anteriormente, os argumentos levantados pelas firmas fusionadas quando fazem a defesa das eficiências apontam justamente que essas hipóteses não estão valendo, ao dizer que os preços dos insumos serão alterados devido à nova escala. Dessa forma, a estimação é conduzida apenas com a função custo apresentada anteriormente.

O sistema composto pela equação de custo de cada firma é estimado conjuntamente. A estimação conjunta é importante em termos da eficiência das estimativas: como nosso objetivo final é obter um intervalo de confiança, eficiência passa a ser um ponto importante a ser obtido. Vale lembrar novamente que os preços dos insumos que estamos utilizando foram calculados por meio de variáveis instrumentais, como indicamos no item anterior.

3) Com os valores estimados dos parâmetros, podemos testar a presença de retornos de escala derivando-se a expressão anterior com respeito às quantidades produzidas dos bens em questão

$$S(q, p) = \frac{C(q, p)}{\sum_{k=1}^M q_k C_k(q, p)} \quad (128)$$

onde o custo marginal  $C_k$  é obtido diretamente da expressão dos custos

$$MC_k = \frac{\partial C}{\partial q_k} = \beta_k + \sum_{l=1}^M \delta_{kl} (q_{1l} + q_{2l}) + \sum_{j=1}^N \rho_{jk} (p_{1j} + p_{2j}) \quad (129)$$

Como a função custo estimada é comum às duas firmas, o valor que utilizamos como sendo  $C(q, p)$  pode ser o estimado do custo de produzir conjuntamente as quantidades observadas  $q_k$  de cada uma das firmas, isto é,  $C(q, p) = C(q_1 + q_2, p)$ . O valor das economias de escala citadas possui sua própria variância (a qual é calculada pelo método delta, uma vez que a especificação é totalmente linear). Com esse valor da variância podemos testar se  $S(q, p)$  é diferente de 1.

A presença ou não de economias de escopo pode ser obtida testando se

$$SC_T(q_1 + q_2) = \frac{C(q_1) + C(q_2) - C(q_1 + q_2)}{C(q_1 + q_2)} \quad (130)$$

é maior do que 1. Os valores  $C(q_1)$  e  $C(q_2)$  são aqueles que observamos como sendo os custos totais das firmas fusionadas. O valor  $C(q_1 + q_2)$  é obtido utilizando-se a função custo estimada e aplicando-se no ponto  $q_1 + q_2$ . Como antes, essa estatística tem uma variância que pode ser calculada pelo método delta, considerando-se a variância dos parâmetros estimados. Então, testa-se a hipótese de  $SC_T(q_1 + q_2)$  ser maior que 1.

4) Uma vez com os parâmetros da função custo estimados, também podemos fazer outros tipos de exercícios. Como indicamos anteriormente, podemos utilizar as informações prestadas pelas firmas a respeito de algumas não-linearidades nos preços dos insumos (por exemplo, desconto por quantidade, ou mesmo incluir os custos de transporte de cada insumo como parte do preço do mesmo) e usar esses novos preços para prever qual seria a nova função custo da firma resultante.

Todas as críticas a esse procedimento que apresentamos anteriormente continuam sendo válidas. Isto é, podemos incorrer em sérios erros de medida, principalmente dos preços dos insumos, a especificação quadrática pode não ser a mais apropriada, a tecnologia das duas firmas pode ser distinta, portanto, a restrição de os parâmetros da função custo ser igual para ambas as firmas pode não estar sendo válida. Entretanto, em um ambiente onde a disponibilidade de dados é crítica, esse é o procedimento mais simples e rápido que um analista possui para responder à pergunta sobre a existência de retornos de escala e escopo.

## 2.4 CONCLUSÃO

Este trabalho teve dois objetivos principais: apresentar como funcionam os modelos de simulação de fusões (suas vantagens e seus problemas) e discutir o que tem sido feito a respeito da verificação de ganhos de eficiência decorrentes da fusão.

À primeira vista, os resultados apresentados não são muito animadores, em nenhuma das duas áreas. A simulação de fusões continua sendo dependente das hipóteses sobre a demanda, custos marginais e interação estratégica entre as firmas. Mesmo com a pesquisa tendo avançado bastante na primeira área (estimação da demanda), os modelos que temos disponíveis hoje em dia para avaliar custos e, principalmente, interação estratégica, ainda são muito preliminares. Cada vez mais

profissionais da área entendem que essas questões somente serão respondidas com bons modelos dinâmicos que possam ser estimados. E isso ainda não é uma realidade.

Mesmo conceitos que se acreditavam ser independentes da forma funcional da demanda (no caso, o CMCR) procuramos mostrar que não o são. Na verdade, o CMCR apresenta dificuldades de interpretação ainda maiores que a mera simulação de preços: uma vez que se determina qual deveria ser a redução em custos marginais devido à fusão, como avaliar se esse número é grande ou pequeno? O uso do CMCR precisa ser acompanhado de uma avaliação cuidadosa dos ganhos de eficiência devido à fusão, o que requer a estimação de funções custo/função de produção.

Por fim, a estimativa de intervalos de confiança para os modelos de simulação ainda não é uma realidade. O procedimento presente em Capps Jr., Church e Love se baseia em hipóteses fortíssimas, sendo a principal delas a exogeneidade dos preços na estimativa do sistema de demanda. Entretanto, acreditamos que é a área com maiores chances de progresso em um curto intervalo de tempo. Se o procedimento do *bootstrap* for realizado tomando-se os devidos cuidados de considerar a endogeneidade dos preços no sistema de demanda (portanto, estimar a demanda usando-se algum tipo de variável instrumental) e se os problemas de correlação serial da demanda forem resolvidos (seja incluindo-se essa especificação no modelo estrutural, seja porque o mercado em análise não apresenta correlação serial), então o *bootstrap* será uma ferramenta útil para a determinação do novo intervalo de confiança.

Então resta a pergunta: pode-se utilizar esse tipo de modelo na instrução de política antitruste? A resposta é um tímido sim. Mesmo com todos os problemas que aqui apontamos, a análise quantitativa dos preços pós-fusão pode ser entendida como mais uma informação à disposição do regulador. Obviamente, calcar a decisão da análise meramente nos resultados obtidos pela simulação é arriscado. Entretanto, cabe ao regulador usar de bom senso para comparar se as hipóteses utilizadas no modelo correspondem ao processo analisado e assim validar os resultados. Além disso, é descobrindo as falhas do método que podemos nos concentrar para aprimorá-lo: dessa forma, ao apontar os revezes do modelo, esperamos estar trazendo à tona uma agenda de pesquisa para adequar os modelos de simulação à realidade do antitruste.

No que diz respeito ao cálculo das eficiências, entendemos haver dois tipos de problemas: disponibilidade de dados e construção de contrafactuals. O primeiro problema acreditamos ser possível de ser resolvido. Como apontamos na seção referente à implementação, as técnicas mais novas de estimativa da função

de produção não serão implementáveis enquanto não dispusermos de dados sobre entrada e saída de firmas nas indústrias. A forma utilizada pelo governo dos Estados Unidos (bem como do Canadá, Chile, Dinamarca, França, Holanda, Israel, Noruega e Suécia) foi implementar um censo industrial. No caso norte-americano, Bartelsman e Doms contam que o governo conduz, a cada cinco anos, uma pesquisa sobre produção, demanda de insumos, estrutura de propriedade e localização contendo dados de aproximadamente 300 mil estabelecimentos. Essa pesquisa é reconduzida anualmente utilizando-se uma subamostra menor (em torno de 50 a 75 mil firmas). Com isso, é possível que o pesquisador acompanhe a vida de várias firmas, desde o seu nascimento até a saída do mercado. Esse tipo de dado é essencial para a identificação da função de produção, como tem sido demonstrado nos estudos mais recentes. Não sabemos dizer qual o custo de implementar um censo dessa natureza, tampouco quais os benefícios (além desses aqui relatados para a instrução de uma análise antitruste), para que essa pesquisa seja implementada.

O segundo ponto, sobre a construção de contrafactuals, é mais delicado. Em um mundo ideal, seríamos capazes de estimar a função de produção da indústria e também a distribuição da produtividade entre as firmas participantes. Então, se construíssemos os novos preços dos insumos após a fusão poderíamos calcular qual seria a função custo da firma fusionada.

O grande problema estaria em determinar como os preços dos insumos variam entre antes e depois da fusão. Mesmo no caso mais simples, onde os preços dos insumos obedecem a uma relação não-linear ditada pelos fornecedores, faz-se necessário estimar como seria essa função dos preços. Como já dissemos anteriormente, pouco tem sido escrito no sentido de medir as interações entre produtores e fornecedores (muito por causa da falta de dados e da complexidade dessas relações). O caso ficaria pior ainda se esse preço fosse resultado da barganha entre produtores e fornecedores. Ainda não temos estratégias empíricas para avaliar esses efeitos.

Com a atual disponibilidade de dados que imaginamos ser possível catalogar, a melhor estratégia para o analista é voltar à estimação da função custo. Não usamos o método proposto por Evans e Heckman porque o problema deles tem uma particularidade: têm dados sobre um monopólio e querem construir o contrafactual do que seria a produção dos bens em firmas diferentes. No nosso caso, observamos a produção em duas firmas e gostaríamos de saber qual seria a produção em apenas uma (a firma fusionada). Mas para isso, podemos utilizar a função custo estimada, dado que ela é comum às duas firmas. Também não indicamos o método de Berry, Kortum e Pakes por um motivo bastante simples: eles se preocupam apenas com a estimação da demanda por materiais. Como uma

fusão pode envolver mudanças significativas no capital instalado e na quantidade de mão-de-obra contratada, esse método não nos parece o mais indicado.

Enfim, acreditamos que a análise antitruste brasileira pode ganhar muito com a implementação de alguns desses modelos e que o entendimento das hipóteses por trás de cada modelo será importante para fundamentar as decisões e esclarecer melhor os estudos apresentados pelas partes durante a instrução dos processos. Esperamos que os modelos apresentados sejam úteis para a instrução dos processos antitruste, bem como as vantagens e problemas de cada método tenham sido bem explicados para a melhor utilização do regulador em suas análises.

## BIBLIOGRAFIA

- ACKERBERG, D. A., CAVES, K. *Structural identification of production functions — an application to the timing of input choice*. University of California at Los Angeles, 2003 (manuscript).
- AGUIRREGABIRIA, V. *Specification and estimation of dynamic factor demand models using a discrete choice approach*. Boston University, 2004 (manuscript).
- ARELLANO, M., BOND, S. Some tests of specification for panel data: Monte-Carlo evidence and an application to employment equations. *Review of Economic Studies*, v. 58, p. 277-297, 1991.
- BAJARI, P., BENKARD, L., LEVIN, J. *Estimating dynamic models of imperfect competition*. NBER, 2004 (Discussion Paper, 10.450).
- BARTELSMAN, E. J., DOMS, M. Understanding productivity: lessons from longitudinal microdata. *Journal of Economic Literature*, v. 38, p. 569-594, 2000.
- BERESTEANU, A. *Nonparametric analysis of cost complementarities in the telecommunications industry*. Duke University, 2004 (manuscript).
- BERRY, S., KORTUM, S. S., PAKES, A. *Environmental change and hedonic cost functions for automobiles*. NBER, 1996 (Discussion Paper, 5.746).
- BERRY, S., LEVINSOHN, J., PAKES, A. Automobile prices in market equilibrium. *Econometrica*, v. 63, p. 841-890, 1995.
- BERRY, S., PAKES, A. Some applications and limitations of recent advances in empirical industrial organization: merger analysis. *American Economic Review*, v. 83, n. 2, p. 247-252, 1993.
- BLUNDELL, R., BOND, S. GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews*, v. 19, p. 321-340, 2000.
- BRESNAHAN, T. F. Duopoly models with consistent conjectures. *American Economic Review*, v. 71, p. 934-945, 1981.
- \_\_\_\_\_. Empirical methods for industries with market power. In: SCHMALENSEE, R., WILLIG, R. (eds.). *Handbook of industrial organization 2*. North-Holland, Amsterdam, 1989.
- BRESNAHAN, T. F., REISS, P. C. Dealer and manufacturer margins. *RAND Journal of Economics*, v. 26, p. 253-268, 1985.

- CAPPS Jr., O., CHURCH, J., LOVE, A. H. Specification issues and confidence intervals in unilateral price effects analysis. *Journal of Econometrics*, v. 113, p. 3-31, 2003.
- CARLTON, D. W., PERLOFF, J. *Modern industrial organization*. 4<sup>th</sup> ed. Addison Wesley Longman, 2004.
- CHAMBERLAIN, G. Multivariate regression models for panel data. *Journal of Econometrics*, v. 18, p. 5-46, 1982.
- CHEONG, K. S., JUDD, K. Mergers and dynamic oligopoly. *Journal of Dynamic Economics and Control*, 2001 (forthcoming).
- CHRISTENSEN, L. R., DALE, W. J., LAU, L. J. Transcendental logarithmic production frontiers. *The Review of Economics and Statistics*, v. 55, p. 28-45, 1973.
- CORTS, K. S. Conduct parameters and the measurement of market power. *Journal of Econometrics*, v. 88, p. 227-250, 1999.
- \_\_\_\_\_. The strategic effects of vertical market structure: common agency and divisionalization in the U.S. motion picture industry. *Journal of Economics and Management Strategy*, v. 10, p. 509-528, 2001.
- CROOKE, P., FROEB, L. M. P., TSCHANTZ, S. *User's guide to simmerger*. Antitrust.org, 2002 (manuscript).
- CROOKE, P., FROEB, L. M., WERDEN, G. J. Effects of assumed demand form on simulated postmerger equilibria. *Review of Industrial Organization*, v. 15, p. 205-217, 1999.
- DENECKERE, R., DAVIDSON, C. Incentives to form coalitions with Bertrand competition. *RAND Journal of Economics*, v. 16, p. 473-486, 1985.
- EPSTEIN, R. J., RUBINFELD, D. L. Merger simulation: a simplified approach with new applications. *Antitrust Law Journal*, v. 69, p. 883-919, 2001.
- ERICSON, R., PAKES, A. Markov-perfect industry dynamics: a framework for empirical work. *RAND Journal of Economics*, v. 95, p. 53-82, 1995.
- EVANS, D. S., HECKMAN, J. J. Multiproduct cost function and natural monopoly tests for the bell system. *Breaking up Bell, essays on industrial organization and regulations*. New York: Elvevier Science Publishing Co., p. 253-283, 1983.
- FARRELL, J., SHAPIRO, C. Horizontal merger: an equilibrium analysis. *American Economic Review*, v. 80, p. 107-126, 1990.
- FREEDMAN, D. A., PETERS, C. Bootstrapping a regression equation: some empirical results. *Journal of the American Statistical Association*, v. 79, p. 97-106, 1984.
- FRIEDLAENDER, A. F., WINSTON, C., WANG, K. Costs, technology and productivity in the U.S. automobile industry. *Bell Journal of Economics*, p. 1-20, 1983.
- FROEB, L. M., TSCHANTZ, S. T., WERDEN, G. J. *Pass-through rates and the price effects of mergers*. Vanderbilt University, 2001 (manuscript).
- GENESOVE, D., MULLIN, W. P. Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *RAND Journal of Economics*, v. 29, p. 355-377, 1998.
- GOWRISANKARAN, G. A dynamic model of endogenous horizontal mergers. *RAND Journal of Economics*, v. 30, p. 56-88, 1999.

- GRILICHES, Z., MAIRESSE, Z. J. *Production functions: the search for identification*. NBER, 1995 (Discussion Paper, 5.067).
- GUILKEY, D. K., LOVELL, C. A. K. On the flexibility of the translog approximation. *International Economic Review*, v. 21, p. 137-147, 1980.
- HASTINGS, J. S. Vertical relationships and competition in retail gasoline markets: empirical evidence from contract changes in Southern California. *American Economic Review*, v. 94, p. 317-328, 2004.
- HAUSMAN, J. A., LEONARD, G. K., ZONA, J. D. Competitive analysis with differentiated products. *Annales d'Economie et Statistique*, v. 34, p. 159-180, 1994.
- HOPENHAYN, H., ROGERSON, R. Job turnover and policy evaluation: a general equilibrium analysis. *Journal of Political Economy*, v. 101, p. 915-938, 1993.
- HOROWITZ, J. L. The bootstrap. *Handbook of econometrics* V. 2001.
- HOSKEN, D. *et alii Demand system estimation and its applications to horizontal merger analysis*. FTC, 2002 (manuscript).
- ICN. *Efficiencies*. Annual Conference of The International Competition Network - ICN. Seul, Coréia do Sul, 2004.
- KATZ, M. Vertical contractual relations. In: SCHMALENSEE, R., WILLIG, R. D. (eds.). *Handbook of industrial organization*, v. I. Amsterdam, North-Holland, p. 655-721, 1989.
- KLETTE, T. J., GRILICHES, Z. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics*, v. 11, p. 343-361, 1996.
- KREPS, D., SCHEINKMAN, J. A. Quantity precommitment and Bertrand competition yield Cournot outcomes. *Bell Journal of Economics*, v. 14, p. 326-337, 1983.
- LEVINSOHN, J., PETRIN, A. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, v. 70, p. 317-341, 2003.
- LINDH, T. The inconsistency of consistent conjectures: coming back to cournot. *Journal of Economic Behaviour and Organization*, v. 18, p. 69-90, 1992.
- MAGGI, G. The value of commitment with imperfect observability and private information. *RAND Journal of Economics*, v. 30, p. 555-574, 1999.
- MARSCHAK, J., ANDREWS, W. Random simultaneous equation and the theory of production. *Econometrica*, v. 12, p. 143-205, 1944.
- McFETRIDGE, D. D. *Efficiency defences. A framework for the design and implementation of competition law and policy*. Paris and Washington, D.C.: OECD and the World Bank, p. 124-140, 1999.
- MORTIMER, J. H. *The effects of revenue sharing contracts on the welfare in vertically separated markets: evidence from the video rental industry*. Harvard University, 2002 (manuscript).
- NEVO, A. Mergers with differentiated products: the case of the ready-to\_eat cereal industry. *RAND Journal of Economics*, v. 31, p. 395-421, 2000.
- OLLEY, S., PAKES, A. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, v. 64, p. 1.263-1.297, 1996.

- PAKES, A. Dynamic strutural models, problems and prospects part II: mixed continuous-discrete control problems, and market interactions. In: SIMS, C. (ed.). *Advances in econometrics*. Cambridge: Cambridge University Press, 1994.
- PAKES, A., MCGUIRE, P. Computing perfect-Markov nash equilibria: numerical implications of a dynamic differentiated product model. *RAND Journal of Economics*, v. 25, p. 555-589, 1994.
- PANZAR, J. C. Technological determinants of firm and industry structure. *Handbook of Industrial Organization*, v. I, p. 3-59, 1989.
- PETERS, C. *Evaluating the performance of merger simulation: evidence from the U.S. airline industry*. Center for the Study of Industrial Organization, Northwestern University, 2003 (Working Paper).
- REISS, P. C., WOLAK, A. Structural econometric modeling: rationales and examples from industrial organization. *Handbook of econometrics*, v. VI. 2005 (forthcoming).
- ROBSON, A. J. Existence of consistent conjectures: comment. *American Economic Review*, v. 73, p. 454-456, 1983.
- SALANT, S., SWITZER, S., REYNOLDS, R. Losses due to merger: the effect of an exogenous change in industry structure on Cournot-Nash equilibrium. *Quarterly Journal of Economics*, v. 98, p. 185-199, 1983.
- SIEG, H. Estimating a bargaining model with asymmetric information: evidence from medical malpractice disputes. *Journal of Political Economy*, v. 108, p. 1.006-1.021, 2000.
- SPADY, R. H., FRIEDLAENDER, A. F. Hedonic cost functions for the regulated trucking industry. *Bell Journal of Economics*, v. 9, p. 159-179, 1978.
- VILLAS-BOAS, S. B. *Vertical contracts between manufacturers and retailers: inference with limited data*. University of California at Berkeley, 2004 (manuscript).
- WERDEN, G. J. A robust test for consumer welfare enhancing mergers among sellers of differentiated products. *Journal of Industrial Economics*, v. 44, p. 409-413, 1996.
- WERDEN, G. J., FROEB, L. M. The effects of mergers in differentiated products industries: logit demand and merger policy. *Journal of Law, Economics & Organization*, v. 10, p. 407-426, 1994.
- \_\_\_\_\_. Simulation as an alternative to structural merger policy in differentiated products industries. In: COATE, M. B., KLEIT, A. N. (eds.). *The economics of the antitrust process*. Kluwer Academic Publishers, Boston, p. 65-88, 1996.
- \_\_\_\_\_. The entry-inducing effects of horizontal mergers: an exploratory analysis. *Journal of Industrial Economics*, v. 46, p. 525-543, 1998.
- WERDEN, G. J., FROEB, L. M., SCHEFFMAN, D. T. *A daubert discipline for merger simulation*. FTC, 2004 (manuscript).

## APÊNDICE

A seguir, apresentamos as três rotinas de simulação, para cada uma das especificações de demanda: CES, AIDS simplificada (sem o termo de efeito renda) e *logit*. Para as duas primeiras é necessário que o usuário forneça:

$p$  = preços pré-fusão de todas as firmas presentes na indústria;

$s$  = participação de mercado pré-fusão de todas as firmas presentes na indústria;

$E$  = matriz de elasticidades pré-fusão de todas as firmas presentes na indústria;

$\text{firmas}$  = vetor com a posição das firmas fusionadas (por exemplo, se os dados das firmas fusionadas encontram-se na posição 1 e 2 do vetor  $p$  ou  $s$ , então  $\text{firmas} = [1 \ 2]$ );

$\text{MCR}$  = vetor que permite ao regulador incluir variações no custo marginal da firma fusionada após a fusão (por exemplo, se o regulador imagina que uma das firmas vai ter seu custo marginal reduzido em 15%, então  $\text{MCR} = [0.15 \ 0 \ 0 \ 0]$ , para o caso de uma indústria com quatro firmas);

Todas as rotinas foram programadas em Matlab.

A implementação das rotinas é uma tarefa bastante simples. Vamos explicar passo a passo como utilizar a rotina CES [as demais rotinas (AIDS e *logit*) obedecem à mesma série de procedimentos].

1. o usuário deve salvar os arquivos ces.m e CPOces.m no diretório *work* do Matlab;

2. uma vez salvas nesse diretório, o usuário deve abri-las. Basta clicar em *open*, selecionar o diretório *work* e então abrir cada uma das duas funções;

3. agora o usuário deve prover as funções os valores de preços, participações e elasticidades da indústria em análise. Isso é feito da seguinte forma:

a) Na janela principal do Matlab, escreva o vetor de preços da indústria. No exemplo considerado no texto, onde todas as quatro firmas possuíam preço igual a 0.07, o usuário deve escrever:  $p = [0.07; 0.07; 0.07; 0.07]$ .<sup>14</sup>

b) Depois inclua os valores de participação das firmas. No exemplo, temos:  $s = [0.634; 0.164; 0.05; 0.152]$ . Importante: a ordem das firmas utilizada ao colocar o primeiro vetor (no caso, preços da AT&T, outras, MCI e Sprint) deve ser mantida para todos os demais vetores e matrizes.

---

14. Para aqueles com poucos conhecimentos da linguagem utilizada no Matlab, os vetores são indicados entre parênteses. Além disso, para obtermos vetores coluna é preciso que os elementos sejam separados por ponto-e-vírgula, como mostramos anteriormente.

c) Inclua a matriz de elasticidades:  $E = [-1.12 \ 0.09 \ 0.03 \ 0.16; \ 0.05 \ -1.33 \ 0.06 \ 0.03; \ 0.61 \ 0.22 \ -1.81 \ 0.03; \ 0.47 \ 0.12 \ 0.04 \ -1.33]$ .<sup>15</sup>

d) Inclua a posição (no vetor de preços) das firmas fusionadas. No nosso caso, como as firmas a se fundirem serão MCI e Sprint, e a posição delas no vetor de preços é 3 e 4, temos que firmas=[3 4].

e) Inclua o MCR. Se o usuário quiser obter qual seria o novo equilíbrio onde as firmas permanecem com o mesmo custo marginal, então MCR=[0 0].

4. Por fim, o usuário deve chamar a função correspondente. Então, para calcular o novo equilíbrio de preços, basta escrever  $x = ces(s, p, E, firmas, MCR)$  e pressionar *enter*. O resultado do novo equilíbrio aparecerá como uma matriz, da seguinte forma:

a) na primeira coluna, o usuário terá o vetor de preços pós-fusão (lembre-se: cada posição do vetor será correspondente à mesma ordem em que são colocados os dados. Então, no nosso exemplo, o primeiro preço será o da AT&T, seguido pelo de outras, e assim por diante);

b) na segunda coluna, o usuário terá o valor das margens pré-fusão;

c) na terceira coluna, o usuário terá o valor do custo marginal após a fusão. Se o valor de MCR fornecido for um vetor de zeros, então o custo marginal pós-fusão será idêntico àquele antes da fusão, pois o usuário estará dizendo que a fusão não vai alterar o CM das firmas; e

d) por fim, a última coluna apresenta o CMCR calculado para as firmas fusionadas. Lembrar que, para as firmas fora da fusão, o CMCR deverá ser zero.

A seguir, disponibilizamos os códigos.

```
function f=ces(s,p,E,firmas,MCR)
global alfa gama MC merger
n=size(E,1);
merger=firmas;
f1=firmas(1);
f2=firmas(2);
gama=E;
```

---

15. Note-se que para separarmos os elementos entre diferentes colunas basta deixarmos um espaço entre eles; a linha seguinte será começada toda vez que colocarmos ponto-e-virgula.

```

for i=1:n
aux=0;
for j=1:n
aux=aux+gama(i,j)*log(p(j));
end
alfa(i)=log(s(i))-aux;
end
A=inv(diag(diag(E,0)))*p;
MCpre=p+A;
MC=diag(1-MCR)*MCpre;
y0=ones(n,1);
preco=fsolve(@CPOces,y0);
margem=(preco-MC)’*inv(diag(preco));
B(1)=p(f1)*(1+E(f1,f1))+p(f2)*E(f2,f1)*s(f2)/s(f1);
B(2)=p(f2)*(1+E(f2,f2))+p(f1)*E(f1,f2)*s(f1)/s(f2);
B=B’;
D=[E(f1,f1) E(f2,f1)*s(f2)/s(f1);E(f1,f2)*s(f1)/s(f2) E(f2,f2)];
MCmerged=inv(D)*B;
MCpos=MCpre;
for i=1:n
if i==f1
MCpos(i)=MCmerged(1);
end
if i==f2
MCpos(i)=MCmerged(2);
end
end
CMCR=inv(diag(MCpre))*(MCpos-MCpre);

```

```
f=[preco margem' MCpos CMCR];
function f=suids(s,p,E,firmas,MCR)
global alfa gama MC merger
s2=s;
p2=p;
E2=E;
merger=firmas;
f1=firmas(1);
f2=firmas(2);
n=size(E,1);
gama=zeros(n,n);
for i=1:n
gama(i,i)=p(i)*s(i);
for j=1:n
gama(i,j)=gama(i,j)+E(i,j)*p(i)*s(i);
end
end
for i=1:n
aux=0;
for j=1:n
aux=aux+gama(i,j)*log(p(j));
end
alfa(i)=s(i)*p(i)-aux;
end
A=inv(diag(diag(E,0)))*p;
MCpre=p+A;
MC=diag(1-MCR)*MCpre;
y0=p;
```

```

preco=fsolve(@CPOsaids,y0);

margem=(preco-MC)*inv(diag(preco));

B(1)=p(f1)*(1+E(f1,f1))+p(f2)*E(f2,f1)*s(f2)/s(f1);

B(2)=p(f2)*(1+E(f2,f2))+p(f1)*E(f1,f2)*s(f1)/s(f2);

B=B';

D=[E(f1,f1) E(f2,f1)*s(f2)/s(f1);E(f1,f2)*s(f1)/s(f2) E(f2,f2)];

MCmerged=inv(D)*B;

MCpos=MCpre;

for i=1:n

if i==f1

MCpos(i)=MCmerged(1);

end

if i==f2

MCpos(i)=MCmerged(2);

end

end

CMCR=inv(diag(MCpre))*(MCpos-MCpre);

f=[preco margem' MCpos CMCR];

```

A função *logit* infere automaticamente a matriz de elasticidades da indústria a partir dos parâmetros  $\beta$ , aqui com o nome beta, e  $\varepsilon$ , aqui com o nome e, que representa a elasticidade-preço da indústria.

Nesse caso, temos uma modificação no passo 3.c

3. c'. inclua a elasticidade da indústria: e = 0.763;

3.c'' . inclua o valor do parâmetro beta. No nosso exemplo, beta = 25;

4. Para calcular o novo equilíbrio de preços, basta escrever x = logit (s, p, e, beta, firmas, MCR) e pressionar *enter*.

```
function f=logit(s,p,e,beta,firmas,MCR)
```

```
global alfa b MC merger
```

```
merger=firmas;
f1=firmas(1);
f2=firmas(2);
b=beta;
n=size(p,1);
pm=p'*s;
POut=e/(b*pm);
P=s*(1-POut);
for i=1:n
    alfa(i)=log(P(i)/POut)+b*p(i);
end
alfa=alfa';
A=b*(1-P);
A=inv(diag(A));
RHS=A*ones(n,1);
MCpre=p-RHS;
MC=diag(1-MCR)*MCpre;
for i=1:n
    for j=1:n
        E(i,j)=beta*p(j)*P(j);
    end
    E(i,i)=E(i,i)-beta*p(i);
end
E
x0=b*ones(n,1);
preco=fsolve(@CPOlogit,x0);
margem=(preco-MC)'*inv(diag(preco));
B(1)=p(f1)*(1+E(f1,f1))+p(f2)*E(f2,f1)*s(f2)/s(f1);
```

```

B(2)=p(f2)*(1+E(f2,f2))+p(f1)*E(f1,f2)*s(f1)/s(f2);
B=B';
D=[E(f1,f1) E(f2,f1)*s(f2)/s(f1);E(f1,f2)*s(f1)/s(f2) E(f2,f2)];
MCmerged=inv(D)*B;
MCpos=MCpre;
for i=1:n
if i==f1
MCpos(i)=MCmerged(1);
end
if i==f2
MCpos(i)=MCmerged(2);
end
end
CMCR=inv(diag(MCpre))*(MCpos-MCpre);
f=[preco margem' MCpos CMCR];

```

As funções a seguir consistem em sub-rotinas que a rotina principal chama para calcular propriamente o equilíbrio pós-fusão.

```

function f=CPOsaids(x)
global alfa gama MC merger
n=length(x);
f1=merger(1);
f2=merger(2);
mi=x-MC;
e=zeros(n,n);
for i=1:n
s(i)=alfa(i);
for j=1:n
s(i)=s(i)+gama(i,j)*log(x(j));

```

```
end
end
for i=1:n
    sys(i)=mi(i)*(gama(i,i)-s(i))/x(i)+s(i);
end
sys(f1)=sys(f1)+mi(f2)*gama(f1,f2)/x(f2);
sys(f2)=sys(f2)+mi(f1)*gama(f2,f1)/x(f1);
f=sys;
function f=CPOces(x)
global alfa gama MC merger
n=length(x);
f1=merger(1);
f2=merger(2);
mi=x-MC;
for i=1:n
    s(i)=alfa(i);
    for j=1:n
        s(i)=s(i)+gama(i,j)*log(x(j));
    end
end
for i=1:n
    sys(i)=mi(i)*gama(i,i)+x(i);
end
sys(3)=sys(3)+mi(4)*gama(4,3)*exp(s(4)-s(3));
sys(4)=sys(4)+mi(3)*gama(3,4)*exp(s(3)-s(4));
f=sys;
function f=CPOlogit(x)
global alfa b MC merger
```

```
n=size(x,1);  
f1=merger(1);  
f2=merger(2);  
mi=x-MC;  
den=1;  
for i=1:n  
den=den+exp(alfa(i)-b*x(i));  
end  
for i=1:n  
q(i)=exp(alfa(i)-b*x(i))/den;  
end  
sys=b*(eye(n,n)-diag(q))*mi-ones(n,1);  
sys(f1)=sys(f1)-mi(f2)*b*q(f2);  
sys(f2)=sys(f2)-mi(f1)*b*q(f1);  
f=sys;
```

# **Estimação de Demanda em Setores Regulados**

**parte 2**



# 3 m o d e l o r a n c i z a ç a o n e s t u d o s e r v i ç o s b a n c á r i o s

Com a estabilidade monetária, ficou mais fácil observar os elevados *spreads* bancários no país. Mais ainda, com um regime monetário de aperto fiscal com elevadas taxas de juros reais, sucessivos e elevados indicadores de lucratividade dos bancos operando no Brasil acabam reforçando a impressão de que os bancos atuam em mercados imperfeitos.

A literatura econômica, por outro lado, mostrou até agora evidências de que o setor é caracterizado por uma estrutura de mercado imperfeita, rejeitando a hipótese de que os bancos brasileiros operam sob a forma de cartel. Nakane, Alencar e Kanczuk, no Capítulo 3, aumentam o grau de conhecimento a respeito dessa polêmica com uma abordagem econométrica que incorpora recentes avanços na modelagem de mercados imperfeitos. Para tal, desenvolvem modelos de demanda *logit* multinomial sob cenários alternativos de concorrência entre bancos — um de concorrência de Bertrand e outro de cartel — para três produtos bancários: *a*) depósitos a prazo; *b*) soma de depósitos à vista e poupança; e *c*) empréstimos.

Quatro modelos de demanda foram estimados, combinando-se diferentes conjuntos de instrumentos e a inclusão (ou não) de interações entre as características do banco e a renda por habitante do município. Os resultados mostram que as elasticidades-preço próprias para a demanda de depósitos a prazo são bastante elevadas, enquanto as elasticidades da demanda por depósitos à vista e de poupança com relação a variações na taxa de juros de depósitos a prazo são bem mais modestas. As demandas por ambos os tipos de depósitos foram encontradas inelásticas com relação a variações nos preços dos serviços bancários.

No lado da oferta de serviços bancários, o teste de poder de mercado apresentado no estudo confronta as margens preço-custo

empiricamente observadas com aquelas preeditas pelos modelos de Bertrand e de cartel. Os resultados apresentados mostram que o modelo de Bertrand descreve bem o comportamento dos preços de serviços bancários, mas superestima o poder de mercado observado na fixação de preços para depósitos a prazo e para empréstimos. Assim, sob bases empíricas e teóricas mais bem fundamentadas que em estudos anteriores, os resultados do estudo de Nakane, Alencar e Kanczuk tendem a confirmar que o setor bancário não atua na forma de cartel.

## CAPÍTULO 3

# DEMANDA POR SERVIÇOS BANCÁRIOS\*

Márcio I. Nakane

Do Departamento de Estudos e Pesquisas do Bacen e do  
Departamento de Economia da USP

Leonardo S. Alencar

Do Departamento de Estudos e Pesquisas do Bacen

Fabio Kanczuk

Do Departamento de Economia da USP

### 3.1 INTRODUÇÃO

O grau de competição a que está sujeito o setor bancário no Brasil é, ainda, uma questão não inteiramente resolvida. A observação de elevados *spreads* bancários no país, bem como a divulgação de generosos indicadores de lucratividade para os principais bancos em operação no Brasil, reforça a opinião popular de que os bancos atuam em mercados imperfeitos.

A literatura disponível que tratou de testar, de forma mais rigorosa, o exercício de poder de mercado pelos bancos brasileiros mostrou evidências de que o setor é caracterizado por uma estrutura de mercado imperfeita [ver Nakane (2002), Belaisch (2003) e Petterini e Jorge Neto (2003)]. Porém, tais testes rejeitam a hipótese de que os bancos brasileiros operam sob a forma de cartel. Nakane (2003) apresenta uma resenha desses trabalhos.

Apesar dessa evidência, ainda não é clara qual seria a contribuição que um eventual exercício de poder de mercado pelos bancos pudesse ter sobre o *spread* bancário cobrado pelos mesmos. Exercícios de decomposição do *spread* bancário realizados pelo Banco Central do Brasil (Bacen) sugerem que 28% deste podem ser

---

\* As opiniões expressas neste documento refletem as posições dos autores e não as do Banco Central do Brasil (Bacen) ou de quaisquer de seus membros. Os autores agradecem, sem implicar, os comentários e sugestões de Eduardo Fiúza, José Gustavo Feres, João Manoel Pinho de Melo, Alberto Salvo, Paulo Furquim de Azevedo, Alkimar Moura, Juliano Assunção, Walter Novaes, Bernardo Blum, Alessandro Oliveira, Oscar Herrera, José Álvaro Rodrigues Neto, Maria Eduarda Tannuri-Pianto, bem como aos participantes de seminários no Bacen, FGV-SP, PUC-RJ, SDE/Ipea/Anpec, Nectar/ITA, Cemla e SBE. Agradecem também ao Defin/Bacen pela disponibilização dos dados utilizados no trabalho.

atribuídos ao que se denomina resíduo do banco<sup>1</sup> que, entre outros fatores, contempla a lucratividade do mesmo nas operações de empréstimo. Entretanto, dado seu caráter residual na decomposição, esse componente fornece uma medida imprecisa do poder de mercado da instituição bancária.

Assim, parece-nos que ainda é largamente desconhecida a relação existente entre poder de mercado e *spreads* bancários no Brasil. O presente estudo pretende colaborar para aumentar o grau de conhecimento a respeito dessa matéria.

Outra área em que o conhecimento empírico no país é parco diz respeito à demanda por produtos bancários. Pouco se sabe, por exemplo, sobre a magnitude de elasticidades-preço para a demanda de diferentes produtos bancários. Neste estudo, pretendemos apresentar estimativas preliminares que procuram suprir essa lacuna. O modelo de demanda apresentado e estimado neste capítulo faz uso dos modelos de escolha discreta e, em especial, da especificação *logit* multinomial.

Assume-se que, na sua dimensão geográfica, o mercado relevante para os produtos bancários seja localmente definido, no nível do município. Essa é outra diferença deste trabalho comparativamente aos trabalhos anteriores que se preocuparam em investigar a estrutura de mercado característica do setor bancário no país. Nestes, a dimensão geográfica do mercado abrangia todo o território nacional. Dada a amplidão dessa definição, os testes realizados até o momento podem estar enviesados na direção de subestimar o poder de mercado realmente existente. Ao adotarmos uma dimensão geográfica muito mais restrita, pretendemos diminuir tal viés.

Do ponto de vista da oferta, o artigo desenvolve dois modelos de comportamento da firma bancária: o primeiro condizente com uma concorrência Bertrand-Nash e o segundo com comportamento cartelizado. As margens preço-custo absolutas compatíveis com cada modelo são derivadas. Tais margens dependem apenas de variáveis observáveis e dos parâmetros de demanda sendo, portanto, passíveis de computação.

O teste de poder de mercado apresentado no estudo confronta as margens preço-custo empiricamente observadas com aquelas preditas pelos modelos de Bertrand e de cartel. Os resultados apresentados mostram que o modelo de Bertrand descreve bem o comportamento dos preços de serviços bancários, enquanto tal modelo superestima o poder de mercado observado na fixação de preços para depósitos a prazo e para empréstimos.

---

1. Dados para dezembro de 2003, última informação disponível. Ver Bacen (2004). Os demais componentes incluídos na decomposição do *spread* bem como suas importâncias relativas são: impostos diretos (12,96%), impostos indiretos mais contribuição ao Fundo Garantidor de Créditos (FGC) (8,09%), compulsório sobre depósitos (5,04%), despesas administrativas (26,37%) e despesas de inadimplência (19,98%).

Além desta introdução, o estudo apresenta o modelo utilizado para o lado da demanda (Seção 3.2), o modelo utilizado para o lado da oferta (Seção 3.3), uma descrição da base de dados (Seção 3.4), os resultados das estimativas de demanda (Seção 3.5), os resultados das simulações (Seção 3.6), considerações finais (Seção 3.7) e um apêndice com os resultados das estimativas.

### 3.2 UM MODELO DE DEMANDA POR PRODUTOS BANCÁRIOS

O modelo descrito aqui segue Dick (2002), que adapta a abordagem de escolha discreta para a demanda por serviços bancários. Dick havia restringido sua análise a serviços de depósitos. Além de depósitos, utilizamos o modelo para especificar a demanda por empréstimos.

O modelo de demanda estimado tem a seguinte estrutura formal: suponha um consumidor (indivíduo ou firma)  $i$  escolhendo os serviços de depósitos de um banco  $j$  em um certo mercado  $t$  ( $t = 1, \dots, T$ ). Seja, então,  $j = 0, 1, \dots, J_t$  o índice referente aos bancos competindo no mercado  $t$ , onde  $j = 0$  representa a opção *outside*, ou seja, a opção de não escolher qualquer banco. A utilidade do consumidor  $i$ , ao escolher o banco  $j$ , é então representada por:

$$u_{ijt} = p_{jt}^d \alpha^d - p_{jt}^s \alpha^s + x_{jt} \beta + \xi_j + \varepsilon_{ijt} \quad (1)$$

onde  $p_{jt}^d$  representa a taxa de juros paga sobre depósitos a prazo;  $p_{jt}^s$  indica a tarifa cobrada sobre os serviços bancários;  $x_{jt}$  é um vetor linha  $K$ -dimensional de características observadas do produto  $j$  no mercado  $t$ ;  $\xi_j$  representa as características não-observadas do produto  $j$ ; e  $\varepsilon_{ijt}$  é um termo estocástico i.i.d., de média zero, com uma distribuição de valor extremo de Tipo 1. O vetor  $K+2$  dimensional  $\theta_D \equiv (\alpha^d, \alpha^s, \beta)$  representa os parâmetros da demanda por serviços de depósitos.

Supõe-se que cada consumidor selecione uma unidade de serviços bancários do banco que lhe gere a maior utilidade. Define-se, então, implicitamente, o conjunto de variáveis não-observáveis que conduzem à escolha dos serviços do banco  $j$ . Formalmente, denote este conjunto por:

$$A_{jt} (x_{.t}, p_{.t}^d, p_{.t}^s, \xi_{.t}; \theta_D) = \{(v_i, \varepsilon_{it}) \text{ tal que } u_{ijt} \geq u_{ilt} \forall l = 0, 1, \dots, J\} \quad (2)$$

onde,  $x_{jt} = (x_{1j}, \dots, x_{Jt})'$ ,  $p_{jt}^d = (p_{1t}^d, \dots, p_{Jt}^d)'$ ,  $p_{jt}^s = (p_{1t}^s, \dots, p_{Jt}^s)'$ ,  $\xi_{jt} = (\xi_{1t}, \dots, \xi_{Jt})'$  e  $v_i$  representa as características não-observáveis do consumidor  $i$ .

Assim, assumindo que empates ocorrem com probabilidade zero, a parcela de mercado do produto  $j$  é dada pela probabilidade de que  $(v_i, \varepsilon_{it})$  pertença à região  $A_{jt}$ , para todo  $i$ . Formalmente, essa parcela de mercado é dada por:

$$s_{jt} = (x_{jt}, p_{jt}^d, p_{jt}^s, \xi_{jt}; \theta_D) = \int_{A_{jt}} dP^*(v, \varepsilon) = \int_{A_{jt}} dP^*(v) dP^*(\varepsilon) \quad (3)$$

onde  $P^*(.)$  denota funções de distribuição populacionais.

Note-se que este modelo não inclui nenhuma interação entre as características individuais não-observáveis  $v_i$  e as características observáveis do produto  $j$ . Ou seja, o modelo pertence à classe dos modelos tradicionais de escolha discreta conhecidos como modelos *logit* multinomiais. Para essa classe de modelos, é possível resolver (3) de maneira analítica, obtendo-se:

$$s_{jt} = \frac{\exp(p_{jt}^d \alpha^d - p_{jt}^s \alpha^s + x_{jt} \beta + \xi_j)}{1 + \sum_{k=1}^J \exp(p_{kt}^d \alpha^d - p_{kt}^s \alpha^s + x_{kt} \beta + \xi_k)} \quad (4)$$

Apesar de sua simplicidade analítica, as limitações da especificação *logit* são bem conhecidas [ver, entre outros, Berry (1994), Berry, Levinsohn e Pakes (1995) e Nevo (2000)]. Em particular, as elasticidades-preço geradas por esse modelo dependem somente das parcelas de mercado, gerando padrões de substituição bastante restritivos. Os modelos de escolha discreta com coeficientes aleatórios introduzem termos de interação entre as características individuais e as características observáveis dos produtos, gerando padrões de substituição mais razoáveis [ver Berry (1994), Berry, Levinsohn e Pakes (1995 e 2004) e Nevo (2001)]. O custo dessa opção é a maior complexidade computacional, uma vez que, nesse caso, a expressão (3) não pode mais ser resolvida analiticamente.

Seja  $\delta_{jt} \equiv p_{jt}^d \alpha^d - p_{jt}^s \alpha^s + x_{jt} \beta + \xi_j$  a utilidade média associada ao produto  $j$ , comum a todos os consumidores e seja  $\delta_t \equiv (\delta_1, \dots, \delta_J)$ . Se  $S_{jt}$  representar a parcela de mercado efetivamente observada para o banco  $j$ , então, o seguinte conjunto de equações deve ser observado para os valores verdadeiros de  $\delta$ :

$$S_{jt} = s_{jt}(\delta), \quad j = 1, \dots, J \quad (5)$$

onde o termo do lado direito de (5) é dado pela expressão (4) e representa a parcela de mercado prevista pelo modelo.

Substituindo a expressão (4) em (5) e normalizando a utilidade média do bem *outside* para zero, obtém-se a equação de demanda por serviços de depósitos a ser estimada:

$$\ln(S_{jt}) - \ln(S_{0t}) = p_{jt}^d \alpha^d - p_{jt}^s \alpha^s + x_{jt} \beta + \xi_j \quad (6)$$

Note que as características não-observáveis do produto  $j$  serão captadas pelo erro da regressão em (6). Dada a correlação entre tais componentes e os preços, técnicas de variáveis instrumentais devem ser utilizadas.

Para o lado dos empréstimos, a especificação é inteiramente análoga. Supõe-se, nesse caso, que a taxa de juros cobrada sobre empréstimos  $p_{jt}^e$  substitua os preços que aparecem em (6).

### 3.3 O LADO DA OFERTA: UM MODELO DE COMPORTAMENTO BANCÁRIO

Esta seção desenvolve o lado da oferta, apresentando um modelo de comportamento da firma bancária.

Suponha que exista um banco  $j$  que atua em  $T$  distintos mercados. Em cada um desses mercados, o banco opera com os seguintes produtos: depósitos à vista ( $DV$ ), depósitos de poupança ( $DPoup$ ), depósitos a prazo ( $DP$ ), empréstimos ( $E$ ), títulos ( $B$ ) e serviços bancários ( $S$ ).

O banco se engaja em concorrência de preços tipo Bertrand com seus concorrentes.<sup>2</sup> Três são os preços objeto de escolha pelo banco  $j$ : taxa de juros de empréstimo ( $p_j^e$ ), taxa de juros sobre depósitos a prazo ( $p_j^d$ ) e preço de serviços ( $p_j^s$ ). Assume-se que o banco não discrimina preços entre os vários mercados, estabelecendo os mesmos preços para todos eles, ou seja, para qualquer mercado  $t$  ( $t = 1, \dots, T$ ), observa-se:  $p_{jt}^e = p_j^e$ ,  $p_{jt}^d = p_j^d$  e  $p_{jt}^s = p_j^s$ .<sup>3</sup>

Além de tais preços, existem dois outros preços que são exógenos ao banco: taxa de juros dos títulos ( $r$ ) e taxa de juros dos depósitos de poupança ( $r^{POUP}$ ). Supõe-se também que os depósitos à vista não são remunerados.

2. Na discussão que segue, assume-se a existência de equilíbrios interiores de estratégia pura, com preços estritamente positivos em tais equilíbrios.

3. Essa hipótese restritiva é assumida em função, especialmente, da disponibilidade de dados para o estudo empírico. Em especial, os preços são observados apenas no nível do banco e não no nível de cada mercado.

A restrição de recursos do banco  $j$  é dada por:

$$B_j + E_j + \rho D V_j = D V_j + D P_j + D P_{oup_j} \quad (7)$$

onde  $\rho$  representa os recolhimentos e direcionamentos compulsórios sobre depósitos à vista. Supõe-se que os recolhimentos compulsórios sobre depósitos a prazo e de poupança, caso eles existam, sejam remunerados à taxa  $r$ , o que neutraliza seus efeitos sobre o banco. Supõe-se também que o banco não disponha de capital (*equity*) como um mecanismo de obter *funding*, o que não é uma hipótese restritiva.

Modelos de demanda foram definidos e estimados para os empréstimos bancários, para os depósitos a prazo e para o agregado de depósitos à vista e de poupança ( $D_j = DV_j + DP_{oup_j}$ ).<sup>4</sup> Para esses produtos, temos as seguintes identidades:

$$\begin{aligned} E_j &\equiv \sum_{t=1}^T E_{jt} \equiv V^E \sum_{t=1}^T M_t s_{jt}^E \\ D P_j &\equiv \sum_{t=1}^T D P_{jt} \equiv V^{DP} \sum_{t=1}^T M_t s_{jt}^{DP} \\ D_j &\equiv \sum_{t=1}^T D_{jt} \equiv V^D \sum_{t=1}^T M_t s_{jt}^D \end{aligned} \quad (8)$$

onde  $E_{jt}$  é o volume de empréstimos do banco  $j$  no mercado  $t$ ,  $D P_{jt}$  é o volume de depósitos a prazo do banco  $j$  no mercado  $t$ ,  $D_{jt}$  é o volume do agregado de depósitos à vista e de poupança do banco  $j$  no mercado  $t$ ,  $V^E$  é o valor médio das operações de empréstimo para todos os bancos,  $V^{DP}$  é o valor médio dos depósitos a prazo para todos os bancos,  $V^D$  é o valor médio dos depósitos à vista e de poupança para todos os bancos,  $M_t$  é o tamanho do mercado  $t$ ,  $s_{jt}^E$  é o *market share* em termos de empréstimos do banco  $j$  no mercado  $t$ ,  $s_{jt}^{DP}$  é o *market share* em termos de depósitos a prazo do banco  $j$  no mercado  $t$ , e  $s_{jt}^D$  é o *market share* em termos do agregado de depósitos à vista e de poupança do banco  $j$  no mercado  $t$ .

---

4. Como os preços de depósitos à vista e de poupança são fixados por lei e não diferem entre os bancos, acreditamos que os determinantes da demanda de cada um desses depósitos sejam similares, permitindo sua agregação.

Os modelos de demanda desenvolvidos no trabalho utilizam a especificação *logit* que, resumidamente, têm a seguinte estrutura:

$$\begin{aligned}\ln(s_{jt}^E) - \ln(s_{0t}^E) &= -\alpha_1 p_j^e + \text{outras variáveis} \\ \ln(s_{jt}^{DP}) - \ln(s_{0t}^{DP}) &= \alpha_2 p_j^d - \alpha_3 p_j^s + \text{outras variáveis} \\ \ln(s_{jt}^D) - \ln(s_{0t}^D) &= -\alpha_4 p_j^s - \alpha_5 p_j^d + \text{outras variáveis}\end{aligned}\quad (9)$$

onde  $s_{0t}^E$  é o *market share* em termos de empréstimos do bem *outside* no mercado  $t$ ,  $s_{0t}^{DP}$  é o *market share* em termos de depósitos a prazo do bem *outside* no mercado  $t$ ,  $s_{0t}^D$  é o *market share* em termos do agregado de depósitos à vista e de poupança do bem *outside* no mercado  $t$ , e  $\alpha_1$  a  $\alpha_5$  são os coeficientes (positivos) dos preços, estimados na Seção 3.5.

Para resolver o problema de otimização do banco sob concorrência de Bertrand, necessitaremos das seguintes derivadas parciais (que podem ser facilmente obtidas a partir do modelo *logit*):

$$\begin{aligned}\frac{\partial s_{jt}^E}{\partial p_j^e} &= -\alpha_1 s_{jt}^E (1 - s_{jt}^E) \\ \frac{\partial s_{jt}^{DP}}{\partial p_j^d} &= \alpha_2 s_{jt}^{DP} (1 - s_{jt}^{DP}) \\ \frac{\partial s_{jt}^{DP}}{\partial p_j^s} &= -\alpha_3 s_{jt}^{DP} (1 - s_{jt}^{DP}) \\ \frac{\partial s_{jt}^D}{\partial p_j^s} &= -\alpha_4 s_{jt}^D (1 - s_{jt}^D) \\ \frac{\partial s_{jt}^D}{\partial p_j^d} &= -\alpha_5 s_{jt}^D (1 - s_{jt}^D)\end{aligned}\quad (10)$$

Para resolver o problema de otimização do banco sob uma estrutura de cartel serão também necessárias as derivadas parciais cruzadas. Assim, para dois bancos distintos  $j$  e  $k$ , obtém-se, a partir de (9):

$$\begin{aligned}
 \frac{\partial s_{jt}^E}{\partial p_k^e} &= \alpha_1 s_{jt}^E s_{kt}^E \\
 \frac{\partial s_{jt}^{DP}}{\partial p_k^d} &= -\alpha_2 s_{jt}^{DP} s_{kt}^{DP} \\
 \frac{\partial s_{jt}^{DP}}{\partial p_k^s} &= \alpha_3 s_{jt}^{DP} s_{kt}^{DP} \\
 \frac{\partial s_{jt}^D}{\partial p_k^s} &= \alpha_4 s_{jt}^D s_{kt}^D \\
 \frac{\partial s_{jt}^D}{\partial p_k^d} &= \alpha_5 s_{jt}^D s_{kt}^D
 \end{aligned} \tag{11}$$

O lucro do banco  $j$  pode ser expresso como:

$$\prod_j = rB_j + p_j^e E_j + p_j^s S_j - p_j^d DP_j - r^{POUP} DP_{oup} j - C(E_j, DP_j, D_j) \tag{12}$$

onde  $C(.)$  representa os custos operacionais do banco. Substituindo a restrição de recursos (7) em (12), obtém-se a seguinte expressão alternativa para o lucro do banco:

$$\prod_j = (p_j^e - r) E_j + (r - p_j^d) DP_j + (r + p_j^s - \theta_j) D_j - C(E_j, DP_j, D_j) \tag{13}$$

onde utilizou-se a hipótese de que o fluxo de serviços  $S_j$  é proporcional ao volume de depósitos  $D_j$  (e, sem perda de generalidade, normaliza-se o coeficiente de proporcionalidade para 1) e onde:

$$\theta_j = \frac{r\rho DV_j + r^{POUP} DP_{oup} j}{D_j}$$

No cômputo das condições de primeira ordem que se seguem, as derivadas parciais de  $\theta_j$  com relação aos preços serão ignoradas.

### 3.3.1 Concorrência de Bertrand

O banco  $j$  determina os preços  $p_j^e$ ,  $p_j^d$  e  $p_j^s$  de forma a maximizar os lucros dados por (13). Na primeira estrutura de mercado a ser considerada (concorrência de Bertrand), esse banco considera que os preços de seus rivais mantenham-se inalterados. A condição de primeira ordem para  $p_j^e$  é dada por:

$$E_j + \left( p_j^e - r - c_j^E \right) V^E \sum_{t=1}^T M_t \frac{\partial s_{jt}^E}{\partial p_j^e} = 0 \quad (14)$$

onde  $c_j^E \equiv \frac{\partial C}{\partial E_j}$  é o custo marginal de prover uma unidade de empréstimo. Subs-

tituindo (10) em (14) e rearranjando os termos obtém-se:

$$p_j^e - r - c_j^E = \frac{E_j}{\alpha_1 V^E \sum_{t=1}^T M_t s_{jt}^E (1 - s_{jt}^E)} \quad (15)$$

Note que, uma vez estimado o coeficiente de demanda  $\alpha_1$ , o lado direito de (15) é observável. Assim, essa expressão pode ser utilizada para derivar a margem preço-custo absoluta do banco  $j$  em suas atividades de empréstimo que seja compatível com o modelo aqui delineado. Adicionalmente, se alguma medida de  $c_j^E$  estiver disponível, pode-se observar também o lado esquerdo de (15) e, portanto, essa expressão pode ser utilizada para ver quão compatível com os dados é o modelo aqui desenvolvido.

A condição de primeira ordem para  $p_j^d$  é dada por [já fazendo uso da expressão (10)]:

$$\begin{aligned} & \left( r - p_j^d - c_j^{DP} \right) \alpha_2 V^{DP} \sum_{t=1}^T M_t s_{jt}^{DP} (1 - s_{jt}^{DP}) - \\ & - \left( r + p_j^s - \theta_j - c_j^D \right) \alpha_5 V^D \sum_{t=1}^T M_t s_{jt}^D (1 - s_{jt}^D) = DP_j \end{aligned} \quad (16)$$

onde  $c_j^{DP} \equiv \frac{\partial C}{\partial DP_j}$  é o custo marginal de prover uma unidade de depósito a prazo e  $c_j^D \equiv \frac{\partial C}{\partial D_j}$  é o custo marginal de prover uma unidade de depósito à vista ou de poupança.

De maneira similar, a condição de primeira ordem para  $p_j^s$  é dada por:

$$\begin{aligned} & \left( r - p_j^d - c_j^{DP} \right) \alpha_3 V^{DP} \sum_{t=1}^T M_t s_{jt}^{DP} \left( 1 - s_{jt}^{DP} \right) + \\ & + \left( r + p_j^s - \theta_j - c_j^D \right) \alpha_4 V^D \sum_{t=1}^T M_t s_{jt}^D \left( 1 - s_{jt}^D \right) = D_j \end{aligned} \quad (17)$$

Pode-se resolver (16) e (17) para encontrar as margens preço-custo absolutas para os dois preços, obtendo-se:

$$r - p_j^d - c_j^{DP} = \frac{\alpha_4 DP_j + \alpha_5 D_j}{(\alpha_2 \alpha_4 + \alpha_3 \alpha_5) V^{DP} \sum_{t=1}^T M_t s_{jt}^{DP} \left( 1 - s_{jt}^{DP} \right)} \quad (18)$$

$$r - p_j^s - \theta_j - c_j^D = \frac{\alpha_2 D_j - \alpha_3 DP_j}{(\alpha_2 \alpha_4 + \alpha_3 \alpha_5) V^D \sum_{t=1}^T M_t s_{jt}^D \left( 1 - s_{jt}^D \right)} \quad (19)$$

Deve-se notar que, assim como para a expressão (15), os lados direitos de (18) e (19) são observáveis, uma vez que os coeficientes de demanda sejam obtidos. Assim, pode-se utilizar tais expressões ou para derivar as margens preço-custo absolutas preditas pelo modelo ou, caso os custos marginais  $c_j^{DP}$  e  $c_j^D$  estejam disponíveis, para checar a plausibilidade do modelo proposto contra os dados.

De (15) e (18) obtém-se o *spread* bancário consistente com o modelo, onde o *spread* é definido como a diferença entre a taxa de juros de empréstimos e a taxa de juros de depósitos a prazo:

$$\begin{aligned} p_j^e - p_j^d &= c_j^L + c_j^{DP} + \frac{E_j}{\alpha_1 V^E \sum_{t=1}^T M_t s_{jt}^E (1 - s_{jt}^E)} + \\ &+ \frac{\alpha_4 D P_j + \alpha_5 D_j}{(\alpha_2 \alpha_4 + \alpha_3 \alpha_5) V^{DP} \sum_{t=1}^T M_t s_{jt}^{DP} (1 - s_{jt}^{DP})} \end{aligned} \quad (20)$$

### 3.3.2 Solução de cartel

O teste de poder de mercado proposto no trabalho compara as alocações sob Bertrand e sob cartel. Para resolver o problema de otimização sob esta última estrutura de mercado, supõe-se que o monopolista encontre os preços de forma a maximizar a soma dos lucros do conjunto dos bancos onde o lucro de cada banco é dado por uma expressão como (13). Assim, a condição de primeira ordem para a taxa de juros de empréstimos de um banco  $j$ ,  $p_j^e$ , é dada por:

$$\begin{aligned} E_j + (p_j^e - r - c_j^E) V^E \sum_{t=1}^T M_t \frac{\partial s_{jt}^E}{\partial p_j^e} + \\ + \sum_{k \neq j}^J \left[ (p_k^e - r - c_k^E) V^E \sum_{t=1}^T M_t \frac{\partial s_{kt}^E}{\partial p_j^e} \right] = 0 \end{aligned} \quad (21)$$

Cabe observar que os dois primeiros termos do lado esquerdo de (21) são idênticos ao do oligopolista de Bertrand. O terceiro termo do lado esquerdo leva em conta o impacto sobre a demanda por empréstimos dos demais bancos quando o banco  $j$  altera seu preço.

Existem  $J$  condições de primeira ordem equivalentes a (21) que devem ser simultaneamente resolvidas para se encontrar as margens preço-custo absolutas

preditas pelo modelo de cartel. Substituindo-se as derivadas parciais (10)-(11) em (21) pode-se encontrar tais margens analiticamente. Para tanto, definam-se as seguintes matrizes:

$$\left( p^e - r - c^E \right) \equiv \begin{bmatrix} p_1^e - r - c_1^E \\ \vdots \\ p_J^e - r - c_J^E \end{bmatrix}, E \equiv \begin{bmatrix} E_1 \\ \vdots \\ E_J \end{bmatrix} \text{ e}$$

$$\Delta_1 \equiv \begin{bmatrix} \sum_{t=1}^T M_t s_{1t}^E (1 - s_{1t}^E) & -\sum_{t=1}^T M_t s_{1t}^E s_{2t}^E & \dots & -\sum_{t=1}^T M_t s_{1t}^E s_{Jt}^E \\ -\sum_{t=1}^T M_t s_{2t}^E s_{1t}^E & \sum_{t=1}^T M_t s_{2t}^E (1 - s_{2t}^E) & \dots & -\sum_{t=1}^T M_t s_{2t}^E s_{Jt}^E \\ \vdots & \vdots & \dots & \vdots \\ -\sum_{t=1}^T M_t s_{Jt}^E s_{1t}^E & -\sum_{t=1}^T M_t s_{Jt}^E s_{2t}^E & \dots & \sum_{t=1}^T M_t s_{Jt}^E (1 - s_{Jt}^E) \end{bmatrix}$$

As margens preço-custo absolutas consistentes com o modelo de cartel são então dadas por:

$$\left( p^e - r - c^E \right) = \left( \frac{1}{\alpha_1 V^E} \right) \Delta_1^{-1} E \quad (22)$$

O procedimento é similar para os outros dois produtos. Assim, para depósitos a prazo, as condições de primeira ordem para os  $J$  bancos na solução de cartel podem ser expressas por:

$$(\alpha_2 V^{DP}) \Delta_2 (r - p^d - c^{DP}) - (\alpha_5 V^D) \Delta_3 (r + p^s - \Theta - c^D) = DP \quad (23)$$

onde:

$$\begin{aligned} (r - p^d - c^{DP}) &\equiv \begin{bmatrix} r - p_1^d - c_1^{DP} \\ \vdots \\ r - p_J^d - c_J^{DP} \end{bmatrix}, (r + p^s - \Theta - c^D) \equiv \\ &\equiv \begin{bmatrix} r + p_1^s - \Theta_1 - c_1^D \\ \vdots \\ r + p_J^s - \Theta_J - c_J^D \end{bmatrix}, DP \equiv \begin{bmatrix} DP_1 \\ \vdots \\ DP_J \end{bmatrix} \end{aligned}$$

$$\Delta_2 \equiv \begin{bmatrix} \sum_{t=1}^T M_t s_{1t}^{DP} (1 - s_{1t}^{DP}) & -\sum_{t=1}^T M_t s_{1t}^{DP} s_{2t}^{DP} & \dots & -\sum_{t=1}^T M_t s_{1t}^{DP} s_{Jt}^{DP} \\ -\sum_{t=1}^T M_t s_{2t}^{DP} s_{1t}^{DP} & \sum_{t=1}^T M_t s_{2t}^{DP} (1 - s_{2t}^{DP}) & \dots & -\sum_{t=1}^T M_t s_{2t}^{DP} s_{Jt}^{DP} \\ \vdots & \vdots & \dots & \vdots \\ -\sum_{t=1}^T M_t s_{Jt}^{DP} s_{1t}^{DP} & -\sum_{t=1}^T M_t s_{Jt}^{DP} s_{2t}^{DP} & \dots & \sum_{t=1}^T M_t s_{Jt}^{DP} (1 - s_{Jt}^{DP}) \end{bmatrix}$$

e:

$$\Delta_3 \equiv \begin{bmatrix} \sum_{t=1}^T M_t s_{1t}^D (1 - s_{1t}^D) & -\sum_{t=1}^T M_t s_{1t}^D s_{2t}^D & \dots & -\sum_{t=1}^T M_t s_{1t}^D s_{Jt}^D \\ -\sum_{t=1}^T M_t s_{2t}^D s_{1t}^D & \sum_{t=1}^T M_t s_{2t}^D (1 - s_{2t}^D) & \dots & -\sum_{t=1}^T M_t s_{2t}^D s_{Jt}^D \\ \vdots & \vdots & \dots & \vdots \\ -\sum_{t=1}^T M_t s_{Jt}^D s_{1t}^D & -\sum_{t=1}^T M_t s_{Jt}^D s_{2t}^D & \dots & \sum_{t=1}^T M_t s_{Jt}^D (1 - s_{Jt}^D) \end{bmatrix}$$

Similarmente, as condições de primeira ordem para os  $J$  bancos na solução de cartel para os preços de serviços podem ser escritas como:

$$(\alpha_3 V^{D_p}) \Delta_2 (r - p^d - c^{DP}) + (\alpha_4 V^D) \Delta_3 (r + p^s - \Theta - c^D) = D \quad (24)$$

onde:

$$D \equiv \begin{bmatrix} D_1 \\ \vdots \\ D_J \end{bmatrix}$$

Resolvendo (23) e (24), simultaneamente, é possível encontrar as margens preço-custo absolutas para depósitos a prazo e para serviços, obtendo-se:

$$(r - p^d - c^{DP}) = \left( \frac{1}{(\alpha_2 \alpha_4 + \alpha_3 \alpha_5) V^{DP}} \right) [\alpha_4 \Delta_2^{-1} DP + \alpha_5 \Delta_2^{-1} D] \quad (25)$$

$$(r - p^s - \Theta - c^D) = \left( \frac{1}{(\alpha_2 \alpha_4 + \alpha_3 \alpha_5) V^D} \right) [\alpha_2 \Delta_3^{-1} D + \alpha_3 \Delta_3^{-1} DP] \quad (26)$$

Finalmente, o *spread* bancário consistente com a solução de cartel vem da comparação de (22) e (25):

$$\begin{aligned} (p^e - p^d) &= c^E + c^{DP} + \left( \frac{1}{\alpha_1 V^E} \right) \Delta_1^{-1} E + \\ &+ \left( \frac{1}{(\alpha_2 \alpha_4 + \alpha_3 \alpha_5) V^{DP}} \right) [\alpha_4 \Delta_2^{-1} DP + \alpha_5 \Delta_2^{-1} D] \end{aligned} \quad (27)$$

### 3.4 DADOS

O modelo de demanda por serviços bancários é estimado para um painel de bancos brasileiros observados em dezembro de 2002 e dezembro de 2003.<sup>5</sup> A amostra contempla todos os bancos comerciais e os bancos múltiplos com carteira comercial operando no país. No total, 134 bancos são incluídos na análise.

#### 3.4.1 Definição do mercado

Para a definição do que seja um mercado, consideramos tanto a dimensão produto quanto a dimensão geográfica. Do ponto de vista da dimensão produto, apresentamos estimativas separadas para três classes de produtos: soma de depósitos à vista e de poupança, depósitos a prazo e empréstimos. Com relação à dimensão geográfica, consideramos um mercado como sendo correspondente a um município, de acordo com a definição utilizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). No total, 3.252 municípios são incluídos, correspondendo aos municípios nos quais havia, em uma das datas-base do estudo, pelo menos um banco operando. Com relação a cada um dos produtos considerados no estudo, temos 3.242 mercados para depósitos à vista e de poupança, 2.781 para depósitos a prazo e 3.242 para empréstimos.

#### 3.4.2 Parcelas de mercado

A estimação da equação (6) requer a definição da parcela de mercado para cada banco, bem como a do bem *outside*. Nesse caso, seguimos o procedimento utilizado por Dick (2002), assumindo que a escolha discreta do consumidor refere-se a unidades de “contas bancárias médias”, calculadas como a razão entre o volume de cada tipo de depósito para todos os bancos do sistema financeiro nacional em um dado ano sobre o número de contas bancárias desse ano. A divisão do volume de cada tipo de depósito de um dado banco em um dado mercado pelas “contas bancárias médias” dá o número destas “produzidas” por esse banco nesse mercado. Sua parcela de mercado será a razão entre o número de “contas bancárias médias” produzidas e o tamanho do mercado potencial. O tamanho potencial do mercado é definido em termos da população local do mercado, como em Berry, Levinsohn

5. Neste trabalho não exploramos estimadores de dados de painel. As duas *cross-sections* disponíveis são tratadas como observações independentes. Como os preços utilizados nas estimativas apresentam variação somente entre os bancos, mas não entre os distintos mercados para cada banco, a identificação dos coeficientes de preços seria extremamente difícil em modelos de painel que procurassem controlar pelo efeito de características não-observáveis para cada banco. De fato, quando versões dos modelos estimados no trabalho foram examinadas utilizando o estimador de efeitos fixos para dados em painel, boa parte dos coeficientes dos preços tornou-se não-significativa (resultados não reportados). A hipótese adotada no trabalho é que a introdução de características do banco que não variam entre os mercados tanto como variáveis explicativas como no conjunto de variáveis instrumentais ajuda a controlar por boa parte dos efeitos individuais não-observáveis. Para uma aplicação em que efeitos fixos são introduzidos no conjunto de características observáveis dos produtos, ver Nevo (2001).

e Pakes (1995) e em Nevo (2001). A parcela de mercado do bem *outside* é encontrada como sendo a diferença entre 1 e a soma das parcelas de mercado dos bancos participantes.<sup>6</sup> O mesmo procedimento foi adotado para se definir os “empréstimos bancários médios” produzidos por cada banco em cada mercado, bem como a parcela de mercado do bem *outside*, no caso de empréstimos bancários.

O sistema de estatísticas bancárias (Estban) do Bacen é a fonte de informações para os volumes de depósitos e de empréstimos para cada banco em cada mercado. Esse sistema de informações traz as principais contas do balancete dos bancos para cada agência em cada município. Agregamos as informações por agência obtendo os volumes de depósitos de cada banco em cada município. A agregação para todos os bancos e todos os municípios fornece os valores para todo o sistema financeiro nacional. Os depósitos à vista correspondem à soma dos depósitos à vista de pessoas físicas e de pessoas jurídicas (conta 410); os depósitos de poupança correspondem à conta 420, os depósitos a prazo correspondem à conta 432 e os empréstimos bancários correspondem à soma dos empréstimos e títulos descontados (conta 161) e financiamentos (conta 162).

O número de contas para cada tipo de depósito foi obtido através dos relatórios semestrais do Fundo Garantidor de Créditos (FGC), produzidos pelo Bacen. Esse relatório consolida os clientes de cada instituição financeira, o que dá uma idéia melhor da decisão de demanda de um agente. Os valores encontrados para as contas bancárias médias foram de R\$ 1.966,47 e R\$ 1.819,24 para o agregado de depósitos à vista e de poupança em dezembro de 2002 e dezembro de 2003, respectivamente; e de R\$ 37.255,17 e R\$ 54.047,71 para depósitos a prazo em dezembro de 2002 e dezembro de 2003, respectivamente. Deve-se também mencionar que para depósitos a prazo utilizou-se o volume total de depósitos conforme relatado pelo relatório do FGC e não a agregação das informações do Estban.<sup>7</sup>

Para empréstimos bancários, seu valor médio foi obtido através de informações do Sistema de Informações de Crédito (SCR) do Bacen para dezembro de 2003. Foram consideradas as operações de empréstimo iniciadas nesse mês de valor unitário superior a R\$ 5 mil e inferior a R\$ 10 milhões. O limite inferior decorre da inexistência de informações individualizadas para tais operações. O limite superior é arbitrário. O valor médio para os empréstimos bancários obtido pela aplicação

6. Esse procedimento gera, para alguns mercados, participações negativas para o bem *outside*. Nesse caso, optamos por excluir tais mercados da análise. O número de mercados excluídos por esse motivo equivale a 0 para depósitos a prazo, 92 (em 2002) e 127 (em 2003) para a soma de depósitos à vista e de poupança e 2 (tanto em 2002 como em 2003) para empréstimos.

7. Existe uma diferença considerável para o volume total de depósitos a prazo informado em cada relatório. A origem dessa discrepância é desconhecida para os autores. Para dezembro de 2002, o volume total de depósitos a prazo é de R\$ 146 bilhões de acordo com o FGC e de R\$ 186 bilhões de acordo com o Estban; para dezembro de 2003, os volumes respectivos são R\$ 170 bilhões e R\$ 202 bilhões.

desses critérios foi de R\$ 21.362,16. Para dezembro de 2002, simplesmente deflacionou-se esse valor pela variação do Índice de Preços ao Consumidor Amplo (IPCA) no período (7,6%), obtendo-se, dessa forma, o valor do empréstimo médio de R\$ 19.853,31.

No que diz respeito à população de cada município, utilizou-se a série “População Residente Total” para o ano 2000, disponibilizado no sistema Sidra do IBGE.

### 3.4.3 Preços

Com relação aos preços  $p_{jt}^d$  e  $p_{jt}^s$  que aparecem em (6) (e também para  $p_{jt}^e$ ), eles são observados apenas no nível do banco e não no nível de cada mercado, isto é, supõe-se que  $p_{jt}^d = p_j^d$ ,  $p_{jt}^s = p_j^s$  e  $p_{jt}^e = p_j^e$  para todo  $t = 1, \dots, T$ .

O preço de depósitos a prazo foi obtido através das taxas de juros pagas a certificados de depósitos bancários em captações prefixadas junto a pessoas físicas e jurídicas. As taxas diárias reportadas ao Bacen por instituição financeira foram ponderadas (de acordo com o volume diário captado), capitalizadas e normalizadas para um mês de 21 dias úteis.

O preço de operações de empréstimo foi obtido junto ao Bacen, a partir das informações primárias utilizadas para gerar as “Notas para Imprensa de Juros e Spread Bancário”. Os dados diárias para operações prefixadas para pessoas físicas e jurídicas foram ponderados (de acordo com o volume diário emprestado), capitalizados e normalizados para um mês de 21 dias úteis.

Para o preço de serviços bancários, utilizou-se como base de informações o sistema Cosif do Bacen, que consiste no Plano Contábil das Instituições Financeiras, através do qual essas instituições devem reportar, mensalmente, seus planos contábeis (balancetes e demonstrativos de resultados). Utilizaram-se as contas referentes aos meses de dezembro de 2002 e dezembro de 2003. O preço dos serviços bancários é definido como a razão entre a renda de prestação de serviços (conta 7.1.7.00.00-9) e o total de depósitos (conta 4.1.0.00.00-7) para cada banco. Esse preço está em base semestral. Dividiu-se, assim, o preço obtido por seis, para se obter o equivalente em termos mensais.

Para o cálculo do termo  $\theta_j$  que aparece em (13), necessitamos também definir o preço pago aos depósitos de poupança. Este foi calculado como a razão entre as despesas de depósitos de poupança (conta 8.1.1.10.00-5, do Cosif) e o volume de depósitos de poupança (conta 4.1.2.00.00-3). Note-se que, assim como os demais preços, o preço de depósitos de poupança também difere entre bancos, porém não difere entre os mercados para cada banco.

### **3.4.4 Características observáveis**

As características observáveis do banco incluem três tipos de variáveis: variáveis observáveis no nível de cada mercado (isto é, no nível do banco no município), variáveis observáveis em nível nacional (isto é, no nível do banco no país) e termos de interação entre características observáveis de cada banco e variáveis econômicas locais (renda e população do município).

As características observáveis de um banco em um mercado incluem: *a)* número de agências do banco no município; *b)* número de postos de atendimento bancário do banco no município; *c)* número de postos de atendimento eletrônico do banco no município; *d)* densidade de agências do banco no município, sendo a densidade definida como a razão entre o número de agências e a área do município; *e)* densidade de postos de atendimento bancário do banco no município; e *f)* densidade de postos de atendimento eletrônico do banco no município.

As características observáveis do banco que não têm variação entre os mercados incluem: *a)* o número de estados em que o banco atua; *b)* a idade do banco; *c)* número de funcionários do banco; *d)* número de funcionários por agência do banco; e *e)* despesas de publicidade do banco.

As informações sobre números de agências, de postos de atendimento bancário, de postos de atendimento eletrônico, de funcionários, idade do banco e despesas de publicidade provêm do Bacen. A renda do município corresponde ao Produto Interno Bruto Municipal a valores constantes de 2000, disponibilizada no sistema Sidra do IBGE. A área de cada município corresponde à área geográfica publicada no Censo de 2000 e também disponibilizada no sistema Sidra do IBGE.

Além de características observáveis de cada banco, também foram incluídas, como variáveis de controle, algumas características do município. Tais características incluem a renda do município, a renda por habitante do município, a área e a densidade populacional do município.

### **3.4.5 Instrumentos**

No que diz respeito aos instrumentos utilizados, incluíram-se variáveis que deslocam os custos dos bancos, tais como custos de mão-de-obra, gastos do banco com ativos fixos, despesas de provisionamento, indicadores de balanço relativos a liquidez, empréstimos e patrimônio líquido. Em especial, utilizamos as seguintes variáveis como instrumentos: *a)* despesas de mão-de-obra; *b)* despesas administrativas; *c)* risco de inadimplência medido como a razão entre despesas de provisionamento líquidas de reversões sobre o volume de operações de crédito do banco; *d)* índice de liquidez do banco medido como a razão entre os ativos líquidos

do banco e seus ativos operacionais;<sup>8</sup> e) razão entre o patrimônio líquido ajustado do banco e seus ativos operacionais; e f) razão entre empréstimos totais do banco e seus ativos operacionais (instrumento utilizado somente para as equações de demanda por depósito).

Todas as informações para a construção das variáveis instrumentais mencionadas anteriormente foram obtidas junto ao plano de contas do *Cosif* do Bacen.

Além das variáveis instrumentais já mencionadas, utilizaram-se também os chamados instrumentos BLP, ou seja, aqueles sugeridos por Berry, Levinsohn e Pakes (1995), que consistem nas características dos outros bancos no mercado. Assim, a lista dos instrumentos BLP utilizada no trabalho inclui as somas das seguintes variáveis dos bancos concorrentes: a) número de agências no município; b) número de agências no Brasil; c) número de postos de atendimento bancário no município; d) número de postos de atendimento bancário no Brasil; e) número de postos de atendimento eletrônico no município; f) número de postos de atendimento eletrônico no Brasil; g) densidade de agências no município; h) densidade de postos de atendimento bancário no município; i) densidade de postos de atendimento eletrônico no município; j) idade; e k) despesas de publicidade.

### **3.5 RESULTADOS DAS ESTIMATIVAS DE DEMANDA**

Nesta seção, apresentam-se as estimativas para a demanda por produtos bancários. Os resultados são apresentados separadamente para cada um dos três produtos considerados neste trabalho: depósitos a prazo, soma de depósitos à vista e de poupança e empréstimos. Para cada um desses produtos, quatro modelos foram estimados combinando-se a utilização de diferentes conjuntos de instrumentos (deslocadores de custos ou tais deslocadores mais instrumentos BLP) e a inclusão (ou não) de interações entre as características do banco e a renda por habitante do município.

Todas as regressões estimadas incluem uma *dummy* de ano (igual a 1 para 2002), para controlar por fatores macroeconômicos. Desvios-padrão robustos à presença de heterocedasticidade são reportados. O teste de sobreidentificação apresentado é a versão do teste robusta à presença de heterocedasticidade, conforme apresentado em Wooldridge (2002, p. 123).<sup>9</sup>

8. Os ativos líquidos do banco correspondem à soma das aplicações *interfinanceiras de liquidez* (conta 1.2.0.00.00-5 do *Cosif*) com títulos públicos livres (conta 1.3.1.00.00-7), subtraída das revendas a liquidar, posição financiada (conta 1.2.1.20.00-2). Ativos operacionais correspondem aos ativos circulante e realizável a longo prazo. Excluíram-se, portanto, os ativos permanentes dessa definição.

9. Entre parênteses, apresenta-se o nível descritivo (*p-value*) do teste.

A Tabela 3.1 mostra os coeficientes dos preços para a demanda por depósitos a prazo. Abaixo do valor de cada coeficiente estimado encontra-se o desvio-padrão estimado (esquerda) e o valor da estatística-t (direita).

Observa-se, conforme esperado, que o preço dos depósitos a prazo tem um impacto positivo sobre sua própria demanda, enquanto o preço dos serviços bancários tem um impacto negativo sobre a demanda por depósitos a prazo. Isso ocorre em todas as especificações, sendo os coeficientes bastante significativos. Além disso, o valor estimado para os coeficientes não difere muito entre as diferentes especificações, mostrando-se razoavelmente robusto.

A Tabela 3.2 traz os resultados para a demanda por depósitos à vista mais poupança.

Os coeficientes estimados para os preços são, conforme esperado, negativos em todas as especificações. Eles também são altamente significativos. Contudo,

TABELA 3.1  
RESULTADOS PARA A DEMANDA POR DEPÓSITOS A PRAZO

Variável	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Preço de depósito a prazo	12,9802 0,598 (21,70)	13,00453 0,540 (24,07)	12,93793 0,594 (21,80)	13,0184 0,538 (24,21)
Preço de serviços	-2,21399 0,311 (-7,11)	-2,788356 0,306 (-9,10)	-2,252178 0,312 (-7,22)	-2,793397 0,307 (-9,09)
Interações com renda <i>per capita</i> ?	Não	Sim	Não	Sim
Instrumentos	Custo	Custo	Custo + BLP	Custo + BLP
Teste de sobreidentificação	406,537 (0,0)	398,737 (0,0)	469,723 (0,0)	593,706 (0,0)
$R^2$ ajustado 1º. estágio (preço de depósito a prazo)	0,8743	0,8799	0,8753	0,8808
$R^2$ ajustado 1º. estágio (preço de serviços)	0,4649	0,4779	0,4662	0,4781
Teste F 1º. estágio (preço de depósito a prazo)	5087,65	3683,12	4177,07	3206,52
Teste F 1º. estágio (preço de serviços)	636,23	461,09	518,55	398,51
Observações	16.086	16.084	16.059	16.057

TABELA 3.2  
RESULTADOS PARA A DEMANDA POR DEPÓSITOS À VISTA MAIS POUPANÇA

Variável	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Preço de depósito a prazo	-1,441257	-1,447618	-1,227275	-1,125488
	0,201 (-7,19)	0,203 (-7,14)	0,197 (-6,22)	0,197 (-5,70)
Preço de serviços	-0,6055965	-0,8701224	-0,609552	-0,8721783
	0,105 (-5,78)	0,109 (-7,99)	0,105 (-5,81)	0,109 (-8,03)
Interações com renda <i>per capita</i> ?	Não	Sim	Não	Sim
Instrumentos	Custo	Custo	Custo + BLP	Custo + BLP
Teste de sobreidentificação	1580,20 (0,0)	1330,39 (0,0)	1738,03 (0,0)	1577,85 (0,0)
$R^2$ ajustado 1º. estágio (preço de depósito a prazo)	0,7907	0,7953	0,7920	0,7968
$R^2$ ajustado 1º. estágio (preço de serviços)	0,3982	0,4016	0,4005	0,4042
Teste F 1º. estágio (preço de depósito a prazo)	3020,33	2174,08	2454,29	1876,77
Teste F 1º. estágio (preço de serviços)	529,74	376,40	431,28	325,53
Observações	16.785	16.783	16.748	16.746

nota-se uma ligeira diferença entre os modelos. O coeficiente para o preço de depósito a prazo mostrou-se superior (em termos absolutos) nas especificações em que somente deslocadores de custos são utilizados como instrumentos (modelos 1 e 2). Para os coeficientes para o preço de serviços bancários, os maiores valores (absolutos) foram encontrados para as especificações que introduzem termos de interação entre a renda *per capita* do município e as características dos bancos (modelos 2 e 4).

A Tabela 3.3 traz os resultados para a demanda por empréstimos. Nesse caso, supõe-se que apenas o preço de empréstimos bancários afete a demanda pelos mesmos.

Em todas as especificações, o coeficiente para o preço é negativo e altamente significativo. Não obstante, existem pequenas variações entre os modelos. O valor (absoluto) do coeficiente é maior para os modelos que incluem somente

TABELA 3.3  
RESULTADOS PARA A DEMANDA POR EMPRÉSTIMOS

Variável	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Preço de empréstimos	-0,3170979 0,021 (-15,28)	-0,3968123 0,018 (-21,64)	-0,2473996 0,019 (-12,77)	-0,3257321 0,018 (-18,55)
Interações com renda <i>per capita</i> ?	Não	Sim	Não	Sim
Instrumentos	Custo	Custo	Custo + BLP	Custo + BLP
Teste de sobreidentificação	113,39 (0,0)	191,63 (0,0)	576,70 (0,0)	600,10 (0,0)
$R^2$ ajustado 1º. estágio (preço de empréstimos)	0,4612	0,4909	0,4688	0,4978
Teste F 1º. estágio (preço de empréstimo)	786,31	600,29	653,63	530,34
Observações	19.267	19.265	19.228	19.226

deslocadores de custo como instrumentos (modelo 1 contra modelo 3 e modelo 2 contra modelo 4), assim como para os modelos que incluem interações entre as características do banco e a renda *per capita* municipal (modelo 2 contra modelo 1 e modelo 4 contra modelo 3).

É importante mencionar que todos os modelos apresentados rejeitam a hipótese nula do teste de sobreidentificação. Não é claro, contudo, se esse resultado é devido à não-validade dos instrumentos utilizados ou se é devido ao grande número de observações em cada regressão. Isso porque o teste utilizado é proporcional ao número de observações. Como contrapartida aos resultados do teste de sobreidentificação, a significância estatística do teste F no primeiro estágio em todas as equações sugere que os instrumentos utilizados têm algum poder explicativo.

A partir dos coeficientes estimados para os preços, é possível calcular as elasticidades-preço para a demanda de cada banco na amostra.<sup>10</sup> As Tabelas 3.4 e 3.5 mostram os resultados para 2002 e 2003, respectivamente. Os resultados são apresentados tanto para o banco mediano da amostra quanto para a média ponderada (pela participação de mercado de cada banco) das elasticidades.

10. É possível, ainda, computar as elasticidades-preço cruzadas para cada par de bancos na amostra. Tais elasticidades não foram, contudo, calculadas neste estudo.

TABELA 3.4  
ELASTICIDADES-PREÇO DA DEMANDA — 2002

Produto	Preço de depósitos a prazo		Preço de serviços		Preço de empréstimos	
	Mediana	Média	Mediana	Média	Mediana	Média
<b>Depósitos a prazo</b>						
Modelo 1	23,5196	21,8817	0,5321	0,7833		
Modelo 2	23,5637	21,9227	0,6701	0,9865		
Modelo 3	23,4430	21,8105	0,5413	0,7968		
Modelo 4	23,5888	21,0461	0,6713	0,9882		
<b>Depósitos à vista + poupança</b>						
Modelo 1	2,1934	2,0096	0,1914	0,2444		
Modelo 2	2,2030	2,0184	0,2750	0,3512		
Modelo 3	1,8677	1,7112	0,1926	0,2460		
Modelo 4	1,7128	1,5693	0,2756	0,3520		
<b>Empréstimos</b>						
Modelo 1					1,2209	1,4314
Modelo 2					1,5278	1,7912
Modelo 3					0,9525	1,1168
Modelo 4					1,2541	1,4704

Os resultados mostram que as elasticidades-preço próprias para a demanda de depósitos a prazo são bastante elevadas, situando-se no patamar de 21,0 a 23,6 em 2002 e de 15,5 a 16,0 em 2003. Os valores encontrados para as elasticidades-preço da demanda por empréstimos e também para a elasticidade da demanda por depósitos à vista e de poupança com relação a variações na taxa de juros de depósitos a prazo são bem mais modestos. Finalmente, a demanda por ambos os tipos de depósitos considerados no trabalho foram inelásticas com relação a variações nos preços dos serviços bancários.<sup>11</sup>

Os valores encontrados para as elasticidades são razoáveis? Devido à ausência de estudos empíricos sobre demanda por produtos bancários, é difícil avaliar essa

11. Tais demandas inelásticas não implicam violação das condições de segunda ordem para máximo. Como fica claro a partir do exame das condições de primeira ordem (16) e (17), tanto a demanda por depósitos a prazo quanto a demanda pelo agregado de depósitos à vista e de poupança devem ser consideradas para a determinação dos preços de serviços bancários e da taxa de juros paga sobre depósitos a prazo. A partir das condições de primeira ordem (14), (16) e (17), obtivemos as condições de segunda ordem e verificamos para cada observação de nossa amostra sua validade. O resultado (não reportado no trabalho) é que essas condições são válidas para a totalidade das observações amostrais.

TABELA 3.5  
ELASTICIDADES-PREÇO DA DEMANDA — 2003

Produto	Preço de depósitos a prazo		Preço de serviços		Preço de empréstimos	
	Mediana	Média	Mediana	Média	Mediana	Média
<b>Depósitos a prazo</b>						
Modelo 1	15.8692	15.5603	0.6253	0.9073		
Modelo 2	15.8989	15.5895	0.7875	1.1427		
Modelo 3	15.8175	15.5096	0.6361	0.9230		
Modelo 4	15.9159	15.6061	0.7889	1.1448		
<b>Depósitos à vista + poupança</b>						
Modelo 1	1.6282	1.4520	0.2313	0.2547		
Modelo 2	1.6354	1.4584	0.3324	0.3659		
Modelo 3	1.3864	1.2365	0.2328	0.2563		
Modelo 4	1.2715	1.1339	0.3331	0.3668		
<b>Empréstimos</b>						
Modelo 1					1.1225	1.2848
Modelo 2					1.4047	1.6078
Modelo 3					0.8758	1.0024
Modelo 4					1.1530	1.3198

questão. Os resultados que melhor se compararam aos aqui apresentados são os de Dick (2002). Essa autora estima a demanda por depósitos nos Estados Unidos utilizando também a especificação *logit* multinomial. Ela encontra valores medianos para as elasticidades-preço de 0,644 para tarifas de serviços e de 5,932 para a taxa paga sobre depósitos. Como a autora não estima separadamente a demanda por diferentes classes de depósitos, é de se imaginar que as elasticidades encontradas por ela devam se situar em valores intermediários aos aqui reportados para cada tipo de depósito. É exatamente isso que observamos, o que permite concluir que os valores estimados para as elasticidades e reportados nas Tabelas 3.4 e 3.5 não estão em conflito com os valores encontrados na literatura. Em particular, ressalte-se que Dick (2002) também encontrou demandas inelásticas com relação ao preço de serviços bancários.

Com relação às demais variáveis explicativas, a Tabela 3.6 resume os sinais encontrados nos vários modelos. O Apêndice traz o significado de cada uma das variáveis. A consideração de diferentes conjuntos de instrumentos não alterou o

TABELA 3.6  
RESULTADOS PARA AS DEMAIS VARIÁVEIS EXPLICATIVAS

Variável	Depósitos a prazo		Depósitos à vista + poupança		Empréstimos	
	Modelos 1 e 3	Modelos 2 e 4	Modelos 1 e 3	Modelos 2 e 4	Modelos 1 e 3	Modelos 2 e 4
agenpabmun	+***	+***	+***	+	+***	-
agenpabbr	+***	+***	+***	+***	+***	+***
paemun	-***	+	-	+	-***	+**
paebr	+***	+*	-***	-***	-***	-***
estados	-***	-***	-	+***	-***	-***
densagenpab	+***	+***	+***	+***	+***	+***
denspae	+	-	+	-***	-***	-***
desppub	+***	+***	+*	-	-***	-**
idadenum	+***	-	+***	+***	+***	+***
funcioag	+***	+***			-	-***
pibmun	-***	-***	-***	-***	-*	-
pibcapita	+***	-***	+***	+***	+***	-
pibcapsq	-***	-***	-***	-***	-***	-***
area	-***	-***	-***	-***	-***	-***
densidade	-***	-***	-***	-***	-***	-***
agpabmunpibcap		-*		+**		+**
agpabbrpibcap		-		+***		-***
paemunpibcap		-*		-		-***
paebrpibcap		-		-***		+***
estpibcap	+***		-***		+**	
densagpabpibcap		-	-***		-***	
denspaepibcap		-		+**		+
desppubpibcap		-**		+		-
idadepibcap	+***		-		+**	
funcagpibcap		+**			+***	

\* Significativo a 10%.

\*\* Significativo a 5%.

\*\*\* Significativo a 1%.

sinal de nenhum coeficiente. Por esse motivo, a tabela agrupa os resultados para os modelos 1 e 3, bem como para os modelos 2 e 4. Quando a significância estatística se altera entre os modelos agrupados, a tabela reporta o caso em que a significância é a menor.

Os resultados da Tabela 3.6 indicam que o número de agências mais postos de atendimento bancário no município, o número de tais estabelecimentos no Brasil e a densidade desses estabelecimentos têm impacto positivo sobre a demanda por produtos bancários de um banco. Esses resultados são esperados, indicando uma valorização dos agentes econômicos de tais características. Outro resultado obtido é que a idade do banco também tem impacto positivo sobre sua demanda de produtos, o que pode indicar algum efeito reputação.

Quanto às demais características bancárias, elas têm efeitos distintos sobre a demanda dos diferentes produtos bancários. Alguns coeficientes não são robustos à introdução dos termos de interação, enquanto outros coeficientes têm sinal oposto ao esperado.

Os coeficientes estimados para os preços e reportados nas Tabelas 3.1 a 3.3 serão utilizados para realizar as simulações na próxima seção.

### **3.6 TESTE DE PODER DE MERCADO**

Nesta seção, utilizaremos as expressões derivadas na Seção 3.4, bem como os coeficientes para a demanda pelos diversos produtos apresentados na Seção 3.5, para realizar um teste de poder de mercado para a indústria bancária no Brasil.

Para cada um dos modelos estimados, computamos a margem preço-custo absoluta compatível com concorrência de Bertrand e com a solução de cartel para cada banco da amostra. Tais margens são dadas pelas expressões que aparecem à direita da igualdade em (15), (18) e (19) para o modelo de Bertrand e (22), (25) e (26) para o modelo de cartel.

As Tabelas 3.7 e 3.8 reportam os resultados, tanto para a mediana dos bancos quanto para a média ponderada (pelo volume de cada produto) dos mesmos. As tabelas também mostram o valor “observado” para a margem preço-custo, dado pelo lado esquerdo das expressões do parágrafo anterior. Note-se que os valores “observados” não dependem dos coeficientes da demanda sendo, portanto, invariantes com relação aos distintos modelos. Entretanto, para o cálculo desses valores observados é necessário ter uma estimativa dos custos marginais respectivos. Os valores observados são então apresentados para dois conjuntos possíveis de estimativas desses custos marginais. O primeiro conjunto de valores observados

TABELA 3.7  
**MARGEM PREÇO-CUSTO ABSOLUTA — 2002**  
[em % a.m.]

Produto	Observado		Bertrand		Cartel	
	Mediana	Média	Mediana	Média	Mediana	Média
Depósitos a prazo	-0,0761	0,0233				
	-0,8857	-1,0377				
Modelo 1			0,0698	0,2024	0,1034	0,2274
Modelo 2			0,0675	0,1638	0,0956	0,1842
Modelo 3			0,0705	0,1879	0,1019	0,2111
Modelo 4			0,0691	0,1486	0,0952	0,1673
Serviços bancários	1,5013	1,4764				
	1,2177	1,1420				
Modelo 1			1,1749	1,2779	3,0033	2,9329
Modelo 2			0,8472	0,9051	2,1884	2,1499
Modelo 3			1,2152	1,3198	3,1088	3,0375
Modelo 4			0,8980	0,9593	2,3198	2,2791
Empréstimos	2,1165	2,8452				
	1,3438	1,9266				
Modelo 1			3,1613	3,2046	3,8225	4,0370
Modelo 2			2,5262	2,5609	3,0546	3,2260
Modelo 3			4,0519	4,1075	4,8994	5,1743
Modelo 4			3,0775	3,1197	3,7212	3,9300

Nota: Na coluna “observado”, a primeira linha assume custos marginais nulos; a segunda linha estima os custos marginais a partir de Fipecafi (2004).

supõe que os custos marginais sejam todos nulos, cujos resultados são apresentados na primeira linha para cada produto. Um segundo conjunto de valores observados é apresentado, utilizando-se estimativas de custos marginais obtidas a partir de Fipecafi (2004), sendo esses resultados mostrados na segunda linha para cada produto.

O estudo da Fipecafi utiliza informações de controladoria a partir de uma amostra de nove bancos para decompor os custos administrativos (custos operacionais e de recursos humanos) associados a três áreas de um banco comercial:

TABELA 3.8  
**MARGEM PREÇO-CUSTO ABSOLUTA — 2003**  
[em % a.m.]

Produto	Observado		Bertrand		Cartel	
	Mediana	Média	Mediana	Média	Mediana	Média
Depósitos a prazo	0,0818	0,0919				
	-0,5683	-0,5530				
Modelo 1			0,0665	0,1870	0,08784	0,2027
Modelo 2			0,0653	0,1526	0,0833	0,1655
Modelo 3			0,0676	0,1742	0,0878	0,1889
Modelo 4			0,0673	0,1393	0,0842	0,1513
Serviços bancários	1,3466	1,1387				
	1,0683	0,7819				
Modelo 1			1,1753	1,2707	3,3701	3,2772
Modelo 2			0,8473	0,8914	2,4717	2,4171
Modelo 3			1,2156	1,3114	3,4904	3,3957
Modelo 4			0,8982	0,9447	2,6202	2,5624
Empréstimos	2,2619	2,8032				
	1,5464	1,9793				
Modelo 1			3,1617	3,2038	3,7750	3,9750
Modelo 2			2,5265	2,5602	3,0167	3,1765
Modelo 3			4,0524	4,1064	4,8385	5,0949
Modelo 4			3,0778	3,1189	3,6750	3,8697

Nota: Na coluna “observado”, a primeira linha assume custos marginais nulos; a segunda linha estima os custos marginais a partir de Fipecafi (2004).

de empréstimos, de tesouraria e área de serviços. Note-se que as atividades de captação não são consideradas uma área específica, estando os custos associados a tais atividades distribuídos entre as três áreas.

Os resultados apresentados no estudo da Fipecafi sugerem que 33% da receita de juros das atividades de empréstimo são destinados à cobertura dos custos administrativos ao passo que para a receita de serviços, a alocação de custos corresponde a 84%. Parte de tais custos deve, contudo, ser alocada às atividades de captação. Supondo que os processos identificados no estudo relativos a “relacionamento com

clientes” e à “captação de recursos” estejam relacionados inteiramente à atividade de captação, temos que 38,89% dos custos administrativos de atividades de empréstimos e 14,33% dos custos administrativos de atividades de serviços estão relacionados a esses dois processos.

Assim, para os propósitos deste trabalho, considera-se que 20,03% da receita de juros das atividades de empréstimo sejam destinados à cobertura dos custos administrativos, que 71,7% da receita de serviços sejam destinados à mesma finalidade e que 52,88% das despesas de juros com depósitos a prazo sejam alocados para cobrir gastos administrativos.

Com essas hipóteses, foram calculados os valores “observados” da margem preço-custo apresentados sob a segunda linha de cada produto nas Tabelas 3.7 e 3.8.

Para depósitos a prazo, os resultados indicam que mesmo o modelo de Bertrand superestima o grau de poder de mercado. Existe também uma diferença considerável entre a mediana dos valores e os valores médios, sendo este último muito maior que o primeiro. Esse resultado revela que a margem preço-custo absoluta, tanto predita pelo modelo quanto observada, é maior para bancos com grande participação de mercado.

Isso, no entanto, não significa que exista exercício de poder de mercado para esse produto. Os resultados mostram que, mesmo sob a suposição de que o custo marginal de tais produtos seja zero, os valores “observados” são inferiores aos respectivos valores preditos pelo modelo de Bertrand, com exceção do valor mediano para o ano de 2003. Neste último caso, o valor observado de 0,082% a.m. para a margem preço-custo é superior ao valor predito sob o modelo de Bertrand (0,067% a.m.), mas ainda é inferior à margem predita sob cartel (0,088% a.m.). Quando se consideram os custos marginais de se prover tais produtos, contudo, as margens observadas tornam-se negativas, indicando que o banco (mediano ou médio) tem prejuízo nessas operações.

Para serviços bancários, os valores observados sob custos marginais zero são maiores que os valores preditos sob o modelo de Bertrand, mas ainda significativamente inferiores aos valores preditos sob o modelo de cartel. Quando se incorpora o custo de provisão desses serviços, os valores observados estão dentro da margem de variação encontrada para os distintos modelos sob a hipótese de Bertrand. Assim, o modelo de Bertrand parece se constituir em uma descrição bastante razoável do padrão de concorrência no mercado de serviços bancários.

Para o mercado de empréstimos bancários, os resultados são similares ao de depósitos a prazo. Ou seja, mesmo o modelo de Bertrand gera valores preditos para a margem preço-custo que superestimam os valores observados. Assim, mesmo

sob a suposição de que os custos marginais de provisão de empréstimos sejam zero, apenas o valor médio para o modelo 2 é capaz de produzir um número que seja inferior ao observado. Para todos os outros casos, o valor observado (mesmo sob custo marginal zero) é inferior ao leque de valores preditos sob Bertrand. Quando se considera o efeito dos custos marginais de provisão de empréstimos, todas as margens preço-custo observadas são significativamente inferiores aos preditos pelo modelo de Bertrand.

Assim, a conclusão geral dos testes de poder de mercado é que mesmo o modelo de Bertrand parece superestimar o poder de mercado observado para depósitos a prazo e para atividades de empréstimo. Para o primeiro mercado, os resultados indicam que o banco (mediano ou marginal) não está operando com lucro econômico. Finalmente, o modelo de Bertrand parece ser adequado para descrever o mercado de serviços bancários. Em nenhum dos mercados, encontrou-se evidência de que o modelo de cartel seria uma representação adequada.

Além dos resultados para a mediana e a média ponderada do conjunto dos bancos, compararamos também a margem preço-custo absoluta observada para cada banco com seu valor predito. Em especial, verificamos se a razão preço-custo absoluta observada para um banco era inferior ao valor predito sob Bertrand; superior a esse, porém inferior ao valor predito sob cartel; ou superior ao valor predito sob cartel. As Tabelas 3.9 a 3.11 resumem os resultados para cada um dos produtos

TABELA 3.9  
DEPÓSITOS A PRAZO: MARGEM PREÇO-CUSTO ABSOLUTA OBSERVADA VERSUS PREDITA

Modelos	Menor que Bertrand		Entre Bertrand e Cartel		Maior que Cartel	
	2002	2003	2002	2003	2002	2003
<b>Modelo 1</b>						
% observado	97,67	100,00	0,00	0,00	2,33	0,00
<i>Market share (%)</i>	99,85	100,00	0,00	0,00	0,15	0,00
<b>Modelo 2</b>						
% observado	97,67	100,00	0,00	0,00	2,33	0,00
<i>Market share (%)</i>	99,85	100,00	0,00	0,00	0,15	0,00
<b>Modelo 3</b>						
% observado	97,67	100,00	0,00	0,00	2,33	0,00
<i>Market share (%)</i>	99,85	100,00	0,00	0,00	0,15	0,00
<b>Modelo 4</b>						
% observado	97,67	100,00	0,00	0,00	2,33	0,00
<i>Market share (%)</i>	99,85	100,00	0,00	0,00	0,15	0,00

TABELA 3.10  
SERVIÇOS BANCÁRIOS: MARGEM PREÇO-CUSTO ABSOLUTA OBSERVADA VERSUS PREDITA

Modelos	Menor que Bertrand		Entre Bertrand e Cartel		Maior que Cartel	
	2002	2003	2002	2003	2002	2003
<b>Modelo 1</b>						
% observado	33,33	47,73	66,67	50,00	0,00	2,27
<i>Market share (%)</i>	90,20	97,68	9,80	2,30	0,00	0,02
<b>Modelo 2</b>						
% observado	0,00	27,27	93,75	70,45	6,25	2,27
<i>Market share (%)</i>	0,00	91,73	99,70	8,25	0,30	0,02
<b>Modelo 3</b>						
% observado	35,42	50,00	64,58	47,73	0,00	2,27
<i>Market share (%)</i>	90,59	97,83	9,41	2,15	0,00	0,02
<b>Modelo 4</b>						
% observado	0,00	34,09	95,83	63,64	4,17	2,27
<i>Market share (%)</i>	0,00	92,49	99,70	7,49	0,30	0,02

bancários mostrando tanto a proporção de observações para cada um dos três intervalos definidos, como a participação de mercado dos bancos em cada intervalo.

Os resultados para depósitos a prazo não mostram divergências entre os diversos modelos estimados nem tampouco para os dois anos considerados. Tanto do ponto de vista do número de bancos quanto com respeito à participação de mercado dos mesmos, a quase totalidade de observações está associada a valores para a margem preço-custo absoluta que são inferiores aos preditos pelo modelo de Bertrand. Assim, condizente com os resultados anteriormente apresentados, conclui-se que o grau de poder de mercado observado para depósitos a prazo é ainda inferior ao que seria compatível com o modelo de Bertrand.

Para serviços bancários (Tabela 3.10), nenhum dos modelos estimados apresenta número significativo de observações para a margem preço-custo absoluta acima dos valores preditos para cartel. Para os dois outros intervalos, contudo, existe grande diferença entre os modelos. Os modelos que não incluem interação de características observáveis do banco com a renda *per capita* do município (modelos 1 e 3) mostram que a maior parte dos bancos gera margens compatíveis com os intervalos observados entre Bertrand e cartel. Do ponto de vista de participação

TABELA 3.11  
**EMPRÉSTIMOS: MARGEM PREÇO-CUSTO ABSOLUTA OBSERVADA VERSUS PREDITA**

Modelos	Menor que Bertrand		Entre Bertrand e Cartel		Maior que Cartel	
	2002	2003	2002	2003	2002	2003
<b>Modelo 1</b>						
% observado	85,85	83,02	6,60	7,55	6,60	9,43
<i>Market share (%)</i>	87,80	97,31	3,19	1,41	9,01	1,29
<b>Modelo 2</b>						
% observado	75,47	75,47	12,26	6,60	12,26	17,92
<i>Market share (%)</i>	55,40	78,71	33,05	14,18	11,55	7,11
<b>Modelo 3</b>						
% observado	96,23	89,62	0,00	2,83	3,77	7,55
<i>Market share (%)</i>	92,31	98,49	0,00	0,82	7,69	0,69
<b>Modelo 4</b>						
% observado	85,85	82,08	4,72	7,55	9,43	10,38
<i>Market share (%)</i>	87,79	97,30	2,90	1,36	9,31	1,34

no mercado, contudo, tais bancos não são significativos. Ou seja, os bancos de maior participação têm margens preço-custo inferiores ao predito pelo modelo de Bertrand.

Para os modelos 2 e 4, por outro lado, existe uma grande diferença entre 2002 e 2003. Enquanto no primeiro ano, quase todos os bancos tinham margens no intervalo de valores entre Bertrand e cartel, para 2003, pelo menos do ponto de vista de participação no mercado, os bancos estavam majoritariamente alocados no intervalo de valores inferior a Bertrand.

Em suma, a conclusão anterior de que o modelo de Bertrand poderia ser uma boa representação do mercado de serviços bancários não é contradita pelos resultados da Tabela 3.10. A Tabela 3.11 apresenta os resultados para o mercado de empréstimos.

Um resultado interessante do mercado de empréstimos é que, do ponto de vista de participação no mercado, houve um aumento na concorrência entre 2002 e 2003. Esse aumento é representado pela elevação da participação de mercado dos bancos cujas margens preço-custo são inferiores aos valores preditos sob Bertrand.

Os resultados para o mercado de empréstimos também apontam no sentido de confirmar os resultados prévios, ou seja, de que a estrutura de mercado parece ser mais competitiva do que a descrita pelo modelo de Bertrand.

É importante ressaltar que os resultados anteriores são fortemente dependentes da correta especificação da demanda pelos produtos bancários. Uma questão associada a esse problema é a escolha da forma funcional adequada. Ou seja, será que os resultados apresentados seriam distintos se a forma funcional adotada fosse outra?

Os resultados de Crooke *et alii* (1999) permitem tecer algumas considerações sobre essa questão. Eles compararam as elasticidades-preço (e também os preços pós-fusões) associadas a quatro formas funcionais distintas para a demanda em mercados de bens diferenciados. Os sistemas de demanda considerados foram Almost Ideal Demand Systems (AIDS), *logit*, linear e log-linear (elasticidade constante). Os resultados indicam que as elasticidades são menores (e, portanto, os preços pós-fusão são maiores) para o caso de demanda log-linear, seguido de AIDS, de *logit* e de demanda linear.

Esses resultados sugeririam que a utilização da forma funcional *logit*, no presente trabalho, poderia subestimar o poder de mercado comparativamente à forma funcional linear. Do ponto de vista de produtos bancários, contudo, esta última forma funcional parece ser muito restritiva, além de carecer de embasamento teórico para fundamentá-la (teoria do consumidor).

Uma comparação talvez mais interessante é o contraste entre a forma funcional *logit* e os modelos de coeficientes aleatórios que introduzem termos de interação entre as características observáveis dos produtos e as características não-observáveis dos indivíduos. Infelizmente, os autores desconhecem trabalhos que utilizem esse modelo para o caso de produtos bancários. Em outros mercados, Nevo (2001) chega ao resultado de que a especificação *logit* gera margens preço-custo preditas inferiores ao modelo de coeficientes aleatórios para o caso de cereais matinais.

Tais resultados indicam que, caso a utilização da forma funcional *logit* gere resultados viesados, o viés será no sentido de superestimar e não de subestimar o poder de mercado.

### **3.7 CONSIDERAÇÕES FINAIS**

O presente trabalho apresentou resultados de um teste de poder de mercado para a indústria bancária no Brasil. A fim de alcançar esse objetivo, o primeiro passo foi a apresentação de um modelo de demanda por produtos bancários e a elaboração de

um modelo de comportamento dos bancos sob as hipóteses de concorrência de Bertrand e de cartel. A segunda etapa foi a estimação de diferentes especificações do modelo de demanda para três produtos bancários: *a)* depósitos a prazo; *b)* soma de depósitos à vista e poupança; e *c)* empréstimos. Por fim, foram apresentados testes de poder de mercado que são inéditos na literatura econômica brasileira.

Os resultados das estimativas da função de demanda para os diferentes produtos bancários foram razoavelmente robustos e as principais variáveis explicativas foram significativas e apresentaram os sinais esperados. Na demanda por depósitos a prazo, a taxa de juros desses depósitos teve um impacto positivo sobre sua própria demanda e o preço dos serviços bancários teve um impacto negativo. Já para a demanda por depósitos à vista mais poupança, tanto o impacto do preço do depósito a prazo quanto o preço dos serviços apresentaram um impacto negativo. Na estimativa da demanda por empréstimos, o impacto da taxa de juros desse produto foi negativo. As magnitudes das elasticidades-preço obtidas a partir das estimativas de demanda foram razoáveis.

Além dos resultados citados anteriormente, também se estimou que o número de agências mais postos de atendimento bancário no município, o número de tais estabelecimentos no Brasil e a densidade desses estabelecimentos têm impacto positivo sobre a demanda por produtos bancários de um banco. Resultados que indicaram uma valorização dos agentes econômicos de tais características. Outro resultado interessante foi o de que a idade do banco também tem impacto positivo sobre sua demanda de produtos, o que pode indicar algum efeito reputação.

Na seção do teste de poder de mercado foi computada a margem preço-custo absoluta compatível com a concorrência de Bertrand e com a solução de cartel para cada banco da amostra e se comparou com os resultados observados. Para depósitos a prazo, os resultados indicaram que mesmo o modelo de Bertrand superestimou o grau de poder de mercado. Além disso, a margem preço-custo absoluta, tanto predita pelo modelo quanto observada, foi maior para bancos com elevada participação de mercado. No entanto, não se pôde concluir que exista exercício de poder de mercado para esse produto. Quanto aos serviços bancários, o modelo de Bertrand pareceu se constituir em uma descrição bastante razoável do padrão de concorrência desse mercado. Por fim, para o mercado de empréstimos, os resultados são similares ao de depósitos a prazo, isto é, mesmo o modelo de concorrência de Bertrand gerou valores preditos para a margem preço-custo que superestimaram os valores observados.

Os resultados aqui apresentados estão alinhados com outros estudos empíricos investigando o poder de mercado no setor bancário brasileiro. Nakane (2003)

apresenta uma resenha desses trabalhos, em que a principal conclusão é de que o setor bancário brasileiro é caracterizado por uma estrutura de mercado imperfeita. Porém, os testes até agora realizados rejeitam a hipótese de que os bancos brasileiros operam sob a forma de cartel. Acreditamos que a contribuição de nosso trabalho a essa literatura tenha sido a de reproduzir o mesmo resultado, porém obtendo-o sob bases empíricas e teóricas mais bem fundamentadas.

## BIBLIOGRAFIA

- BACEN. *Economia bancária e crédito — avaliação de 5 anos do projeto juros e spread bancário*. 2004.
- BELAISCH, A. *Do Brazilian banks compete?* International Monetary Fund, 2003 (Working Paper 03/113).
- BERRY, S. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, v. 25, p. 242-262, 1994.
- BERRY, S., LEVINSOHN, J., PAKES, A. Automobile prices in market equilibrium. *Econometrica*, v. 63, p. 841-890, 1995.
- . Differentiated products demand systems from a combination of micro and macro data: the new car market. *Journal of Political Economy*, v. 112, p. 68-105, 2004.
- CROOKE, P. et alii. Effects of assumed demand form on simulated postmerger equilibria. *Review of Industrial Organization*, v. 15, p. 205-217, 1999.
- DICK, A. *Demand estimation and consumer welfare in the banking industry*. Federal Reserve Board, 2002 (Finance and Economics Discussion Paper, 2002-58).
- FIPECAFI. *Estudo sobre a estrutura da taxa de juros no Brasil*. Relatório preparado para a Febraban, 2004.
- NAKANE, M. I. Concorrência e spread bancário: uma revisão da evidência para o Brasil. In: BANCO CENTRAL DO BRASIL. *Economia bancária e crédito — avaliação de 4 anos do projeto juros e spread bancário*, 2003.
- NEVO, A. A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of Economics and Management Strategy*, v. 9, p. 513-548, 2000
- . Measuring market power in the ready-to-eat cereal industry. *Econometrica*, v. 69, p. 307-342, 2001.
- PETTERINI, F. C., JORGE NETO, P. de M. *Competição bancária no Brasil após o plano Real*. Caen UFC, 2003, mimeo.
- WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. MIT Press, 2002.

## **APÊNDICE**

### **Siglas das variáveis explicativas apresentadas na Tabela 3.6**

#### **1 Características do banco**

- agenpabmun: soma do número de agências e do número de postos de atendimento bancário que um banco tem em um determinado município;
- agenpabbr: soma do número de agências e do número de postos de atendimento bancário que um banco tem no Brasil;
- paemun: número de postos de atendimento eletrônico que um banco tem em um determinado município;
- paebr: número de postos de atendimento eletrônico que um banco tem no Brasil;
- estados: número de estados onde um banco tem agência;
- densagenpab: densidade da soma de agências e de postos de atendimento bancário de um dado banco em um determinado município;
- denspae: densidade de postos de atendimento eletrônico de um dado banco em um determinado município;
- desppub: despesas de publicidade de um banco (valores negativos);
- idadenum: idade do banco em dias;
- funcioag: número de funcionários dividido pelo número total de agências de um banco.

#### **2 Características do município**

- pibmun: PIB do município;
- pibcapita: PIB por habitante do município;
- pibcapsq: quadrado do PIB por habitante do município;
- area: área do município;
- densidade: densidade populacional do município.

### 3 Interações

- agpabmunpibcap: agenpabmun\*pibcapita;
- agpabbrpibcap: agenpabbr\*pibcapita;
- paemunpibcap: paemun\*pibcapita;
- paebrpibcap: paebr\*pibcapita;
- estpibcap: estados\*pibcapita;
- densagpabpibcap: densagenpab\*pibcapita;
- denspaepibcap: denspae\*pibcapita;
- desppubpibcap: desppub\*pibcapita;
- idadepibcap: idadenum\*pibcapita;
- funcagpibcap: funcioag\*pibcapita.



# Capítulo 4

Cerca de 38 milhões de indivíduos (ou 21% da população residente brasileira) são beneficiários de algum tipo de plano ou seguro de saúde. O Capítulo 4 analisa diferentes aspectos da demanda de saúde suplementar no Brasil. Inicialmente, no Subcapítulo 4A, Andrade discute a regulamentação dos planos e seguros de saúde no Brasil quanto às regras de determinação de preços e de benefícios definidas na Lei 9.656 de 1998.

Em seguida, três estudos econométricos tratam com originalidade os diversos aspectos da demanda do setor, que lida com duas falhas de mercado derivadas da assimetria de informação entre seguradora e segurado: risco moral e seleção adversa. Quanto maior a cobertura do plano ou seguro de saúde, maior é o incentivo na escolha de mais utilização do cuidado — um problema de risco moral. A seleção adversa, *ex ante* ao estabelecimento do contrato, consiste na seleção, por parte da seguradora, de clientes com risco maior que o risco médio da sociedade. Em ambos os casos a provisão dos serviços é inefficiente.

Andrade e Maia, no Subcapítulo 4B, medem uma *proxy* do risco moral através de medidas de sobreutilização dos serviços de saúde suplementar em relação aos serviços públicos de saúde, aplicando modelos de contagem (binomial negativo) com dados da PNAD de 1988. Embora as autoras indiquem a existência de sobreutilização, esta pode refletir tanto a presença de problemas nos incentivos dos contratos de saúde suplementar quanto um problema de acesso ao setor público de saúde, ou mesmo as duas situações simultaneamente. Ainda com base na PNAD de 1998, Andrade e Maia estimam, no Subcapítulo 4C, a presença de seleção adversa dentro da população que possui plano de saúde privado e entre esta e a população coberta apenas pelo Sistema Único de Saúde (SUS). A seleção adversa é parametrizada pela correlação positiva entre variáveis de risco e pela

probabilidade de terem plano de saúde, num modelo de contagem estimado por Máxima Verossimilhança com Informação Completa (FIML). Os principais resultados indicam que atributos de risco observáveis, como sexo e idade, são importantes na decisão de adquirir plano de saúde. As características de risco não-observáveis não foram significativas.

No Subcapítulo 4D, Andrade, Paraíso e Brito, estimam diretamente a demanda por serviços de saúde considerando a população beneficiária do plano oferecido pela Sabesp. O comportamento dos indivíduos é modelado através de quatro equações para identificar separadamente gastos ambulatoriais e de internação. Conclui-se que os atributos de risco, idade, sexo e estado de saúde passada são bastante relevantes na determinação dos gastos com serviços de saúde.

## CAPÍTULO 4

# DEMANDA NO SETOR DE SAÚDE SUPLEMENTAR NO BRASIL

Mônica Viegas Andrade

Do Centro de Desenvolvimento e Planejamento Regional da UFMG

Ana Carolina Maia

Do Centro de Desenvolvimento e Planejamento Regional da UFMG

Daniela Goes Paraiso

Do Centro de Desenvolvimento e Planejamento Regional da UFMG

Rubens José Amaral de Brito

Do Centro de Desenvolvimento e Planejamento Regional da UFMG

## 4A. REGULAMENTAÇÃO DO SETOR

### 4A.1 INTRODUÇÃO

O objetivo deste trabalho é apresentar uma discussão sobre a regulamentação dos planos e seguros de saúde no Brasil, salientando dois aspectos específicos: regra de determinação de preços e especificação do desenho de benefícios. O texto está organizado em mais quatro seções além desta: a segunda e terceira seções tratam da caracterização do sistema de saúde brasileiro e do mercado de bens e serviços de saúde, respectivamente; a quarta seção apresenta os principais aspectos introduzidos com a regulamentação do setor de saúde suplementar e a quinta seção procura discutir os prováveis impactos dessa regulamentação.

### 4A.2 BREVE CARACTERIZAÇÃO DO SISTEMA DE SAÚDE BRASILEIRO

O sistema de saúde brasileiro passou por transformações importantes desde os anos 1960. Entretanto, somente a partir de 1988, com a Constituição, quando foi determinada a garantia do acesso aos serviços de saúde como um direito universal e igualitário de todos os brasileiros, e estabelecido o papel de cada instância governamental (municipal, estadual e federal) no provimento, financiamento e gerenciamento dos serviços de saúde, esse sistema se concretizou como um sistema

de saúde nacional.<sup>1</sup> Os serviços passam a integrar uma rede regionalizada e hierarquizada, constituindo um Sistema Único de Saúde (SUS), definido de acordo com as seguintes diretrizes: *descentralização*, isto é, a execução dos serviços deve ser comandada pelos estados e municípios, cabendo à instância federal as funções de planejamento, fiscalização e controle; e *atendimento integral* com prioridade nas ações preventivas e participação social.<sup>2</sup> A assistência à saúde privada é livre e as instituições privadas podem participar de forma complementar ao SUS mediante contrato de direito público ou convênio, de preferência com entidades filantrópicas e sem fins lucrativos.<sup>3</sup>

Os recursos para financiamento do SUS integram o Orçamento da Seguridade Social. Suas principais fontes de financiamento advêm de contribuições sociais — contribuição sobre a folha salarial, contribuição sobre o lucro líquido das empresas (CSLL), Contribuição para o Financiamento da Seguridade Social (Cofins), contribuição sobre o Programa de Integração Social (PIS), contribuições sobre o programa de formação do Patrimônio do Servidor Público (Pasep) e, mais recentemente, foi criada a Contribuição Provisória sobre Movimentações Financeiras (CPMF).

A principal mudança apresentada pelo SUS é a descentralização da gestão dos serviços ou a *municipalização da saúde*, o que transfere a responsabilidade de gerir a oferta de serviços de saúde da União e dos estados para o município. A *municipalização* define o município como responsável pelo provimento e pela gestão dos serviços de saúde no seu território. Essa estratégia de descentralização significa uma tentativa de transformar os municípios em unidades gestoras de fato, e de tornar a identificação e a busca de soluções dos problemas de saúde mais próximas dos lugares onde eles acontecem.

A operacionalização dessa descentralização tem se dado através da edição de Normas Operacionais Básicas (NOBs), pelo Ministério da Saúde. As NOBs estabelecem regras e procedimentos a serem adotados pelos municípios para se tornarem gestores, possibilitando-lhes assumir a gestão e o provimento dos serviços de saúde locais de acordo com suas possibilidades. Em termos de competência, caberia ao

1. Segundo a Constituição Federal, Título VIII, Capítulo II, artigo 196, a saúde é um direito de todos e dever do Estado, garantido mediante políticas sociais e econômicas que visem à redução do risco de doença e de outros agravos e ao acesso universal igualitário às ações e serviços para sua promoção, proteção e recuperação.

2. A participação da comunidade no SUS é realizada através de duas instâncias colegiadas: a Conferência de Saúde e o Conselho de Saúde. A Conferência de Saúde contém representação de vários segmentos sociais e se reúne a cada quatro anos a fim de propor diretrizes e planos na gestão da saúde. O Conselho de Saúde é um órgão colegiado composto de representantes do governo, prestadores de serviço, profissionais de saúde e usuários do sistema, atuando na formulação de políticas e estratégias e no controle da execução da política de saúde.

3. O SUS foi regulamentado através da Lei 8.080, de 19/09/1990, também denominada Lei Orgânica do SUS, e da Lei 8.142, de 28/12/1990, que dispõe sobre a participação da comunidade na gestão do SUS e sobre as transferências de recursos do governo federal para as instâncias estadual e municipal.

governo federal apenas a intermediação de algumas das ações entre os estados no campo da saúde pública e da vigilância sanitária. Os estados deveriam realizar a intermediação das ações entre os municípios no que tange à saúde pública e à vigilância sanitária, bem como na criação de consórcios municipais.<sup>4</sup> Os municípios seriam as unidades básicas de gestão e prestação dos serviços. Os municípios podem ser enquadrados em tipos de gestão diferenciada, dependendo da infra-estrutura local existente. A transferência de recursos financeiros para custeio dos serviços médicos é realizada de acordo com o tipo de gestão a que o município se habilita.

Recentemente, o governo federal editou uma nova norma operacional, a Norma Operacional da Assistência à Saúde (Noas-SUS), que pretende dar ênfase à regionalização dos serviços de saúde. A Noas redefine o papel dos estados na organização dos serviços de saúde, que passam a ser responsáveis pela execução de um plano diretor regional que permita a compatibilização da rede de serviços de saúde regional entre os municípios que a compõem. Além disso, a Emenda Constitucional 29, editada em setembro de 2000, assegura um nível mínimo de recursos destinados à saúde nas três esferas de governo. Atualmente, o governo federal só tem uma política de financiamento definida para os gastos de custeio, ficando os gastos de investimentos, em princípio, a cargo dos estados e municípios. Com a edição da Emenda Constitucional os estados e municípios estarão obrigados a destinar recursos para a ampliação da oferta de serviços de saúde.<sup>5</sup>

O sistema de saúde brasileiro se caracteriza como um sistema misto. Os setores privado e público coexistem no provimento e no financiamento dos serviços de saúde. A medicina privada no Brasil está organizada em diversas modalidades que se diferenciam na forma de acesso e sistema de pagamento, como também nos benefícios ofertados. As principais formas institucionais de compra e venda de serviços privados de assistência médica no Brasil são:

*a) A medicina de grupo, constituída de empresas médicas que administram planos de saúde para empresas, indivíduos e famílias, com estrutura de atendimento*

4. Para garantir a integração e harmonia na organização das ações de saúde do sistema, existem a Comissão Intergestores Tripartite (CIT), composta por representantes do gestor federal e dos gestores estaduais e municipais, e a Comissão de Intergestores Bipartite (CIB), composta por representantes do gestor estadual e dos gestores municipais. As CIB e CIT são instâncias de negociação e pactuação.

5. Emenda Constitucional 29 do Senado Federal, de 13 de setembro de 2000. De acordo com essa emenda, a União deveria aplicar na saúde em 2000 o equivalente ao montante empregado em ações e serviços públicos de saúde no ano anterior (1999), acrescido de no mínimo 5%. E de 2001 a 2004, deverá ser aplicado o valor apurado no ano anterior, corrigido pela variação nominal do PIB. Aos estados e ao Distrito Federal cabe aplicar 12% da arrecadação dos impostos a que se referem os artigos da lei. No caso dos municípios, devem ser aplicados 15% do produto da arrecadação dos impostos a que também se referem os artigos da lei. Tal emenda define também que dos recursos da União, no mínimo 15%, serão aplicados nos municípios, segundo o critério populacional, em ações e serviços básicos de saúde. Assim, entre 2000 e 2004 os municípios terão de elevar seus gastos com a saúde de 7% para 15% da arrecadação. Os estados aumentarão os gastos de 7% para 12% da arrecadação. Em 2000 a União elevou a aplicação de recursos na saúde em 5%. Entre 2001 e 2004, os gastos federais deverão ser corrigidos pela variação nominal do PIB.

baseada principalmente na utilização de serviços e uma rede credenciada. É o modelo mais próximo das Health Maintenance Organizations americanas [ver Andrade e Lisboa (2000)].

*b)* As cooperativas médicas, instituições nas quais os médicos são simultaneamente sócios e prestadores de serviços. A principal representante dessa modalidade é a Unimed.

*c)* Planos geridos pelas próprias empresas, também denominados autogestão. Em geral a autogestão combina a administração própria com outras modalidades, por exemplo, o acesso a determinados serviços de rede credenciada. Na maior parte dos casos, o sistema de pagamento nessa modalidade consiste de tarifação segundo o risco da população coberta (*community rating*) e não como realizado usualmente, considerando as características próprias dos indivíduos (*experience rating*). Além disso, em alguns casos a participação em planos básicos é compulsória e a parcela de contribuição dos empregados é progressiva, já que na maior parte das vezes é determinada como percentual dos salários.

*d)* Seguro-saúde. Esta modalidade segue as características usuais das seguradoras. A instituição provedora realiza a intermediação financeira entre determinado grupo de indivíduos através da prática do reembolso para os serviços utilizados. A mudança recente na legislação brasileira permite às seguradoras, além de realizarem a intermediação financeira, contratarem a rede credenciada de serviços.<sup>6</sup>

A grande expansão do setor de medicina suplementar no Brasil se deu no período entre 1987 e 1994, quando houve crescimento de 73,4% da população coberta, que passou de 24,4 milhões de segurados para 42,3 milhões. As estimativas existentes apontam hoje um contingente de cerca de 25% da população total no Brasil como compradora dos planos e seguros de saúde. O estado com maior participação é São Paulo. O acesso aos serviços privados, entretanto, é bastante desigual entre os grupos sociais e entre as regiões do Brasil. A expansão da medicina privada se deu de forma diferenciada entre as diversas modalidades. Muito embora o setor de medicina privada tenha sido sempre importante no provimento de serviços de saúde no país, a relação entre o setor público e privado não foi concebida no desenho do sistema de saúde nacional proposto a partir da Constituição de 1988. Apenas recentemente, com a regulamentação do setor de medicina suplementar, o governo federal propôs uma intervenção direta neste setor.<sup>7</sup>

---

6. Circular 05 da Susep de 1989.

7. Os planos e seguros de saúde foram regulamentados através da Lei 9.656, de 3 de junho de 1998.

#### **4A.3 ASSIMETRIA DE INFORMAÇÃO E BENS MERITÓRIOS: CARACTERIZAÇÃO DO SETOR SAÚDE**

O setor de bens e serviços de saúde apresenta pelo menos quatro características principais que o diferenciam da maioria dos demais setores econômicos, e que podem resultar em perda de bem-estar social na ausência de regulação. Em primeiro lugar, diversos serviços e produtos de saúde se caracterizam como bens credenciais: nesse caso é necessário, em geral, a certificação de um profissional especializado para indicar o produto ou serviço a ser consumido em cada caso específico, e para atestar sua qualidade. Assim, na ausência de uma certificação pública reconhecida pelos consumidores como confiável, a reputação do provedor do bem ou serviço passa a ser relevante na determinação tanto das decisões de consumo por parte dos pacientes, quanto de prescrição médica por parte dos profissionais de saúde.

Em segundo lugar, o consumo de produtos e serviços de saúde se caracteriza pela dissociação entre o consumidor final e o agente responsável pela indicação terapêutica. Essa dissociação pode resultar em problemas de agência: por um lado, o paciente pode desejar maximizar sua utilidade esperada, levando em consideração tanto os possíveis efeitos terapêuticos dos diversos tratamentos opcionais existentes quanto os custos necessários; por outro, o médico, ou o responsável pela prescrição de tratamento, pode estar preocupado apenas com os efeitos terapêuticos ou com que o paciente utilize seus serviços.

Em terceiro lugar, alguns produtos e equipamentos do setor saúde se caracterizam por elevados gastos com pesquisa e desenvolvimento de novos processos e, sobretudo, novos produtos. A especificidade dos novos produtos no caso do setor de medicamentos, por exemplo, novas moléculas, permite maior eficácia na implementação de lei de patentes do que na maioria dos demais setores econômicos. Mansfield (1986), por exemplo, estima que na ausência dessa lei nos Estados Unidos, cerca de 60% dos produtos do setor não seriam desenvolvidos e 65% não seriam introduzidos no mercado.

Essas características específicas do setor de bens e serviços de saúde resultaram, no caso dos Estados Unidos, em regulações específicas tanto da lei de patentes, visando garantir os incentivos necessários ao desenvolvimento de novos produtos pela indústria farmacêutica, quanto no estabelecimento de relações contratuais entre seguradoras, provedores de serviços médicos e consumidores, visando minorar os eventuais problemas de agência no consumo de serviços médicos.

Por fim, em quarto lugar, o consumo de diversos produtos de saúde por alguns indivíduos gera externalidades difusas sobre o resto da sociedade. Diz-se

que um bem gera externalidade difusa quando o consumo médio da sociedade tem impacto direto no bem-estar de cada indivíduo. Assim, por exemplo, a taxa média de vacinação na sociedade contra determinadas doenças afeta a probabilidade de um agente não-vacinado contrair a doença. No caso de externalidades difusas positivas, a escolha de mercado pelo consumo do bem é, em geral, menor do que a eficiente, pois os agentes apenas internalizam os benefícios privados do consumo do bem mas não seu impacto sobre o bem-estar dos demais agentes [Andrade (2000)].

A adoção de políticas públicas para o setor de saúde nos principais países desenvolvidos decorre, além das quatro possíveis fontes de ineficiência já discutidas, de uma quinta razão, esta relacionada a aspectos distributivos. O acesso a determinados serviços médicos em geral, e a medicamentos em particular, é considerado em diversos países como direito de cidadania, o que resulta na classificação desses bens e serviços como *meritórios*, isto é, bens e serviços a que todo cidadão deve ter acesso, sendo responsabilidade da política pública garantir o acesso universal a esses bens.

Entretanto, os gastos individuais em saúde, em alguns casos, e sobretudo os gastos com medicamentos, são relativamente inelásticos à renda: em grande medida, são condicionados pelo estado de saúde do paciente. Essa regressividade dos gastos com medicamentos, aliada a sua natureza meritória, levou diversos países a adotarem políticas públicas visando garantir o acesso a diversos produtos e serviços de saúde, ao menos pela população de menor renda. A forma específica dessa política não é uniforme entre os diversos países, incluindo desde o reembolso de certa fração de determinados gastos com saúde, até o seu provimento gratuito pelo setor público.<sup>8</sup>

A consideração dos bens e serviços de saúde como meritórios não implica, entretanto, a necessidade de financiamento e/ou provimento público. O acesso a esses bens e serviços pode ser garantido mediante diversos outros mecanismos, por exemplo, a implementação de um sistema de seguridade social compulsório público; o provimento público dos bens e serviços considerados meritórios: políticas de subsídio cruzado que permitam aos grupos com menores rendas o acesso ao seguro-saúde privado; e o financiamento público dos gastos com saúde para grupos específicos, entre outros.

Empiricamente, a experiência internacional mostra que os países têm optado por formas bastante diferenciadas de garantir esse acesso. Nos Estados Unidos, por exemplo, a maior parte da população tem acesso aos serviços médicos através

---

8. Para uma sistematização das principais políticas públicas de acesso a medicamentos, ver Lisboa *et alii* (2001).

do seguro privado, cabendo ao estado prover bens e serviços apenas aos grupos sociais mais vulneráveis (idosos, através do Medicare, e crianças e população de baixa renda, através do Medicaid). O sistema inglês, por sua vez, se caracteriza pelo financiamento e provimento público dos serviços médicos. Já no caso canadense, o Estado tem o monopólio do seguro de saúde. Desse modo, a primeira dificuldade apresentada pelo mercado de bens e serviços de saúde é a necessidade de garantir à população o acesso a esses serviços. A principal diferença entre os serviços de saúde e demais bens meritórios, como educação, está relacionada à estrutura de mercado do setor saúde, à incerteza dos indivíduos sobre o seu estado de saúde, e à existência de formas específicas de assimetria de informação.

O consumo de bens e serviços de saúde varia significativamente com os estados individuais da natureza. Como a maior parte desse risco é idiosincrático, não estando em geral associado a ocorrência de choques agregados na economia, a existência de um sistema de seguridade social, privado ou público, resulta em ganhos de bem-estar para os indivíduos. O estabelecimento de mecanismos de seguro, na ausência de problemas de assimetria de informação, permite aos indivíduos suavizar o seu consumo entre os diversos estados da natureza e períodos, reduzindo os gastos necessários na ocorrência de choques de saúde adversos.

Entretanto, uma importante característica do setor saúde é, precisamente, a existência de formas específicas de assimetria de informação entre provedores de serviços médicos, financiadores dos gastos com saúde — sejam públicos ou privados —, e os potenciais pacientes. Podem-se distinguir duas incertezas associadas ao consumo desses bens: incerteza quanto à necessidade de uso dos serviços médicos e incerteza quanto ao diagnóstico [Arrow (1963)]. Aliada à assimetria informacional existente entre os agentes nesse mercado, a incerteza torna o consumo dos serviços médicos bastante diferenciado dos demais. As relações de contrato entre os agentes desse mercado devem considerar os problemas de incentivos existentes, de modo a garantir o uso eficiente dos recursos disponíveis.

No que concerne ao comportamento estratégico dos agentes envolvidos nesse mercado, existem problemas de risco moral e seleção adversa que em sua maior parte independem da estrutura de propriedade associada ao provimento desses bens. O estabelecimento de um contrato de seguro-saúde, público ou privado, envolve, necessariamente, três grupos de agentes: o segurador, que pode ser o governo ou seguradoras privadas; os provedores de cuidados médicos, que, uma vez mais, podem ser públicos ou privados; e os segurados (pacientes potenciais). A relação do segurador, seja com os provedores, seja com os segurados, apresenta problemas de risco moral.

Na relação entre paciente e provedor, o problema de risco moral ocorre quando há o seguro pleno, na medida em que, nesse caso, os agentes tendem a sobreutilizar os serviços, já que o custo marginal do serviço demandado é zero. Esse problema de risco moral determina um *trade-off* entre os ganhos de bem-estar obtidos com a cobertura de seguro e a perda de bem-estar decorrente da sobreutilização dos serviços. Em outras palavras, a presença da cobertura de seguro gera um ganho de bem-estar para os indivíduos na medida em que permite diminuir a incerteza; por outro lado, quanto mais extensa for a cobertura de seguro, maiores os incentivos à utilização excessiva dos serviços, o que resulta em perda de bem-estar social, isto é, em filas, rationamentos ou em gastos sociais acima dos necessários à implementação do sistema de saúde, sejam eles financiados através de seguros-saúde privados, sejam através de impostos [ver Newhouse (1996)].

O desenho de mecanismos de incentivos capazes de garantir o provimento de bens e serviços de saúde e que não resultem na perda de recursos sociais desnecessários é um dos problemas centrais na regulação desse setor. Para isso é necessário entender o impacto de sistemas opcionais de incentivos sobre o comportamento dos diversos agentes envolvidos. Essa necessidade de análise do desenho dos mecanismos de incentivo ocorre tanto em um sistema privado de saúde quanto em um sistema público.

A relação entre provedores de serviços médicos e financiadores dos gastos realizados, sejam estes últimos os próprios pacientes, seguradoras privadas ou o Estado, também não está isenta de assimetria de informação e problemas de incentivo. O problema de risco moral ocorre porque os provedores podem induzir uma maior utilização dos serviços, já que possuem mais informação sobre o estado de saúde do paciente e sobre a necessidade de cuidados médicos do que os agentes financiadores. Os provedores podem sobreutilizar o volume de serviços tanto para elevar seus rendimentos diretos, se o sistema de reembolso for por tarefa, como para tentar garantir a certeza de um diagnóstico correto, caso existam ganhos de reputação.<sup>9</sup> Por um lado, o provedor de serviços médicos tem informação privilegiada sobre o diagnóstico realizado e os tratamentos mais indicados. Além disso, a remuneração do provedor pode depender do tipo de diagnóstico ou tratamento proposto, seja porque o diagnóstico ou tratamento pode resultar na contratação dos seus serviços, seja porque o sucesso do tratamento pode resultar em ganhos de reputação que significam maior renda no futuro. Os financiadores, por outro lado, podem contrabalançar os benefícios de diversos tipos de tratamento aos custos relativos requeridos. Por essa razão, o tratamento proposto pelos provedores

---

9. Esse argumento é formalizado em Lisboa e Moreira (2000).

pode não maximizar o bem-estar social, e até resultar em maiores gastos dos potenciais pacientes do que o socialmente ótimo, seja através de seguros-saúde mais caros, seja pela necessidade de maiores impostos para financiar um sistema público de saúde mais oneroso do que o necessário [Lisboa e Moreira (2000)].

Os problemas decorrentes da assimetria de informação existente entre os agentes nesse mercado podem ser ainda exacerbados pelo contexto em que as decisões de consumo são realizadas. Em sua maioria, os consumidores, ao decidirem comprar bens e serviços de saúde, dispõem de tempo limitado e em geral se encontram em momentos de vulnerabilidade, em que a capacidade de adquirir informações pode estar reduzida. Além disso, o consumo de alguns bens e serviços de saúde não permite o aprendizado através da repetição.

A assimetria de informação existente entre os agentes pode ainda resultar em problemas de seleção de risco. Nesse caso podem ocorrer duas situações: a seleção adversa de consumidores e o *cream skimming*. A seleção adversa se dá em situações em que o segurador não pode distinguir entre os indivíduos de diferentes classes de risco. Nesses casos, o segurador taxa o seguro com base no risco médio da sociedade. A seleção adversa de consumidores ocorre quando os indivíduos de risco baixo saem do *pool* de risco, uma vez que em conhecendo o seu risco, não estariam dispostos a pagar mais do que o prêmio atuarialmente justo. Desse modo, permanecem no *pool* do seguro apenas os indivíduos de risco elevado. Em alguns casos, a assimetria de informação pode resultar não só na ausência de mercado de seguros para alguns grupos, mas até na inexistência de mercado.

Problemas de seleção adversa podem ainda ser gerados através da imposição de mecanismos regulatórios propostos pelo governo com o objetivo de impedir a discriminação de indivíduos de riscos diferentes. Os governos, a fim de proteger determinados grupos sociais, em alguns casos, impedem que a seguradora venha a discriminar todos os grupos de risco, tentando forçar a implementação de um subsídio cruzado entre os grupos de risco diferente. Os agentes de risco baixo, entretanto, a menos que estejam sob situação de compra compulsória de seguro, preferem sair do *pool* de risco, configurando o problema de seleção adversa. [Andrade (2000), Newhouse (1996) e Newdeck e Podczeck (1996)].

O *cream skimming* ocorre na situação inversa, quando as seguradoras, através dos contratos propostos, conseguem segmentar o mercado e ofertar serviços diferenciados para os indivíduos com risco distinto. A oferta de contratos com auto-seleção, entretanto, em geral implica perda de bem-estar social para os agentes com baixo risco com relação à solução eficiente.

Os problemas de assimetria informacional existentes no mercado de bens e serviços independem da estrutura de propriedade do sistema de saúde. As relações entre os provedores de serviços médicos, seguradores, públicos ou privados, e potenciais pacientes são acometidas pelos mesmos problemas, independentemente de o sistema ser financiado/provado privada ou publicamente. Em ambos os casos, a existência de cobertura plena tende a aumentar a demanda por serviços médicos, e, portanto, os gastos sociais necessários, sejam financiados privadamente, mediante o seguro-saúde, sejam financiados por recursos públicos.

Desse modo, o aspecto distributivo — garantia de acesso aos bens e serviços de saúde — pode e deve ser, em certa medida, isolado do aspecto da eficiência dos mercados, que depende da capacidade de se implementar os incentivos corretos para os agentes no mercado. São dois problemas de natureza distinta, ainda que interligados. O primeiro se relaciona à definição de mecanismos que viabilizem o acesso a esses serviços para toda a população; o segundo diz respeito aos desenhos de contratos estabelecidos entre os agentes provedores e os consumidores desses serviços que tentem, na medida do possível, incorporar os problemas de assimetria informacional existentes.

Eventualmente, critérios de acesso podem resultar na inviabilidade de certos mecanismos de incentivo assim como considerações de incentivos podem resultar em alguma restrição de acesso. Entretanto, e esse é o ponto essencial, considerações sobre a garantia de acesso de modo algum implicam a necessidade de provimento público, nem considerações de incentivo necessariamente implicam restrições de acesso. Sobretudo, a discussão quanto às formas de regulação do setor saúde deve incorporar seus possíveis impactos sobre o comportamento dos diversos agentes envolvidos decorrentes das inevitáveis fontes de assimetria de informação existentes.

#### **4A.4 REGULAMENTAÇÃO DO SETOR DE SAÚDE SUPLEMENTAR**

Nesta seção descrevemos os principais aspectos do marco regulatório do sistema de saúde suplementar, de acordo com a Lei 9.656/98, que dispõe sobre o assunto. A regulação no mercado de saúde suplementar tem por objetivo corrigir os entraves que desequilibram a relação entre consumidores e operadores dos planos, procurando conciliar a dimensão assistencial e a econômico-financeira, fiscalizando a prestação de serviços à saúde e a viabilidade financeira das carteiras. Dividimos esta seção em três subseções, seguindo a tipologia utilizada por Mesquita (2002). A primeira trata do padrão de cobertura assistencial e da relação com o consumidor, em que destacamos a instituição do plano-referência, de oferta obrigatória por todas as operadoras. A segunda subseção refere-se à relação entre os setores público e

privado no mercado de atenção à saúde. A terceira subseção destaca o funcionamento das operadoras de planos de saúde e o papel que a ANS exerce nesse mercado.

#### **4A.4.1 Padrão de cobertura assistencial e relação com o consumidor**

O padrão de cobertura assistencial e a relação com o consumidor serão analisados destacando-se três aspectos: cobertura assistencial, desenho dos contratos e adaptação dos contratos à regulamentação criada.

##### **4A.4.1.1 Cobertura assistencial**

Um dos principais objetivos da Lei 9.656/98 foi o estabelecimento de cláusulas e coberturas básicas para o consumidor, com a instituição do plano-referência de assistência básica à saúde (artigo 10). Todas as operadoras, à exceção das autogestões, devem, obrigatoriamente, fornecer esse tipo de plano, que inclui: cobertura assistencial médico-ambulatorial e hospitalar, compreendendo partos e tratamentos realizados no Brasil com padrão de enfermaria, centro de terapia intensiva, ou similar, e quando necessária, a internação hospitalar.<sup>10</sup> Inclui-se na cobertura contratual a prestação de cirurgia plástica para reconstrução de mama no caso de câncer e excluem-se alguns tratamentos previstos no artigo 10, como tratamento clínico ou cirúrgico experimental, entre outros, com fins estéticos.

Além do plano-referência, as operadoras, não importa a modalidade do plano de saúde — individual, familiar ou coletivo — podem contratar um ou mais segmentos de cuidado definidos pela ANS, desde que respeitem as respectivas amplitudes de cobertura de serviços especificadas na regulamentação. São definidos os seguintes segmentos: ambulatorial, hospitalar com ou sem obstetrícia, e odontológico; que se diferenciam pela cobertura assistencial e padrão de atendimento (artigo 12 da Lei 9.656, de 3.06.98). Por último, ainda sobre a cobertura de serviços, a regulamentação estabelece também obrigatoriedade de atendimento nos casos de emergência e urgência pelos planos de saúde conforme artigo 35-C.

##### **4A.4.1.2 Desenho dos contratos**

É garantido ao consumidor, pelo artigo 13 da Lei 9.656/98, o direito à renovação do contrato, sem a cobrança de taxas ou qualquer outro valor no ato da renovação. O consumidor tem também a garantia de participar de planos privados de assistência à saúde, não podendo ser rejeitado em função da idade ou por ser portador de doenças (artigo 14). No que se refere ao período de carência, o artigo

10. A cobertura deve compreender todas as doenças citadas na Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID-10), da Organização Mundial de Saúde.

11 institui um período de 24 meses para doenças ou lesões preexistentes, ou seja, anteriores à vigência do contrato. O período de carência estabelecido no artigo 12, inciso V, é um período predeterminado no início do contrato, com um prazo máximo de 300 dias para partos, 180 dias para consultas, internações, cirurgias etc., e prazo máximo de 24 horas para a cobertura dos casos de urgência e emergência.

Além dos períodos de carência e cláusulas estabelecendo as condições de renovação dos contratos, a regulamentação específica ainda os itens que obrigatoriamente devem estar presentes nos contratos (artigo 16):

I — as condições de admissão;

II — o início da vigência;

III — os períodos de carência para consultas, internações, procedimentos e exames;

IV — as faixas etárias e os percentuais a que alude o *caput* do artigo 15;

V — as condições de perda da qualidade de beneficiário;

VI — os eventos cobertos e excluídos;

VII — o regime, ou tipo de contratação:

a) individual ou familiar;

b) coletivo empresarial; ou

c) coletivo por adesão;

VIII — a franquia, os limites financeiros ou o percentual de co-participação do consumidor ou beneficiário, contratualmente previstos nas despesas com assistência médica, hospitalar e odontológica;

IX — os bônus, os descontos ou os agravamentos da contraprestação pecuniária;

X — a área geográfica de abrangência;

XI — os critérios de reajuste e revisão das contraprestações pecuniárias;

XII — número de registro na ANS.

Ao firmar o contrato de seguro com o consumidor, as operadoras deverão se comprometer a manter o nível de sua rede assistencial durante toda a vigência do contrato. Conforme artigo 17, a substituição de entidade hospitalar ou redução de sua rede ocorrerá apenas mediante informação à ANS. O consumidor deverá ser informado e, se em período de internação, terá a despesa custeada até receber alta hospitalar.

No caso de o consumidor aposentar-se ou ser demitido pela empresa em que trabalha (a qual lhe oferecia um plano de seguro), ele terá direito a continuar como beneficiário, nas mesmas condições de cobertura assistencial de que dispunha quando da vigência de contrato de trabalho, desde que assuma o pagamento integral do plano (artigos 30 e 31).

#### **4A.4.1.3 Adaptação de contratos e reajuste de preços**

Em relação à adaptação de contratos, o artigo 35 da Lei 9.656/98, permite que os consumidores com contratos anteriores à lei decidam pela adaptação ou não à lei vigente, podendo permanecer no contrato anterior por prazo indeterminado. Os consumidores não podem ser prejudicados nas adaptações, principalmente na contagem dos períodos de carência, nos prazos para atendimentos de doenças preexistentes e nos prazos na aquisição de benefícios (rescisão de contrato de trabalho ou aposentadorias). Sendo assim, ao optar pela mudança, as garantias contratuais e de cobertura são asseguradas aos consumidores dos planos anteriores.

O artigo 15 da lei estabelece que os reajustes nos contratos, em razão da idade do consumidor, somente poderão ocorrer caso estejam previstas as faixas etárias e os percentuais de reajustes incidentes em cada uma, conforme normas expedidas pela ANS. Não será permitido o reajuste de mensalidades por mudança de faixa etária para consumidores com mais de 60 anos ou que participem do plano há mais de dez anos. O artigo 35-E dispõe do reajuste por faixa etária para os contratos anteriores à lei. Entre os principais direitos estão as necessidades de autorização prévia para reajuste das mensalidades para consumidores com mais de 60 anos, a impossibilidade de rescisão de contratos de seguro individuais ou familiares pelas operadoras, e a vedação à interrupção de tratamento hospitalar em leito clínico, cirúrgico ou em centro de terapia intensiva.

A RN 63, de 22 de dezembro de 2003, define os limites a serem observados para a adoção de variação de preço por faixa etária nos planos privados contratados a partir de 1º de janeiro de 2004. Os percentuais de variação em cada mudança de faixa etária deverão ser fixados pela operadora, observadas as seguintes condições: o valor fixado para a última faixa etária não poderá ser superior a seis vezes o valor da primeira, e a variação acumulada entre a sétima e a décima faixa não poderá ser superior à variação acumulada entre a primeira e a sétima faixa.

#### **4A.4.2 A relação público-privado**

Nesta subseção, nos propomos a apresentar os aspectos concernentes à relação entre os setores público e privado no sistema de saúde brasileiro a partir dos dispositivos

que regulamentam a assistência suplementar: os artigos 20 e 32 da Lei 9.656/98 foram eleitos como os que mais permeiam essa discussão. O primeiro trata das informações a serem fornecidas à ANS pelas operadoras de planos. Enquanto o segundo, do resarcimento ao SUS.

A experiência internacional recente em reformas dos sistemas de saúde tem apresentado uma tendência de incentivar a participação privada no financiamento/provimento dos bens e serviços do setor. Nesse contexto, o Estado passa a exercer um novo papel, com reorientação para regulação, controle e avaliação da atuação do setor privado.

No Brasil, a expressiva atuação do sistema suplementar de saúde aponta a ineficácia do sistema público em prover assistência integral e de qualidade a toda a população. Dessa forma, o investimento privado é essencial e deve ser incentivado de modo a contribuir para que a oferta pública e privada se complementem. A dupla cobertura dos indivíduos com plano, entretanto, teve um importante desdobramento com a implementação da Lei 9.656/98: o resarcimento da assistência prestada pelo SUS, mas de responsabilidade das operadoras de plano de saúde.

Entre as competências da ANS, destacamos (Lei 9.961):

- Estabelecer normas para o resarcimento ao SUS.
- Proceder a integração de informações com os bancos de dados do SUS.

Para atender ao artigo 32 da Lei 9.656/98, as operadoras de planos de assistência à saúde são obrigadas a fornecer, periodicamente, à ANS, todas as informações e estatísticas relativas a suas atividades, em especial aquelas que permitam a identificação de seus consumidores (artigo 20). O artigo 32 dispõe sobre o resarcimento, pelas operadoras, dos serviços de atendimento à saúde, previstos nos contratos, prestados a seus consumidores, em instituições públicas ou privadas, conveniadas ou contratadas, integrantes do SUS.

A ANS disponibilizará às operadoras a discriminação dos procedimentos realizados por cada consumidor e enviará a cobrança. A comparação dos usuários é feita pelo cadastro de beneficiários de planos de saúde com as Autorizações de Internação Hospitalar (AIHs). O resarcimento será cobrado de acordo com os procedimentos estabelecidos pela Tabela Única Nacional de Equivalência de Procedimentos (Tunep) aprovada pela ANS (RDC 18). A Tunep identificará os procedimentos e definirá os valores de referência. Dos valores estabelecidos pela Tunep, o Fundo Nacional de Saúde é reembolsado do montante pago pelo SUS, enquanto o prestador do serviço do SUS recebe a diferença entre a Tunep e a Tabela SUS.

#### **4A.4.3 Funcionamento das operadoras de planos de saúde e o papel da ANS<sup>11</sup>**

Nesta subseção analisaremos a regulamentação no que tange aos aspectos gerais de funcionamento das operadoras e fiscalização desse mercado a partir da Lei 9.656/98, bem como o papel a ser exercido pela ANS, criada pela Lei 9.961/00.

##### **4A.4.3.1 Funcionamento das operadoras de planos de saúde**

Entre as novidades trazidas pela lei podemos destacar: a introdução de ordem jurídica única para todos os operadores de planos de saúde; a instituição de multas e da fiscalização governamental; o estabelecimento de normas de funcionamento relacionadas à capacidade de prestação dos serviços e da viabilidade econômico-financeira; a proibição do requerimento de falência ou concordata por parte das operadoras e a possibilidade de o regulador promover a alienação da carteira para sanar irregularidades e preservar os direitos dos consumidores; a co-responsabilização dos dirigentes das operadoras por prejuízos causados a terceiros pelo descumprimento da lei.

A Lei 9.656/98 foi posteriormente alterada pela MP 2.177-44, de 24/08/2000, a qual visava aprimorar a regulação e fiscalização no setor. A principal alteração introduzida pela MP no texto original buscou o combate aos segmentos que desejavam transformar-se em exceção, com a modificação do alcance da lei, que passou do negócio da empresa (operação de planos e seguros) para o produto comercializado (produtos com características de planos de saúde). Todas as referências a seguros e seguradoras foram suprimidas da lei e ficou estabelecida a equivalência entre seguro de saúde e plano de saúde como um produto só. Todas as empresas que operam esse produto foram obrigadas a se constituírem como seguradoras especializadas. A constituição de pessoa jurídica independente passa a ser obrigatória para as instituições que exerçam outras atividades distintas da operação de planos de saúde.

Outras alterações introduzidas pela MP 2.177-44: criação da autorização para as operadoras encerrarem suas atividades com a exigência da transferência prévia da carteira; a comercialização passou a ser permitida apenas com o cadastramento das operadoras e produtos no órgão regulador; a ANS passa a ter competência para determinar o tipo de regime a que estará submetida a operadora que apresentar dificuldades; os prestadores de serviços tornam-se os credores preferenciais no caso da liquidação da operadora; as multas passam a variar conforme o tamanho da operadora e os recursos passam a ser depositados para a ANS; passa

---

11. Esta subseção é baseada em Mesquita (2002).

a ficar estabelecido o registro de bens garantidores, fundos e provisões, com sua alienação proibida; possibilidade da contratação de resseguro junto a empresas autorizadas pela Superintendência de Seguros Privados (Susep).

#### 4A.4.3.2 O papel da ANS

A ANS é criada no contexto de regulação do mercado de saúde suplementar nos termos da Lei 9.656/98. Entre os principais objetivos da agência, destacamos a proposição de políticas a fim de gerar equilíbrio nas decisões econômicas, proteger consumidores e promover um padrão de competitividade desejável no setor. A Lei 9.656/98 criou o Conselho Nacional de Saúde Suplementar (Consu) como órgão colegiado integrante da estrutura regimental do Ministério da Saúde e estabeleceu a composição interministerial do Consu. Com a MP 2.177-44 as competências normativas do Consu, bem como as do CNSP e da Susep foram transferidas para a ANS, com o conselho mantendo as funções de fixar as diretrizes de constituição, organização, funcionamentos e fiscalização das operadoras e supervisionar a execução dessas políticas. A referida MP também alterou a composição do Consu e vinculou a Câmara de Saúde Suplementar à ANS.

A Lei 9.961/00 criou a ANS, que absorveu as competências normativas, fiscalizadoras, de monitoramento e controle relativas à assistência de saúde suplementar anteriormente devidas à Susep, Ministério da Saúde, CNSP e Consu, como autarquia especial com autonomia financeira, administrativa, patrimonial, de gestão de recursos humanos e decisões técnicas, vinculada ao Ministério da Saúde e com atuação em todo o território nacional. O artigo 3º da referida lei estabelece que “A ANS terá por finalidade institucional promover a defesa do interesse público na assistência suplementar à saúde, regulando as operadoras setoriais, inclusive quanto à suas relações com prestadores e consumidores, contribuindo para o desenvolvimento das ações de saúde no país”.

Entre as principais competências da ANS listadas no artigo 4º da lei, podemos destacar: o estabelecimento das características gerais dos instrumentos contratuais utilizados nas atividades das operadoras; a elaboração do rol de procedimentos e eventos em saúde; a fixação de critérios para credenciamento e descredenciamento de prestadores de serviços junto às operadoras; o estabelecimento de normas para o resarcimento ao SUS; o estabelecimento de normas relativas à adoção e utilização, pelas operadoras de planos de assistência à saúde, de mecanismos de regulação do uso dos serviços de saúde; a normatização dos conceitos de doença e lesão preexistente; a autorização de reajustes e revisões de contraprestações pecuniárias dos planos privados de assistência à saúde, ouvido o Ministério da Fazenda; a autorização de funcionamento, cisão, fusão, incorporação,

alteração ou transferência de controle societário das operadoras de planos privados de assistência à saúde; a fiscalização do cumprimento e aplicação das penalidades pelo descumprimento das disposições da Lei 9.656/98 e suas regulamentações; o procedimento de liquidação extrajudicial e autorização ao liquidante para o requerimento de falência ou insolvência civil das operadoras; a determinação ou promoção da alienação da carteira de planos privados de assistência à saúde das operadoras.

#### **4A.5 POSSÍVEIS IMPACTOS DA REGULAMENTAÇÃO**

O objetivo desta seção é discutir alguns impactos da regulamentação, dando ênfase às regras que norteiam a especificação dos planos e seguros de saúde, segundo grupos de risco. Como apresentado na Seção 4A.3, o mercado de bens e serviços de saúde apresenta uma série de características específicas que o difere dos demais. Algumas dessas características acabam resultando em falhas de mercado, as quais tornam ineficiente a alocação de equilíbrio. Além dessas falhas de mercado, a natureza meritória dos bens e serviços de saúde torna o aspecto da eqüidade uma recorrente justificativa na literatura para que a oferta e financiamento dos serviços sejam realizados através do setor público. No Brasil, como visto na seção de caracterização do sistema, optamos por um sistema misto, em que os setores público e privado se complementam tanto no provimento como no financiamento. Uma das vantagens de sistemas públicos universais é a inexistência de problemas de seleção adversa. A escolha do desenho institucional de sistema de saúde tem sido um tema recorrente na literatura econômica, não existindo, entretanto, um consenso acerca do melhor desenho institucional. Na América Latina vários países — Colômbia, Chile, Argentina e Brasil — têm implementado reformas nos sistemas de saúde, com o objetivo comum de buscar garantir o acesso aos bens e serviços de saúde [Jack (2002)].

Um primeiro aspecto da regulamentação dos planos e seguros de saúde que chama a atenção é a relação entre os setores público e privado. Essa é uma discussão difícil e envolve a própria escolha do desenho institucional que realizamos, ou melhor, que estamos construindo. Desse modo, não pretendemos esgotar o assunto, mas, apenas iniciar uma reflexão sobre um tema tão relevante. O sistema de saúde brasileiro, diferentemente da maior parte dos sistemas dos países que optam por um modelo público universal, permite a coexistência dos setores público e privado, não havendo nenhuma regra ou arranjo institucional que impeça aos indivíduos terem duplo acesso ao sistema. Um dos aspectos mais relevantes introduzidos pela regulamentação, sem dúvida, é o resarcimento ao SUS dos serviços que este financia para a população que tem cobertura privada. Esse tipo de mecanismo, sem dúvida, ajuda a reduzir a regressividade do acesso aos serviços de saúde no Brasil,

mas é ainda bastante tímido, considerando-se outras experiências institucionais. A interação entre os setores público e privado nacionais se distingue dos modelos adotados na maior parte dos países, pelo fato de no Brasil grande parte dos serviços financiados publicamente serem ofertados por rede privada contratada. Na maior parte dos países em que coexistem os sistemas público e privado, ou o acesso aos serviços privados se limita a um pacote de serviços médicos específicos, ou existe uma segmentação no acesso: aos indivíduos que optam por ter saúde suplementar não é permitido o acesso aos serviços públicos. Tanto um desenho como outro podem ser interessantes no sentido de reduzir as iniquidades. No primeiro caso, o sistema público pode se especializar na oferta de serviços médicos básicos para toda a população, e fica a cargo do setor de saúde suplementar ofertar outros serviços especializados e apresentar para os consumidores maior possibilidade de escolha. No segundo desenho, os indivíduos são obrigados a optar por um sistema ou outro, o que inclusive é interessante para garantir a qualidade dos serviços médicos nos dois setores.

Um segundo aspecto relevante da regulamentação, aspecto que também perpassa a discussão da relação entre os setores público e privado no Brasil, diz respeito à padronização do desenho de benefícios proposta na regulamentação. Como já apresentado, a regulamentação propõe alguns tipos de contratos específicos e estabelece o rol de procedimentos mínimos obrigatoriamente contidos em cada um deles. Esse tipo de regulação possui vantagens e desvantagens. Entre os aspectos favoráveis, essa padronização de contratos reduz o problema de informação existente entre o consumidor e a seguradora/operadora em relação ao contrato de serviços no momento em que este é vendido. Na maior parte das vezes o consumidor/paciente não tem capacidade de avaliar se o rol de procedimentos propostos no contrato são suficientes para solucionar os possíveis problemas de saúde que ele venha a sofrer. Além disso, em muitos casos, não existe consistência intertemporal nas escolhas dos consumidores: *ex post* a realização da incerteza, estando doente, o consumidor prefere ter acesso a um grupo de serviços mais amplo. Uma prática importante e complementar ao mecanismo de padronização do desenho de benefícios é a criação de um sistema de acreditação dos serviços médicos e das operadoras de planos e seguros de saúde. O sistema de acreditação é fundamental para reduzir a assimetria informacional acerca da qualidade do serviço ofertado e para garantir um patamar mínimo de qualidade para toda a população. Esse sistema de acreditação ainda não foi implementado pela ANS. A desvantagem da padronização é que ela reduz as possibilidades de desenhos de contratos a serem ofertados para os consumidores/pacientes, de forma que o paciente não pode ter um contrato ajustado totalmente ao seu risco específico esperado. Esse não-ajustamento pode

determinar níveis de cobertura subótimas<sup>12</sup> e discriminação dos planos através da qualidade dos serviços ofertados. O contrato ofertado por uma seguradora/operadora apresenta três dimensões: o rol de procedimentos, a qualidade do serviço ofertado (rede contratada) e a quantidade de serviços disponibilizados. A padronização dos tipos de planos proposta na regulamentação estabelece parâmetros mínimos para o rol de procedimentos e para a quantidade, mas a disposição de regras sobre a qualidade do serviço ofertada é mais limitada. Nesse sentido, um mecanismo de discriminação factível para as seguradoras é oferecer contratos para grupos de risco específicos que se distinguem pela qualidade dos serviços. Em algumas situações a alocação final de contratos ofertada pode ser pior para os consumidores do que na ausência de regulação.

Um terceiro aspecto que merece destaque diz respeito à regra de precificação. A regulamentação permite uma discriminação de preços em que se considera unicamente o critério idade. Atualmente existem dez grupos de faixas etárias, de modo que indivíduos pertencentes a cada grupo etário pagam o mesmo prêmio mensal.<sup>13</sup> Além dos grupos etários especificados, a regra de precificação impõe ainda uma razão entre os prêmios cobrados entre os grupos extremos. Esses dispositivos têm dois problemas fundamentais: *a)* esses grupos etários só podem ser especificados se, de alguma forma, refletirem os grupos de risco; e *b)* a razão dos prêmios só pode ser proposta se também refletir o risco entre os grupos extremos. Caso essas duas premissas não se verifiquem, o resultado em termos de alocação tende a ser pior do que seria na ausência de regulamentação. Na verdade, o objetivo implícito da ANS ao propor esses mecanismos de preços é implementar um sistema de subsídio cruzado entre grupos de risco diferente na população, em particular entre a população mais jovem e a população idosa. Infelizmente, a evidência empírica e a teoria econômica já evidenciaram que a tentativa de implementar subsídio cruzado através da oferta privada de seguro saúde não é factível. Se o governo brasileiro pretende subsidiar os indivíduos de risco mais elevado, sejam os idosos, ou mesmo os grupos com endemias específicas, deve fazê-lo mediante um mecanismo direto de taxação da renda dos mais ricos e com menor risco de saúde e transferência de recursos para os mais pobres e de maior risco da saúde. Esse tipo de mecanismo de taxação, entretanto, também não é viável. A questão óbvia que se coloca na discussão de nossa regulamentação é: em que medida essa divisão de grupos etários é suficiente ou capaz de representar os grupos de risco existentes na população brasileira?

---

12. Além disso, vale ressaltar, com a padronização do rol de procedimentos a ANS se torna responsável pela determinação de introdução de novas tecnologias no setor.

13. São dez grupos etários: 0-18, 19-23, 24-28, 29-33, 34-38, 39-43, 44-48, 49-53, 54-58, 59 e mais. Esses grupos foram alterados recentemente de modo a compatibilizar com o Estatuto do Idoso.

Alguns estudos já mostram que os grupos escolhidos não retratam grupos de risco homogêneo, mesmo quando se considera apenas o critério idade [Ribeiro (2005)]. Esse é um aspecto fundamental da regulamentação que deve ser averiguado com mais cuidado pela ANS. Os impactos desses dispositivos de especificação podem ser maiores ou menores, dependendo da forma de adesão ao plano. No caso de adesão familiar, é mais factível um membro da família subsidiar o outro; entretanto, no caso de adesão individual, não é de se esperar que um jovem hoje aceite pagar mais porque acredita com isso gerar um subsídio para si próprio quando estiver idoso. De imediato, é de se esperar que ocorra evasão de indivíduos mais jovens dos planos, ou seja, a regulamentação vai acirrar a possibilidade de seleção adversa no setor de saúde suplementar. Além disso, espera-se que ocorra um nível de cobertura subótima. Dependendo da estrutura de mercado vigente, a presença de seleção adversa pode ainda determinar redução da qualidade de alguns serviços ofertados, uma vez que este pode ser o mecanismo de discriminação que as seguradoras implementam.

Diversos países utilizam o ajustamento ao risco considerando também, além da variável idade, a variável sexo. Sexo e idade são as duas variáveis observáveis mais usuais para sistemas de ajustamento ao risco [Van de Ven (2000)]. Uma alteração interessante seria a desagregação de cada grupo etário em dois tipos: planos para homens e mulheres. Essa desagregação é importante principalmente para o intervalo da idade em que a mulher se encontra no período reprodutivo. Atualmente, o único critério presente na regulamentação que considera de forma diferente homens e mulheres é a opção pelo plano obstétrico.

Um quarto aspecto da regulamentação que merece ser destacado diz respeito aos dispositivos introduzidos para garantir e controlar a sustentabilidade financeira das seguradoras/operadoras. Essa regulamentação é fundamental, uma vez que os consumidores/pacientes não têm nenhum mecanismo de controle e informação da saúde financeira da empresa quando escolhem o plano ao qual irão aderir.

Por último, cabe ainda ressaltar que a regulamentação não tem nenhum dispositivo que impede a introdução nos contratos de práticas de divisão de custos entre as seguradoras e pacientes, com o objetivo de minorar o risco moral e também de práticas de cuidado gerenciado. Já existem alguns trabalhos evidenciando a presença de risco moral no sistema suplementar brasileiro [Maia (2004)].

#### **4A.6 CONSIDERAÇÕES FINAIS**

Neste trabalho discutimos de forma bastante sucinta alguns possíveis impactos da regulamentação brasileira. Esse tipo de contribuição é importante, haja vista o

papel do setor de saúde suplementar no Brasil. Sem dúvida, a expansão da cobertura privada é uma forma de minorar as responsabilidades do setor público. O setor suplementar no Brasil teve um grande crescimento nos anos 1980, mas nos anos 1990 apresentou expansão praticamente nula. Esse congelamento da parcela da população coberta é certamente um aspecto importante que tem de ser analisado pelos nossos reguladores. O trabalho mostra que parece existirem espaços para a expansão da cobertura privada no Brasil. A análise dos possíveis impactos da regulamentação pode ser bem melhor realizada a partir do acesso aos bancos de dados da própria ANS. Nesse caso, é possível cruzar informações de utilização dos pacientes com informações acerca dos provedores e do tipo de contrato de seguro do paciente. Infelizmente, até o momento da finalização deste trabalho, essas informações não estavam disponíveis. Fica a sugestão de organização de um banco de dados e sua disponibilização para trabalhos acadêmicos que possam fornecer insumos importantes para auxiliar na gestão dos planos e seguros de saúde.

## 4B. A SOBREUTILIZAÇÃO DOS SERVIÇOS

### 4B.1 INTRODUÇÃO

O setor suplementar de saúde brasileiro constitui uma importante e significativa parcela do mercado de assistência médica no Brasil. Entretanto, somente em 1998, com a criação da Lei 9.656, o Estado passa a regulamentar esse setor, e a expansão da assistência médica suplementar se dá, principalmente, entre os anos de 1987 e 1994.

Diversos estudos têm analisado o comportamento dos indivíduos quanto à utilização dos serviços de saúde diante de diferentes formas de financiamento desses serviços [Newhouse (1996) e Manning *et alii* (1987)]. Em geral, se observa que quanto maior a cobertura do plano ou seguro de saúde, maior é o incentivo para os indivíduos utilizarem com maior freqüência os cuidados médicos, resultando, dessa forma, em um problema de risco moral.

A evidência mostra que o risco moral é um problema importante no mercado de saúde e este tem sido um tema recorrente na literatura internacional [Newhouse (1996), Sapelli e Vial (2003) e Cameron *et alii* (1988)]. A introdução de mecanismos que minimizem as distorções resultam em ganhos de bem-estar para a população, daí a importância de se conhecer a magnitude desses efeitos.

Ainda são poucos os trabalhos que investigam as inter-relações entre os setores de saúde público e privado no Brasil, bem como a atuação da assistência suplementar no mercado de bens e serviços de saúde. O objetivo deste trabalho é estimar

a sobreutilização dos serviços de saúde no sistema suplementar *vis-à-vis* o sistema público de saúde. Foram realizadas três estimativas: na primeira, consideramos todos os indivíduos que possuem algum tipo de plano ou seguro de saúde; na segunda e na terceira, separamos os indivíduos com plano entre os que possuem plano corporativo (adesão via trabalho) e não-corporativo (adesão individual). Essa desagregação da população com plano de saúde nesses dois grupos, adesão via trabalho e adesão individual, é importante na medida em que os incentivos à entrada no plano podem ser bastante distintos, influenciando o comportamento do consumidor e, por conseguinte, a sobreutilização.

A sobreutilização é entendida como a utilização excessiva dos serviços observada entre usuários do sistema suplementar de saúde e usuários do sistema público. Optamos por denominar sobreutilização essa diferença, já que não é uma medida exata do que se denomina risco moral na literatura. A diferença poderia ser interpretada como medida do risco moral, conforme definido na teoria econômica, se supusermos que os custos envolvidos na utilização do sistema público correspondem aos custos monetários com que os indivíduos se defrontariam na ausência de um sistema de seguros. Por outro lado, essa diferença poderia refletir também as dificuldades de acesso ao serviço público se admitirmos que o padrão ótimo de utilização é o observado no setor privado. Provavelmente, a diferença estimada se deve tanto a um problema de risco moral, quanto a diferenças no acesso entre os dois setores.<sup>14</sup>

A estimativa da magnitude da sobreutilização é realizada através de uma análise contrafactual, em que a diferença entre a utilização média estimada dos indivíduos que possuem plano menos a utilização média estimada para esses mesmos indivíduos, caso não possuissem plano privado, nos fornece o valor esperado da sobreutilização no sistema de saúde suplementar.

Utilizamos os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), de 1998, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A pesquisa fornece uma ampla variedade de informações individuais e domiciliares. Em 1998, foi a campo um suplemento que indaga a respeito das características de saúde dos moradores.

As categorias de cuidado médico selecionadas foram: *a)* número de consultas realizadas no último ano; *b)* número de dias na última internação realizada nos últimos 12 meses; e *c)* número de internações no último ano.

14. A possível diferença no acesso aos serviços públicos e privados pode também ser interpretada como uma diferença na qualidade do serviço provido em cada sistema, na medida em que o acesso é uma dimensão da qualidade do serviço. As demais dimensões da qualidade do serviço médico são supostas iguais nos dois sistemas.

Os principais resultados encontrados apontam a presença de sobreutilização no sistema de saúde suplementar para os três tipos de medida de utilização dos serviços médicos. Na primeira estimativa, a sobreutilização de consultas médicas foi estimada em 0,78 consulta *per capita* por ano, representando cerca de 24% da utilização estimada de consultas. No cuidado hospitalar, a sobreutilização estimada de dias internado foi de 0,12 dia *per capita* por ano, enquanto no número de internações, a sobreutilização foi de 0,03 internação por ano por pessoa. Esses valores representam, respectivamente, 22% e 30% da utilização estimada de cada cuidado hospitalar.

Na segunda e terceira estimativas da sobreutilização separamos os indivíduos possuidores de plano segundo o tipo de adesão: plano corporativo e adesão individual. Os resultados também apontaram para a presença de sobreutilização. Para planos não-corporativos, a sobreutilização estimada foi de 0,96 consulta *per capita* por ano (28% da utilização média estimada), 0,13 dia de internação *per capita* por ano (25%) e 0,03 internação por ano por indivíduo (30%). Para planos corporativos estimamos: 0,63 consulta *per capita* por ano (22% da utilização média estimada), 0,14 dia de internação *per capita* por ano (25%) e 0,03 internação por ano por indivíduo (30%).

A interpretação desses resultados como uma medida do risco moral sugere que os contratos no sistema suplementar podem não estar desenhados adequadamente. A sociedade pode estar incorrendo em perdas de bem-estar, uma vez que os indivíduos com cobertura privada escolhem mais cuidado do que escolheriam na ausência de seguro e as seguradoras, por sua vez, antecipam esse comportamento elevando o prêmio. Em consequência, uma parcela menor da população tem acesso à cobertura privada. Uma mudança institucional importante envolveria a incorporação de mecanismos de divisão de custos com consumidores ou até mesmo a adoção de práticas de cuidado gerenciado.

Além desta introdução, o trabalho está dividido em mais seis seções. Na Seção 4B.2, caracterizamos o mercado de bens e serviços de saúde. Uma breve revisão da literatura empírica internacional e nacional é apresentada na Seção 4B.3. A Seção 4B.4 resume os principais aspectos do sistema de saúde suplementar brasileiro. A abordagem empírica é apresentada na Seção 4B.5. O banco de dados e a análise descritiva estão presentes na Seção 4B.6. Por fim, a Seção 4B.7 apresenta os resultados e a conclusão.

## **4B.2 CARACTERIZAÇÃO DO MERCADO DE BENS E SERVIÇOS DE SAÚDE**

O mercado de bens e serviços de saúde é bastante particular, quando comparado aos demais. Podemos ressaltar três aspectos importantes: o ambiente de incerteza,

a presença de informação assimétrica e o ambiente de mercado no qual os bens e serviços de saúde são consumidos.

Os bens e serviços de saúde são, em sua maioria, consumidos em um ambiente de incerteza. A incerteza está presente em dois momentos da decisão de consumo individual. Primeiro, existe incerteza quanto à necessidade do uso dos serviços médicos, uma vez que em sua maior parte esse uso é decorrente de choques estocásticos. Segundo, existe incerteza quanto ao diagnóstico clínico e quanto à eficácia do tratamento. Além disso, a distribuição das despesas médicas tanto na população quanto ao longo do ciclo de vida dos indivíduos não é normal, assumindo valores muito altos para poucos episódios.

A assimetria de informações está presente na relação entre os três agentes envolvidos no estabelecimento de um contrato de seguro de saúde: o consumidor, a seguradora e o provedor dos bens e serviços de saúde, e pode resultar em problemas de agência. Os consumidores e a seguradora correspondem à demanda de mercado, podendo ser, estes últimos, público ou privado; enquanto provedores correspondem à oferta de mercado, podendo também ser público ou privado.

A relação do segurador seja com o provedor, seja com o consumidor, apresenta problemas de risco moral. Com o provedor o problema ocorre na medida em que este pode induzir uma sobreutilização dos serviços, já que possui maior quantidade de informação que os financiadores. A sobreutilização pode ocorrer tanto para elevar os rendimentos diretos dos provedores, quanto para reduzir a incerteza do diagnóstico, e com isso obter também ganhos de reputação. Entre o consumidor e o segurador, o risco moral pode ocorrer na presença de seguro pleno. Nesse caso, os consumidores tendem a sobreutilizar os serviços, já que o custo marginal do produto demandado é 0. Um outro problema, decorrente da assimetria de informações nesse mercado, é a seleção adversa de consumidores, que pode ser entendida como a seleção, por parte da seguradora, de um grupo específico de indivíduos que difere em determinados atributos de risco da média observada na população total.<sup>15</sup>

O ambiente de mercado nos quais os bens e serviços de saúde são consumidos apresenta, pelo menos, quatro características que o diferem dos demais: *a)* caracterização de diversos bens e serviços de saúde como bens credenciais, ou seja, que necessitam da certificação pública de um profissional especializado, sendo esta

---

15. Tradicionalmente, a literatura emprega o termo seleção adversa quando ocorre uma seleção desfavorável de consumidores. Uma outra possibilidade é a ocorrência do *cream-skimming*, que consiste numa seleção adversa contrária, ou seja, em favor do grupo de indivíduos considerado.

reconhecida e confiável para consumidores; *b*) dissociação entre consumidor final e agente responsável pela indicação terapêutica, que pode resultar em problemas de agência; *c*) presença de produtos e equipamentos com elevados gastos (pesquisa e desenvolvimento de novos processos e produtos); *d*) por fim, presença de externalidades difusas.<sup>16</sup> Uma outra característica importante é a consideração de alguns bens e serviços de saúde como *meritórios*, isto é, bens e serviços a que todo cidadão deve ter acesso, sendo responsabilidade da política pública a garantia de acesso universal [Andrade e Lisboa (2001)].

#### **4B.3 REVISÃO DA LITERATURA**

Nesta seção faremos uma breve revisão da literatura que investiga a presença do risco moral e a sobreutilização nos serviços de saúde.

A evidência empírica já apontou, em diversos estudos, que o risco moral ou a sobreutilização dos serviços de saúde é um dos principais problemas no mercado de saúde. O risco moral pode gerar perdas de bem-estar, na medida em que parte da população não estaria apta a adquirir um plano ou seguro de saúde devido ao aumento dos custos, com a conseqüente elevação dos preços, envolvendo, dessa forma, um *trade-off* fundamental entre divisão de riscos e incentivos adequados. O aumento do grau de cobertura do seguro implica uma divisão de riscos mais ampla, mas também leva ao aumento de perda de bem-estar, porque os indivíduos escolhem mais cuidado (risco moral) e os provedores ofertam mais cuidado (indução de demanda) [Cutler e Zeckhauser (2000)].

Uma das principais referências na literatura foi um estudo coordenado por Newhouse (1996). A base do trabalho é o RAND Health Insurance Experiment, um experimento longitudinal que, entre os vários objetivos, analisa o comportamento do indivíduo diante de diferentes arranjos de seguro de saúde. Os resultados mostram que a demanda por serviços de saúde responde negativamente a variações no preço, ou seja, diferentes mecanismos de incentivos financeiros resultam em decisões diferentes entre os consumidores, com menor utilização diante de despesas financeiras mais elevadas, diretamente desembolsadas por consumidores. Manning *et alii* (1987), usando os mesmos dados de painel (RAND Health Insurance Experiment) estimam a elasticidade de preço da utilização do serviço médico a partir de uma mudança na taxa de copagamento. Os autores encontram valores entre -0,1 e -0,2, dependendo do tipo de cuidado considerado e da variação na taxa de co-pagamento.

---

16. Um bem gera externalidade difusa quando o consumo médio da sociedade tem um impacto direto no bem-estar de cada indivíduo.

A diversidade dos sistemas de saúde que determina a inserção do setor privado de forma diferenciada entre os países e a não-padronização dos bancos de dados de saúde têm resultado em formas particulares de mensurar ou apenas detectar a presença de risco moral. Em geral, a estimativa parte da comparação da magnitude da utilização estimada dos serviços de saúde entre dois grupos: público e privado, privado com e sem mecanismo de divisão de custos, diferentes segmentos do setor privado etc.

Uma dificuldade associada aos trabalhos empíricos que analisam o problema de risco moral é a presença da endogeneidade entre a escolha de adquirir plano de saúde e a utilização dos serviços de saúde. Essa endogeneidade se deve ao fato de a decisão de adquirir um plano ser realizada com base na expectativa de utilização dos serviços. Dessa forma, características específicas dos indivíduos que adquirem plano podem estar influenciando sua utilização, resultando em um problema de viés de seleção amostral.

Cameron *et alii* (1988) são os pioneiros no desenvolvimento de um modelo que interage a demanda por seguro de saúde e a utilização do cuidado de saúde. A implementação empírica do trabalho, a partir de dados *cross-section* do *Australian Health Survey* de 1977-1978, consiste em utilizar variáveis instrumentais da escolha do seguro de saúde como variável dependente do modelo de utilização do cuidado de saúde. São consideradas sete variáveis de utilização do cuidado de saúde.<sup>17</sup> É estimado um modelo para a probabilidade de ter um determinado tipo de seguro (quatro tipos de seguro de saúde são considerados), em que o valor predito da probabilidade é a variável instrumental da escolha de ter seguro e é incluída no modelo de utilização. Os autores constatam que o estado de saúde parece afetar mais a utilização do cuidado que a demanda pelo seguro, enquanto a renda é mais importante na determinação da escolha do seguro que na utilização do cuidado de saúde. Além disso, a utilização do serviço parece ser maior para aqueles que possuem uma cobertura mais ampla, resultado que indica a presença tanto de risco moral quanto de auto-seleção.

Sapelli e Vial (2003) estimam a presença de sobreutilização e seleção adversa no sistema de saúde chileno. São analisadas duas variáveis de utilização do cuidado médico, o número de consultas médicas e o número de dias internado. Para mensurar a presença de sobreutilização compara-se a utilização do serviço de saúde na presença do seguro com a utilização caso o indivíduo não houvesse adquirido

---

17. Número de consultas nas duas últimas semanas, número de internações nos últimos 12 meses, número de noites internado na mais recente internação, número de consultas com um profissional de saúde que não médico nas últimas quatro semanas, número de medicamentos prescritos e não prescritos, número de medicamentos prescritos, número de medicamento não-prescritos usados nos últimos dois dias.

o seguro. A sobreutilização é encontrada não-significativa para dias internado, mas é encontrada para consultas médicas.

A utilização de bases longitudinais é uma forma alternativa para avaliar a presença de risco moral no uso dos serviços de saúde que solucionariam o problema da endogeneidade. Essas bases acompanham o mesmo indivíduo ao longo do tempo, identificando o impacto das mudanças nas características individuais sobre a utilização. Além disso, é possível verificar o momento de adesão ao plano e controlar melhor o estado de saúde do indivíduo. Essas informações são importantes na medida em que a primeira permite modelar o problema com base na decisão intertemporal de adquirir o plano e utilizar os serviços de saúde. Enquanto a segunda permite controlar o estado de saúde após a utilização, ou seja, é possível verificar o efeito da utilização sobre o restabelecimento do estado de saúde.

Utilizando dados longitudinais durante dois anos, na França, Chiappori, Durand e Geoffard (1998) avaliam a presença de risco moral na utilização do cuidado médico. O trabalho analisa uma mudança na estrutura de incentivos dos contratos entre seguradores e pacientes, que consiste na mudança na taxa de copagamento de utilização dos serviços. A análise considera apenas a decisão de realizar ao menos uma consulta, para três tipos de cuidado: consultas realizadas no consultório, consultas realizadas em casa e consultas no consultório de um médico especialista. Os principais resultados indicam a presença de risco moral para consultas em casa, mas não para consultas no consultório, sugerindo a presença de risco moral na demanda pelo cuidado médico quando custos não-monetários também são importantes. Ou seja, custos como o de deslocamento e de tempo podem representar uma importante fração do custo total considerado pelo consumidor, tendendo a diminuir o impacto da mudança de preços sobre a utilização.

#### **4B.4 CARACTERIZAÇÃO DO SISTEMA DE SAÚDE SUPLEMENTAR BRASILEIRO**

O sistema de saúde brasileiro é caracterizado como um sistema misto, em que os setores público e privado atuam no provimento e no financiamento dos bens e serviços de saúde. A Constituição de 1988 garante acesso aos serviços de saúde como um direito universal e igualitário de todos os brasileiros, mas permite a coexistência, paralelamente ao sistema público, de um sistema de saúde suplementar. Em princípio, cabe ao setor público a cobertura de toda a população residente, embora se estime que, aproximadamente, 25% da população possuam algum tipo de plano de saúde privado [PNAD (1998)], caracterizando a assistência médica suplementar.

A participação do setor privado se dá de forma bastante generalizada no sistema de saúde brasileiro. Além de um sistema suplementar, uma parte dos serviços

financiados pelo setor público é ofertada por instituições privadas (75% dos leitos, 83% das clínicas médico-ambulatoriais e 40% das unidades de complementação diagnóstica e terapêutica) [Viacava e Bahia (1996) citados por Reis (2000)].

As estimativas do gasto nacional com saúde para o ano de 1998 indicam 7,8 como percentagem do PIB, em que 3,2 correspondem ao gasto público e 4,6 ao gasto privado.<sup>18</sup>

A Tabela 4B.1 sistematiza o gasto total (público e privado) com saúde e o gasto público com saúde (ambos como percentual do PIB) em países selecionados.

A importância social e econômica do setor de saúde traz à luz a discussão sobre quais os arranjos institucionais mais eficientes na alocação dos recursos e que melhor atendam aos indivíduos de diferentes classes de risco. Nesse sentido, o monitoramento do gasto em saúde, independentemente de ser o financiamento público ou privado, é crucial e está sempre acompanhado da análise sobre o padrão de utilização da população.

Podemos distinguir quatro modalidades de atuação da assistência médica suplementar no Brasil: a medicina de grupo, as cooperativas médicas, os planos próprios das empresas (autogestão) e as seguradoras. Elas se distinguem quanto à forma de contratação dos serviços de saúde.

A medicina de grupo constitui empresas médicas que administram planos de saúde para empresas, indivíduos e famílias, com atendimento baseado, principalmente, numa rede credenciada de serviços. Nos planos destinados a pessoas jurídicas, a adesão dos empregados é automática em caso de custos financiados

**TABELA 4B.1  
GASTO COM SAÚDE COMO % DO PIB E GASTO PÚBLICO COMO % DO GASTO TOTAL EM PAÍSES SELECIONADOS**

País	Gasto total 1995 % do PIB	Gasto público % do gasto total — 1995	Gasto total 2001 % do PIB	Gasto público % do gasto total — 2001
Brasil	7,4	43,24	7,6	42,10
Canadá	8,9	69,66	9,5	71,57
Estados Unidos	13,0	45,38	13,9	44,60
Inglaterra	6,8	80,88	7,6	82,89

Fonte: Elaboração própria a partir de dados do Banco Mundial.

18. Fonte: MS/SIS/DPI/Siops, IBGE/PNAD de 1998 e Banco Central (Bacen). Elaboração de Sérgio F. Piola – Curso de especialização em políticas de recursos humanos para a gestão do SUS.

integralmente pelo empregador, e em caso de divisão de custos com o empregado, a participação é opcional. As cooperativas são entidades em que os médicos são sócios e prestadores de serviço simultaneamente, com pagamento proporcional à quantidade de serviços prestados. As Unimed representam quase a totalidade desse segmento.

Os planos próprios das empresas constituem planos fechados, restritos apenas a funcionários e seus dependentes, em geral de uma única empresa, responsável pela administração direta do plano (autogestão), podendo ainda dispor de uma rede credenciada. Algumas empresas financiam integralmente o plano básico para seus funcionários, embora na maioria dos planos próprios das empresas exista a participação do empregado, que pode, também, optar por adicionais, como planos complementares e diferentes formas de acesso: livre-escolha e reembolso. Por fim, as seguradoras, que originalmente no Brasil funcionavam oferecendo unicamente o seguro-saúde por reembolso, atualmente podem, também, lançar mão de rede credenciada.

A grande expansão da assistência médica suplementar no Brasil se deu entre 1987 e 1994, quando houve crescimento de 73,4% da população coberta, que passou de 24,4 milhões para 42,3 milhões [Andrade e Lisboa (2001)]. Embora o setor suplementar de saúde abranja uma parcela expressiva da população brasileira, com evidente importância no provimento dos serviços de saúde, só recentemente, com a regulamentação do setor de medicina suplementar, a relação público e privado passa a ser concebida no desenho institucional do sistema de saúde brasileiro.

Apesar de a regulamentação do setor de saúde suplementar ter sido instituída em 1998, somente a partir de meados de 2000 suas diretrizes passam efetivamente a ser cumpridas. Logo, os efeitos da regulamentação são ainda obscuros: restam diversas questões a serem entendidas na relação entre os setores público e privado.

A regulamentação do setor de saúde suplementar não apresenta nenhuma restrição ao uso de mecanismos de divisão de custos (por exemplo, o co-pagamento) e também ao cuidado gerenciado.<sup>19</sup> Esses mecanismos são usualmente implementados para racionalizar o uso dos serviços médicos, de modo a evitar a sobreutilização decorrente do comportamento tanto dos provedores como dos pacientes. Evidências empíricas mostram que a elasticidade de demanda é sensível à alteração dos preços dos serviços médicos, e os valores encontrados variam de -0,14 a -1,5 dependendo do tipo de cuidado considerado e do método de estimativa [Cutler e Zeckhauser (2000)].

---

19. O termo cuidado gerenciado se refere, de uma forma genérica, às práticas adotadas para administrar os serviços prestados pelos provedores contratados pelos planos e seguradoras de saúde.

No sistema brasileiro, esses mecanismos ainda são pouco difundidos no setor privado e inexistentes no setor público. Embora, na prática, o sistema público restrinja a utilização ao impor, por exemplo, a presença de filas de espera para atendimento, institucionalmente esses mecanismos não estão implementados. Segundo dados da PNAD de 1998, apenas 28% dos indivíduos com plano de saúde pagam, além da mensalidade, algum valor pelos atendimentos a que têm direito, ou seja, uma parcela ainda pequena da população com plano está sujeita a algum mecanismo de divisão de custos.<sup>20</sup>

No caso de sistemas públicos, a introdução desses mecanismos financeiros é pouco usual, e quando existe, incide apenas sobre alguns grupos de serviços. A dificuldade de implementação de mecanismos de divisão de custos em sistemas públicos, ou seja, racionalização através de custos monetários impostos aos pacientes, está na questão distributiva. Os impactos desses mecanismos são distintos, dependendo do nível de renda individual. Esses sistemas em geral utilizam mecanismos de racionalização indiretos/não-monetários, por exemplo, a fila de espera.

No Brasil, existem ainda poucos estudos que investigam empiricamente a presença do risco moral no sistema de saúde. Um trabalho com dados da PNAD de 1998 foi desenvolvido por Stancioli e Zylberstajn (2002) onde os resultados apontam a presença de risco moral. A escassez desses trabalhos se deve, provavelmente, à restrita disponibilidade de bases que contemplam dados de utilização do serviço de saúde associados a características individuais, como a cobertura por algum plano de saúde, características socioeconômicas, de estado de saúde, entre outras.

#### **4B.5 MODELO EMPÍRICO**

A abordagem empírica deste trabalho consiste em uma análise contrafactual<sup>21</sup> na qual estimamos a utilização dos serviços de saúde para diferentes segmentos do sistema de saúde brasileiro, considerando o *status* de cobertura de plano de saúde dos indivíduos.

A utilização dos serviços de saúde é mensurada através de três indicadores: número de consultas médicas nos últimos 12 meses, número de dias na última internação realizada nos últimos 12 meses e número de internações no último ano. Esses indicadores representam diferentes tipos de cuidado de saúde e são

---

20. As perguntas da PNAD são insuficientes para permitir uma comparação entre os grupos de indivíduos com plano de saúde sujeitos a mecanismos de divisão de custos e os indivíduos com plano de saúde sem mecanismos de divisão de custos. O nível de detalhamento não informa sobre quais tipos de serviços o co-pagamento incide e de que forma.

21. Para a validade desse exercício, supomos que não existem variáveis relevantes omitidas.

decorrentes de diferentes processos de tomada de decisão. No caso de consultas médicas, a decisão de consumir o serviço é, na maior parte das vezes, realizada pelo próprio indivíduo, enquanto, no caso da internação, a decisão é usualmente tomada pelos médicos.

Entendemos como sobreutilização dos serviços de saúde a utilização excessiva que ocorre devido à presença do seguro. Em um plano de saúde com seguro pleno ou quase pleno, os indivíduos tendem a sobreutilizar os serviços, dado que não arcam com o custo marginal total de cada procedimento, ou seja, no ato da realização do serviço médico, em geral, o indivíduo tem custo monetário marginal igual ou próximo a 0.<sup>22</sup> Além do custo monetário, os indivíduos, ao receberem cuidado médico, incorrem também em custos decorrentes do tempo de espera para a realização do serviço, em custo de oportunidade de procurar o serviço e em custos de deslocamento, os quais alteram a freqüência de utilização dos serviços médicos.

Foram realizadas três estimativas de sobreutilização para cada indicador. Na primeira, o objetivo é obter a sobreutilização dos cuidados de saúde comparando os indivíduos possuidores de plano à situação caso esses mesmos indivíduos não possuíssem plano. Separamos, dessa forma, a amostra entre os indivíduos que possuem plano de saúde e os que não possuem. O modelo é estimado em duas etapas. Na primeira etapa é usado um modelo *probit* para explicar a decisão de compra do plano de saúde, e na segunda etapa, um modelo de contagem a fim de explicar a utilização dos serviços.

Na segunda e terceira estimativas separamos o banco de dados em três subamostras: indivíduos sem plano de saúde, com plano de saúde corporativo e com plano de saúde não-corporativo. O objetivo é obter a sobreutilização dos cuidados de saúde comparando a utilização daqueles que possuem plano corporativo como se não possuíssem plano nenhum e o dos que possuem plano não-corporativo como se não possuíssem plano.

#### **4B.5.1 1<sup>a</sup> estimativa: sobreutilização dos serviços médicos no sistema suplementar**

Para estimar a utilização dos serviços médicos, aplicamos modelos de dados de contagem. Essa escolha se justifica na medida em que a utilização é mensurada através do número de visitas médicas, número de dias internado e número de internações, os quais consistem de valores não-negativos e inteiros.

---

22. No caso de co-pagamento, existe um custo monetário marginal positivo.

O modelo de Poisson é o modelo básico no arcabouço dos modelos de dados de contagem. Ele pressupõe que a variável dependente, dado o vetor de covariadas, possui uma distribuição de Poisson e que a densidade da variável dependente é completamente determinada pela média condicional. Essa hipótese impõe algumas restrições que, quando violadas, podem tornar inconsistentes os parâmetros estimados. A mais importante delas é a eqüidispersão, onde a média e a variância condicional são iguais (essa suposição deriva da hipótese de que o parâmetro de intensidade do modelo é determinístico). Essa restrição torna o modelo inadequado em muitas situações reais, que em geral tendem a apresentar sobredispersão.

Um modelo alternativo ao modelo de Poisson é o modelo Binomial Negativo, em que a distribuição permite mais flexibilidade na modelagem da variância. O modelo Binomial Negativo é o modelo paramétrico-padrão para o caso de sobredispersão dos dados, ou seja, o caso em que a variância condicional excede a média.

A fim de verificar se os dados são sobredispersos, realizamos o teste da razão de verossimilhança. O resultado do teste indicou a sobredispersão dos dados.<sup>23</sup> Assim, a segunda etapa do modelo é estimada por um modelo binomial negativo.

O modelo é estimado em duas etapas. Na primeira é usado um modelo *probit* para explicar a decisão de compra do plano de saúde. E, na segunda etapa, um modelo de contagem, a fim de explicar a utilização dos serviços. Ou seja:

$$\text{Prob}(D_{Si} = 1) = \text{Prob}(Z_i\alpha > \mu_i) \quad (1)$$

$$E[y_i | D_{Si}, \varepsilon_i] = \exp(X_i\beta + \varepsilon_i) \quad (2)$$

onde:

$i = 1, 2, 3, \dots$  indivíduos;

$D_s$  = variável binária igual a 1 se o indivíduo possui plano de saúde, e igual a 0 se não possui plano;

$\alpha_K$  = parâmetros estimados para as  $K$  variáveis;

$\beta_M$  = parâmetros estimados para as  $M$  variáveis;

---

23. Para uma referência mais completa do teste, ver Cameron e Trivedi (1998).

$y_i$  = utilização do serviço de atenção médica. O mesmo modelo é estimado para as três variáveis dependentes separadamente;

$\mu_i$  = erro aleatório; e

$Z$  e  $X$  = vetores de características individuais.

A equação (1) corresponde à probabilidade de ter um plano de saúde. A equação (2) corresponde ao modelo de utilização, que é estimado separadamente para a subamostra com e sem plano de saúde. A segunda etapa do modelo tem como variável explicativa o valor predito da probabilidade. Esse procedimento procura controlar a presença do viés de seleção.

Como os dados contêm informações sobre os indivíduos de uma mesma família, utilizamos a correção intrafamiliar nas duas etapas do modelo.

A estimativa da sobreutilização consiste em uma análise contrafactual, em que o valor esperado é a diferença entre *a*) o número esperado de serviços consumidos pelo indivíduo  $i$  quando possui um plano de saúde e *b*) o número esperado de serviços que o indivíduo  $i$  poderia consumir, caso ele não possuísse plano de saúde.

Temos, portanto, que:

$$\text{Sobreutilização} = E(y_s/x, D_s=1) - E(y_s/x, D_s=0)$$

onde:

$y_s$  é a utilização da subamostra com plano de saúde;

$D_s=1$  são indivíduos que possuem algum plano de saúde (sistema suplementar);

$D_s=0$  são indivíduos que não possuem nenhum tipo de plano de saúde; e

$x$  é vetor de atributos dos indivíduos com plano de saúde.

Em outras palavras, a estimativa da sobreutilização corresponde à diferença entre o valor esperado da utilização dos indivíduos com plano de saúde, calculado a partir do produto interno do vetor de coeficientes estimado para a amostra da população com plano de saúde e o vetor de características desses indivíduos, e o valor esperado da utilização dos indivíduos com plano de saúde caso não possuíssem plano, calculado a partir do produto interno do vetor de coeficientes estimado para a amostra da população sem plano de saúde e o vetor de características dos indivíduos com plano.

No caso de um valor positivo para a estimativa de sobreutilização, o valor esperado da utilização é maior para aqueles que adquirem um plano, do que para os mesmos indivíduos, caso não tivessem comprado um plano de saúde.

#### **4B.5.2 2<sup>a</sup> estimativa: sobreutilização dos serviços médicos no setor de saúde suplementar para os planos de adesão individual**

Na segunda e terceira estimativas separamos os indivíduos que possuem plano entre aqueles que possuem plano corporativo (adesão via trabalho) e os que possuem plano não-corporativo (adesão individual). Para estimar a sobreutilização, considerando-se apenas os indivíduos que possuem plano individual, o modelo econométrico é similar ao utilizado na 1<sup>a</sup> estimativa. A diferença consiste apenas no universo amostral dos indivíduos que têm plano, universo que se restringe aos que possuem plano por adesão individual.

#### **4B.5.3 3<sup>a</sup> estimativa: sobreutilização dos serviços médicos no setor de saúde suplementar para os planos corporativos**

Nessa estimativa o modelo é composto apenas da equação de utilização dos serviços médicos. Nesse caso não estimamos a equação de probabilidade uma vez que os indivíduos com plano corporativo não estão diretamente envolvidos na decisão de ter plano. Além disso, esperamos que o viés de seleção não seja importante na medida em que em geral as empresas são compostas de diferentes trabalhadores, e em praticamente todos os setores existem planos corporativos.

### **4B.6 BANCO DE DADOS E ANÁLISE DESCRIPTIVA**

Nesta seção, descrevemos o banco de dados usado e as principais variáveis incluídas na estimação.

#### **4B.6.1 Banco de dados**

A base de dados utilizada é a PNAD do ano de 1998, realizada pelo IBGE. A pesquisa possui um suplemento com informações a respeito das características de saúde dos moradores.

A seguir descrevemos o efeito esperado das variáveis do modelo de utilização. O quadro, em anexo, explica como as variáveis foram incluídas na primeira e na segunda etapas da estimação.

Variáveis explicativas usadas nas equações de utilização:

*a) Renda familiar per capita.* A variável renda pode apresentar dois efeitos: o primeiro se refere a um aumento na utilização na presença de seguro, devido ao baixo custo marginal de utilização, que dá ao indivíduo a possibilidade de aumentar o consumo de bens e serviços de saúde. E o segundo, na direção contrária, que aumenta o custo de oportunidade (custo de tempo) de utilização dos serviços.

*b) Grupos de idade.* Consideramos nove grupos de idade: de 0 a 2 anos, de 3 a 5, de 6 a 10, de 11 a 25, de 26 a 40, de 41 a 50, de 51 a 60, de 61 a 70 e de 71 ou mais anos. O grupo de referência é de 6 a 10 anos de idade. Essa segmentação procura contemplar diferentes grupos de risco. A divisão dos grupos etários procurou contemplar padrões de utilização homogêneos.

*c) Gênero.* A variável sexo busca controlar a diferença de necessidade de utilização do serviço de saúde entre homens e mulheres. Espera-se que mulheres utilizem mais o serviço de saúde que homens.

*d) Área de residência.* As *dummies* para área de residência (região metropolitana, urbana e rural) têm como objetivo captar diferentes densidades populacionais. Espera-se que áreas de baixa densidade tenham efeito negativo tanto na utilização dentro do segmento dos planos de saúde, quanto do segmento da população coberta apenas pelo SUS. O efeito está associado ao custo de tempo de viajar para utilizar os serviços e à reduzida oferta de provedores nessas áreas.

*e) Estados.* As *dummies* de estado [unidades da federação (UF)] também têm como objetivo controlar o diferencial de acesso ao provimento dos serviços de saúde e captar dinâmicas estruturais distintas, como mercado de trabalho, acesso a meios de transporte etc. Além disso, tem o objetivo de captar os diferenciais de oferta dos serviços de saúde (provedores) e controlar o comportamento atinente aos provedores entre os diferentes estados.<sup>24</sup>

*f) Anos de estudo.* Quanto maior a escolaridade, maior deve ser a utilização do serviço de saúde, uma vez que os indivíduos mais escolarizados têm maior capacidade de valorizar o cuidado recebido. O grau de informação tende a impactar principalmente a utilização preventiva. Desse modo, o efeito esperado é diferente para os modelos de consultas e de internação. A escolaridade usada é o máximo de anos de estudo entre a educação do chefe de família e do indivíduo.

*g) Número de filhos por faixas etárias.* O controle do número de filhos por faixa etária pode alterar a utilização dos serviços de saúde, na medida em que o tempo disponível para realizar um atendimento de saúde (custo de oportunidade) diminui com o aumento do número de filhos.

*h) Saúde auto-avaliada.* A pergunta sobre saúde auto-avaliada refere-se a uma medida do estado de saúde auto-avaliado. São cinco categorias de resposta: muito

---

24. Apesar de reconhecermos que o provedor tem um papel importante na freqüência de utilização de cuidados médicos, não existem no questionário da PNAD informações sobre os tipos de planos e contratos a que os provedores estão sujeitos. A *dummy* de estado tenta captar muito mais as diferenças de oferta pública e privada existentes em cada região. Dentro de uma mesma região existem provedores com diferentes estruturas de incentivo. Apesar dessa fragilidade, isso não invalida a importância do exercício realizado uma vez que não é o objetivo do trabalho decompor a sobreutilização em componentes decorrentes do comportamento do provedor e do paciente.

bom, bom, regular, ruim ou muito ruim. As respostas foram agregadas e transformadas numa *dummy*. É esperado que, quanto pior for esse indicador de saúde, maior será a utilização dos serviços de saúde.

*i) Presença de doença crônica.* As *dummies* que indicam a presença de determinada doença crônica devem aumentar a utilização do serviço de saúde (fator de risco).

*j) Acesso.* A variável de acesso foi construída com o objetivo de se tentar controlar o diferencial de acesso aos serviços de saúde entre a população com plano e a população com cobertura apenas do sistema público de saúde.

*l) Qualidade.* A *proxy* de controle da qualidade refere-se a uma medida sobre como o indivíduo considera o atendimento de saúde recebido. São cinco categorias de resposta: muito bom, bom, regular, ruim ou muito ruim. Essa pergunta refere-se apenas àqueles indivíduos que receberam algum atendimento de saúde no último ano. Agrupamos as categorias de resposta em uma variável categórica na qual os indivíduos que não respondem à questão porque não receberam atendimento no último ano correspondem a 0; aqueles que reportam o atendimento recebido como regular, ruim e muito ruim, correspondem a 1; e os que reportam muito bom e bom, correspondem a 2.

*m) Probabilidade de ter um plano de saúde.* Quanto maior a probabilidade de possuir um plano de saúde, maior deve ser a utilização dos serviços de saúde. Essa variável aparece na primeira e na segunda estimativas.

#### 4B.6.2 Análise descritiva

A amostra com plano corresponde a 24,5% do total, sendo 84.419 observações. Os indivíduos com plano representam 260.452 observações. Esses dados estão sistematizados na Tabela 4B.2.

TABELA 4B.2  
PROPORÇÃO DA AMOSTRA SEGUNDO COBERTURA DE PLANO DE SAÚDE

	Com plano	Sem plano	Total
Número de observações	84.419	260.452	344.871
Proporção (%)	24,48	75,52	100

Fonte: Elaboração própria a partir de dados da PNAD de 1998.

Entre os indivíduos com plano temos que cerca de 33% possuem plano não-corporativo contra 67% de planos corporativos.<sup>25</sup>

A Tabela 4B.3 mostra a média de consultas entre indivíduos com plano de saúde, sem plano, corporativos e não-corporativos e total da população. O número médio de consultas é cerca de duas vezes maior entre indivíduos com plano em relação aos que não o possuem.<sup>26</sup> É interessante observar que o desvio-padrão em relação à média é mais que o dobro na amostra sem plano, enquanto na com plano essa relação é cerca de uma vez e meia. Ou seja, os dados indicam maior dispersão em relação à média da freqüência de consultas médicas entre os que não possuem plano do que entre os que possuem. Outra observação interessante: indivíduos com planos não-corporativos (de adesão individual) têm maior média de consultas do que os que possuem plano corporativo. Isso indica, talvez, um possível fenômeno de seleção para planos caracterizados como individuais.

A média de dias de internação é maior entre os indivíduos com plano que entre os que não o possuem, sendo essa diferença estatisticamente não-significativa.<sup>27</sup> Observamos também que os indivíduos com plano de saúde corporativo apresentam maior média de dias de internação.

A Tabela 4B.5 indica a média de internações segundo a cobertura por plano de saúde e planos corporativos e não-corporativos. O teste de médias mostra que as médias entre indivíduos com e sem plano também não são diferentes ao nível de significância de 1%.

A seguir, sistematizamos a média das variáveis de utilização do cuidado segundo o sexo por cobertura de plano de saúde. O primeiro aspecto que chama a atenção é que, na média, a utilização, dentro do mesmo grupo de cobertura de

**TABELA 4B.3  
NÚMERO MÉDIO DE CONSULTAS NO ÚLTIMO ANO**

	Com plano	Sem plano	Planos corporativos	Planos não-corporativos	Total
Média	3,18	1,80	2,85	3,32	2,14
Desvio-padrão	4,85	3,72	4,52	4,95	4,07

Fonte: Elaboração própria a partir de dados da PNAD de 1998.

25. A separação entre planos corporativos e não-corporativos é feita a partir de duas perguntas da PNAD: se esse plano ao qual se tem direito é um plano público: se sim, o plano é caracterizado como corporativo. A questão seguinte é sobre quem paga esse plano a que se tem direito. As respostas: o indivíduo através do emprego, o empregador ou o empregado. Todas caracterizam o plano de adesão via trabalho. As demais respostas caracterizam o plano de adesão individual.

26. O teste de médias revela que as médias são diferentes ao nível de significância de 1%.

27. O teste de médias revela que as médias são iguais ao nível de significância de 1%.

**TABELA 4B.4  
NÚMERO MÉDIO DE DIAS INTERNADO NA ÚLTIMA INTERNAÇÃO DO ÚLTIMO ANO**

	Com plano	Sem plano	Planos corporativos	Planos não-corporativos	Total
Média	0,45	0,43	0,47	0,42	0,44
Desvio-padrão	3,97	3,89	3,87	3,98	3,91

Fonte: Elaboração própria a partir de dados da PNAD de 1998.

**TABELA 4B.5  
NÚMERO MÉDIO DE INTERNAÇÕES NOS ÚLTIMOS 12 MESES**

	Com plano	Sem plano	Planos corporativos	Planos não-corporativos	Total
Média	0,10	0,09	0,10	0,10	0,09
Desvio-padrão	0,44	0,45	0,42	0,43	0,45

Fonte: Elaboração própria a partir de dados da PNAD de 1998.

plano, é sempre maior entre as mulheres do que entre os homens. Esse comportamento está associado, provavelmente, a diferenças no grau de risco entre mulheres e homens. Se mulheres têm uma propensão maior a apresentar problemas de saúde, sua utilização deve ser mais elevada do que entre os homens.

O segundo aspecto importante é que a utilização do cuidado hospitalar é maior entre as mulheres sem plano em relação aos homens com plano. Enquanto para o cuidado ambulatorial (consultas médicas) homens com plano utilizam mais. Ainda analisando-se o cuidado hospitalar e ambulatorial, observamos que a diferença na utilização entre homens com e sem plano é relativamente maior para consultas que nos dois tipos de cuidado hospitalar, comportamento similar se repete entre as mulheres. Como descrito anteriormente, a média não controlada por sexo do cuidado hospitalar entre indivíduos com plano e sem é estatisticamente igual. Esse padrão pode ser explicado pelas diferenças entre os dois tipos de cuidado: no hospitalar, a decisão sobre a permanência (dias de internação) e sobre

**TABELA 4B.6  
MÉDIA DE CONSULTAS, DIAS DE INTERNAÇÃO E INTERNAÇÕES SEGUNDO O SEXO E COBERTURA POR PLANO DE SAÚDE**

	Consultas		Dias de internação		Número de internações	
	Com plano	Sem plano	Com plano	Sem plano	Com plano	Sem plano
Mulheres	3,75	2,25	0,48	0,45	0,12	0,11
Homens	2,51	1,35	0,41	0,41	0,08	0,07

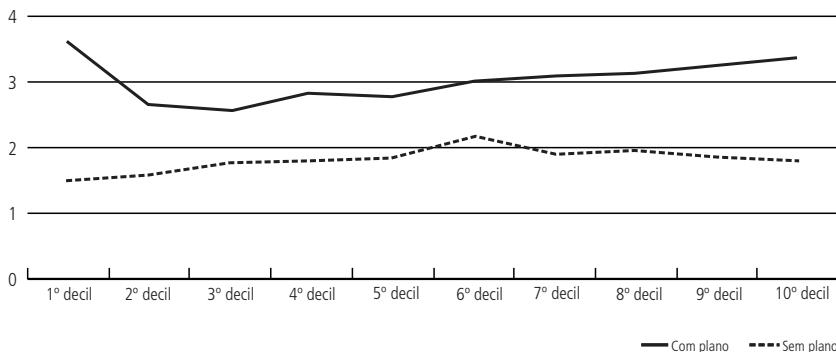
Fonte: Elaboração própria a partir de dados da PNAD de 1998.

a necessidade de utilização (internação) é usualmente tomada pelos médicos. Enquanto a decisão de consultar está muito mais relacionada à decisão do paciente.

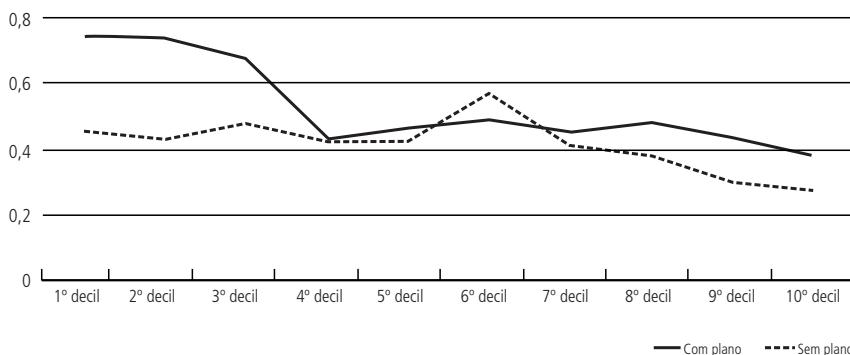
No Gráfico 4B.1 temos a média de consultas segundo decis de renda. Ao longo de todos os decis a quantidade de consultas é maior entre aqueles que têm plano de saúde. Chama a atenção a elevada utilização no primeiro decil com plano. Do segundo ao décimo decil da amostra com plano existe uma tendência de aumento na utilização, conforme aumenta a renda. Na amostra de indivíduos sem plano, a média de consultas tende a ser relativamente mais constante entre os decis de renda.

A média de dias de internação por decis de renda está ilustrada no Gráfico 4B.2. Observamos uma tendência de redução na média de dias de internação na amostra

**GRÁFICO 4B.1  
NÚMERO MÉDIO DE CONSULTAS SEGUNDO DECIS DE RENDA**



**GRÁFICO 4B.2  
NÚMERO MÉDIO DE DIAS INTERNADO POR DECIS DE RENDA**



com plano à medida que aumenta a renda. Exceto pelo sexto decil, a utilização é sempre maior entre indivíduos com plano de saúde. Na amostra sem plano observamos um padrão relativamente constante entre o primeiro e o quinto decil, com um aumento da utilização no sexto e uma tendência de decréscimo a partir daí.

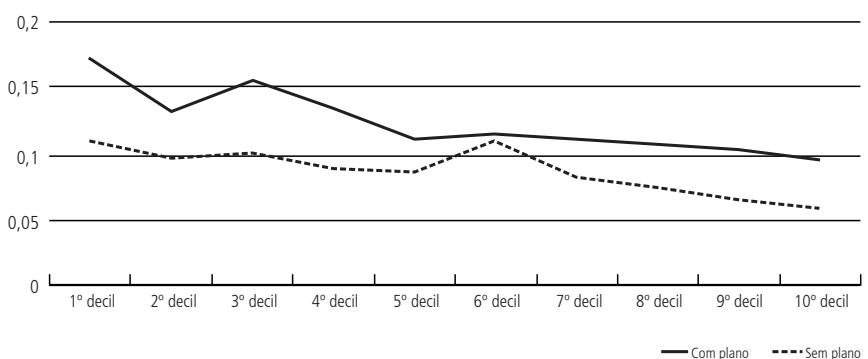
O Gráfico 4B.3 mostra a média de vezes de internação no último ano por decil de renda. Ao longo dos decis essa média é maior entre indivíduos com plano, com uma tendência de decréscimo à medida que aumenta a renda.

Um aspecto comum às três variáveis de utilização é a média mais elevada no primeiro decil de renda da amostra de indivíduos com plano de saúde. Quando consideramos o tipo de cuidado, também observamos que em consultas existe tendência de aumento na utilização com os decis de renda, enquanto no cuidado hospitalar parece haver decréscimo da utilização com a renda.

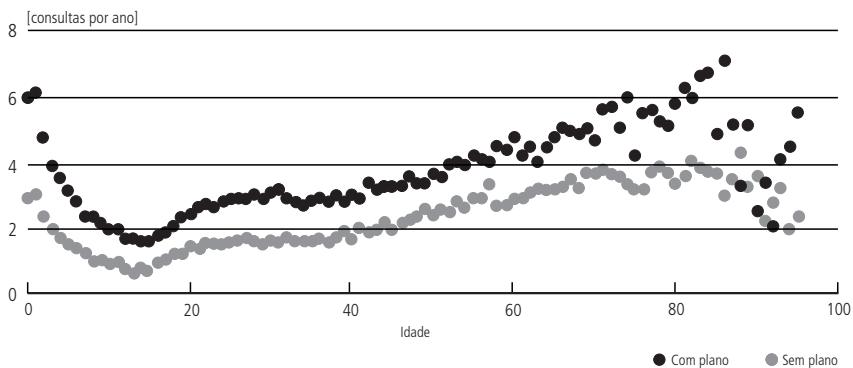
Nos Gráficos 4B.4, 4B.5 e 4B.6 temos a média da utilização por idade na amostra de indivíduos com e sem plano de saúde. O primeiro aspecto que chama a atenção é que não existe um padrão de utilização depois de, aproximadamente, 70 anos. Os dados mostram que essa utilização tanto pode ser muito elevada, quanto muito baixa, sem apresentar nenhuma tendência. Por um lado, esse comportamento pode ser explicado pelo reduzido número de observações nas faixas etárias mais elevadas.

O segundo aspecto interessante é que as diferenças na utilização do cuidado hospitalar entre indivíduos com e sem plano, e que estejam na mesma idade, são muito pequenas. Ao passo que, no cuidado ambulatorial, indivíduos com plano utilizam mais consultas em todas as idades até cerca de 85 anos. Mais uma vez, a

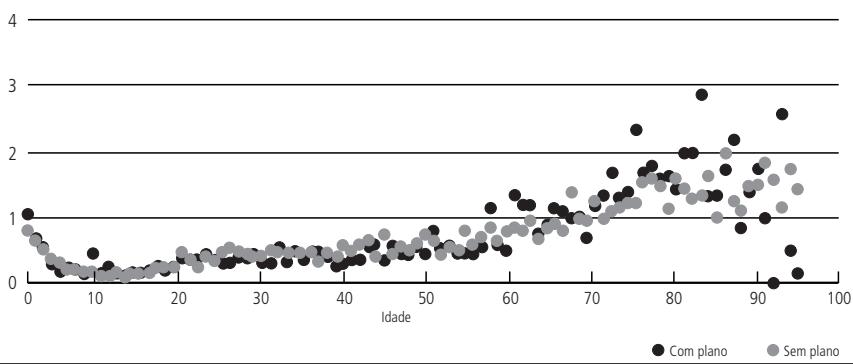
**GRÁFICO 4B.3  
QUANTIDADE DE INTERNAÇÕES SEGUNDO DECIS DE RENDA**



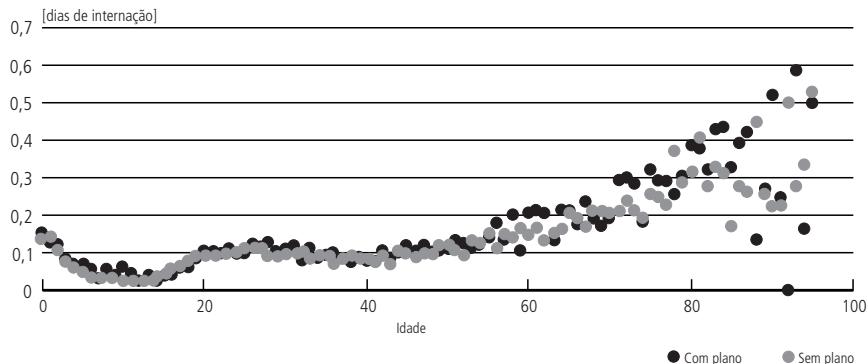
**GRÁFICO 4B.4**  
**MÉDIA DE CONSULTAS POR IDADE**



**GRÁFICO 4B.5**  
**MÉDIA DE DIAS INTERNADO POR IDADE**



**GRÁFICO 4B.6**  
**MÉDIA DE INTERNAÇÕES POR IDADE**



explicação para esse comportamento está na diferença da natureza da atenção entre os dois tipos de cuidado.

#### **4B.7 RESULTADOS**

Nesta seção apresentamos os resultados. Como explicado no modelo empírico, estimamos a sobreutilização dos cuidados de saúde em três momentos, a partir de uma análise contrafactual:

1º resultado:

Nesse resultado, a sobreutilização estimada é a diferença entre a média dos valores estimados da utilização dos indivíduos com plano e a média da utilização para esses mesmos indivíduos como se eles não tivessem plano de saúde.<sup>28</sup> Ou seja, a diferença entre:

$$E(y_s/x, D_S = 1) - E(y_s/x, D_S = 0)$$

Obtivemos que:

Sobreutilização de consultas:  $3,24 - 2,46 = 0,78$  consulta *per capita* por ano.

Sobreutilização de dias de internação:  $0,54 - 0,42 = 0,12$  dia de internação *per capita* na última internação do ano.

Sobreutilização de internações:  $0,10 - 0,07 = 0,03$  internação *per capita* por ano.<sup>29</sup>

---

#### **SOBREUTILIZAÇÃO COMO % DA UTILIZAÇÃO MÉDIA ESTIMADA**

---

Consultas médicas (0,78 como % de 3,24)	24%
Dias de internação (0,12 como % de 0,54)	22%
Número de internações (0,03 como % de 0,10)	30%

---

28. Para obter o contrafactual plotamos os coeficientes estimados para o grupo sem plano de saúde nas características do grupo de indivíduos com plano de saúde.

29. Embora o modelo que estima o número de dias internado e o número de internações apresente diversos coeficientes não-significativos, sobretudo para as *dummies* de estado da federação, os valores estimados da sobreutilização são significativos. Ou seja, o modelo sugere que existe também sobreutilização desse tipo de cuidado quando estimamos um modelo de demanda controlando pelas variáveis de risco individual. Infelizmente não se dispõe nem no banco de dados da PNAD, nem em outros bancos com registros de utilização existentes no Brasil, de informações atinentes aos provedores e aos seus contratos, de modo a possibilitar que a sobreutilização desse tipo de cuidado fosse estimada através de um modelo de oferta de serviços que pode ser mais adequado, dado o processo de decisão de utilização desses serviços. O exercício realizado nesse trabalho tem a limitação de não controlar para as variáveis de oferta que seriam fundamentais no caso dos serviços de internação, uma vez que a decisão de internar depende em grande medida do provedor e não do paciente. Entretanto, como estimamos um modelo de demanda em que a medida é a utilização realizada, entendemos que de alguma forma os valores de utilização refletem as características da oferta.

Os resultados revelam que existe sobreutilização nos serviços de saúde para os três tipos de cuidado.<sup>30</sup> Em termos percentuais, a sobreutilização de número de internações foi maior, embora os resultados sejam bastante parecidos.

### 2º e 3º resultados:

Esses resultados se referem a 2<sup>a</sup> e 3<sup>a</sup> estimativas, na qual estimamos a sobreutilização para indivíduos com plano corporativo e não-corporativo.

#### Planos não-corporativos

Sobreutilização de consultas:  $3,39 - 2,43 = 0,96$  consulta *per capita* por ano.

Sobreutilização de dias de internação:  $0,51 - 0,38 = 0,13$  dia de internação *per capita* na última internação do ano.

Sobreutilização de internações:  $0,10 - 0,07 = 0,03$  internação *per capita* por ano.

---

#### **SOBREUTILIZAÇÃO COMO % DA UTILIZAÇÃO MÉDIA ESTIMADA**

---

Consultas médicas (0,96 como % de 3,39)	28%
Dias de internação (0,13 como % de 0,51)	25%
Número de internações (0,03 como % de 0,10)	30%

---

É interessante observar que a sobreutilização estimada do cuidado de consultas médicas foi maior entre aqueles que possuem plano de adesão individual do que entre aqueles que possuem plano corporativo. Esse resultado se manteve para número de internações, mas não para dias de internação.

#### Planos corporativos

Sobreutilização de consultas:  $2,90 - 2,27 = 0,63$  consulta *per capita* por ano.

Sobreutilização de dias de internação:  $0,56 - 0,42 = 0,14$  dia de internação *per capita* na última internação do ano.

---

30. Para todos os valores de sobreutilização estimados efetuamos dois testes. O primeiro consiste de um teste de média simples que compara, a partir dos valores de sobreutilização estimados para cada indivíduo, se o valor esperado médio de consultas/internações/número de dias internado para os indivíduos que possuem planos é igual ao valor esperado médio de consultas/internações/número de dias internado, caso esses indivíduos não possuissem plano de saúde. O segundo teste consiste também de um teste de médias, sendo que o desvio-padrão do valor estimado no contrafactual foi reestimado utilizando-se a técnica do *bootstrap*. Esses testes foram realizados para todos os valores estimados no contrafactual. Para todos os testes as médias foram estatisticamente diferentes a 1%.

Sobreutilização de internações:  $0,10 - 0,07 = 0,03$  internação *per capita* por ano.

#### **SOBREUTILIZAÇÃO COMO % DA UTILIZAÇÃO MÉDIA ESTIMADA**

Consultas médicas (0,63 como % de 2,90)	22%
Dias de internação ( 0,14 como % de 0,56)	25%
Número de internações ( 0,03 como % de 0,10)	30%

A sobreutilização dos serviços pode refletir tanto a presença de problemas nos incentivos dos contratos no setor de saúde suplementar, quanto um problema de acesso ao setor público de saúde, ou mesmo as duas situações simultaneamente. A hipótese de que a estimativa da sobreutilização estaria mensurando totalmente a utilização excessiva dos dois cuidados é válida apenas se o pressuposto de igualdade de acesso entre os dois setores se mantenha, caso contrário, esse valor estaria sobreestimado e poderia refletir, em parte, uma dificuldade de acesso ao setor público de saúde.

Uma importante conclusão é que a eficiência dos contratos de plano ou seguro de saúde pode estar comprometida, gerando perdas de bem-estar, na medida em que os preços praticados estariam acima do nível ótimo. O nível ótimo pode ser entendido como o nível de preços no qual o número máximo de indivíduos estaria apto a adquirir o plano ou seguro de saúde sem que qualquer indivíduo incorra em perdas. Nesse caso, uma das implicações mais significativas é a não-cobertura pelo sistema suplementar de saúde de indivíduos que não adquirem plano devido a sua restrição orçamentária.

Uma das formas usuais de inibir a sobreutilização dos serviços é a implementação de instrumentos de contenção de demanda. No caso de um contrato privado, o meio mais tradicional de contornar o problema é a divisão de custos com os consumidores. Os instrumentos usuais são co-pagamentos, co-seguros e franquias. Os co-pagamentos e co-seguros constituem um valor específico ou percentual, cobrado pelo serviço prestado e pago pelo segurado. A franquia é um limite inferior de dispêndio até o qual o consumidor é responsável pelo pagamento integral dos serviços. Caso esse limite seja ultrapassado, o seguro passa a cobrir os gastos parcialmente ou integralmente, dependendo do tipo de contrato estabelecido. Os efeitos da imposição desses mecanismos já foram testados empiricamente e verificou-se que a demanda por serviços médicos é elástica em relação aos preços, embora a redução de procedimentos afete indivíduos de diferentes classes de risco de forma diferenciada. Pelo lado do provedor não existe um mecanismo ainda consolidado de redução do volume de procedimentos.

Um aumento de eficiência dos serviços financiados pelo sistema suplementar de saúde tenderia a impactar indiretamente o sistema público de saúde. Desde que houvesse queda nos preços, a ampliação da população com plano de saúde privado reduziria a parcela da população que utiliza exclusivamente o sistema público de saúde, com implicações importantes de política pública: *a)* fortalecimento do Estado no papel de regulamentação do setor; e *b)* possibilidade de alteração do foco dos gastos públicos em saúde, com a implementação de políticas específicas para indivíduos de baixa renda e/ou sujeitos a piores condições de saúde. Além disso, um aumento do número de indivíduos com cobertura por algum plano de saúde poderia diminuir o risco médio da população coberta, o que tende a alterar a estrutura e contratos no longo prazo, seja através de uma queda no nível de preços, seja pela ampliação do grau de cobertura do plano ou seguro-saúde.

Esse resultado é importante na medida em que o Brasil é um país em desenvolvimento, com restrições na sua capacidade de financiamento e provimento dos bens e serviços considerados meritórios. Além disso, a sustentabilidade de um sistema público de saúde universal e integral, conforme previsto na legislação do SUS, pode ser pouco viável, sobretudo se forem consideradas as inovações tecnológicas contínuas na área da saúde. O setor privado é certamente uma opção para viabilizar o acesso aos serviços de saúde para parte da população brasileira, e a atuação conjunta entre os setores público e privado está proposta desde a concepção do SUS em 1988. Estudos como esse, que de alguma forma se propõem a analisar a adequação dos contratos no setor privado, podem oferecer subsídios para políticas de gestão e possibilitar a extensão da cobertura.

## **4C. A SELEÇÃO ADVERSA NO SISTEMA**

### **4C.1 INTRODUÇÃO**

Atualmente, cerca de 38 milhões de indivíduos são beneficiários de algum tipo de plano ou seguro de saúde no Brasil, montante que corresponde aproximadamente a 21% da população residente.<sup>31</sup> O mecanismo de provimento e financiamento dos bens e serviços de saúde, seja no setor público ou no privado, está relacionado à discussão sobre quais as formas mais eficientes de garantir acesso a esses bens e serviços. Nesse sentido, os impactos decorrentes da estrutura de incentivos presentes no sistema suplementar refletem diretamente sobre a gestão dos recursos de saúde, tanto no sistema público, quanto no sistema privado de atenção à saúde.

---

31. Fonte: Agência Nacional de Saúde Suplementar (ANS) — dados referentes a outubro de 2003.

Dois problemas, decorrentes da assimetria de informações entre os agentes no estabelecimento de um contrato de seguro de saúde, seja público ou privado constituem uma preocupação crescente dos gestores de política de saúde: o risco moral e a seleção adversa. O primeiro pode ser caracterizado como a sobreutilização dos serviços após o estabelecimento do contrato de seguro de saúde. Enquanto o segundo, *ex ante* ao estabelecimento do contrato, consiste na seleção, por parte da seguradora, de indivíduos com risco maior que o risco médio da sociedade. O problema de seleção adversa decorre da impossibilidade de discriminação dos indivíduos segundo seu grau de risco e do fato de que a identidade do segurado afeta fortemente os custos. A seleção adversa pode gerar perdas de bem-estar, na medida em que os indivíduos podem não ter acesso aos planos tanto quanto gostariam de obter.

O objetivo deste trabalho é analisar a presença de seleção adversa de consumidores no mercado de assistência médica suplementar brasileiro. Entendemos como seleção adversa a auto-seleção, por parte dos próprios consumidores que compram plano de saúde compondo uma carteira de indivíduos com determinadas características que tornam o risco dessa carteira superior ao risco médio da população. Essas características podem ser atributos de risco observáveis ou não-observáveis. Apesar de utilizarmos o termo seleção adversa, não é possível testar empiricamente a seleção conforme proposto na teoria. Do ponto de vista teórico é necessário que haja alguma característica que a seguradora não possa discriminar. Até 1998 não existia regulamentação no mercado suplementar brasileiro, sendo de alguma forma possível para a seguradora estabelecer todas as formas de discriminação.

Para analisar a presença de seleção adversa no sistema de saúde suplementar, consideramos diferentes segmentações do sistema de saúde brasileiro. Primeiro distinguimos indivíduos com e sem plano de saúde. Uma segunda segmentação, dentro da população que possui plano de saúde, considera três subgrupos: indivíduos com plano de saúde privado adquirido através do empregador, indivíduos que possuem plano de saúde de adesão voluntária e aqueles que possuem plano de saúde adquirido por intermédio do vínculo trabalhista com o setor público sendo esse plano gerido por uma instituição pública. Nesse caso, a estimativa de seleção adversa é entre um subgrupo específico e os demais indivíduos que possuem plano.

A base de dados utilizada é a Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1998 realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Em 1998 foi a campo um suplemento inédito que indaga a respeito das características de saúde dos moradores.

Os dados utilizados se referem apenas aos chefes de família. A opção se justifica por serem estes os agentes diretamente envolvidos na decisão de adquirir plano de saúde. Para analisar a seleção não-observável, estimamos conjuntamente a utilização dos serviços de saúde e a probabilidade de ter plano (modelo de seleção), onde a correlação entre os erros do modelo de utilização e o de seleção indica a presença de seleção não-observável. Duas categorias de cuidado médico foram selecionadas: o número de consultas realizadas nos últimos 12 meses e o número de dias internado na última internação nos últimos 12 meses. Para avaliar a seleção com base em atributos de risco observáveis, analisamos os resultados do modelo estimado para a probabilidade de ter plano.

O trabalho está dividido em mais quatro seções, além desta introdução. A Seção 4C.2 caracteriza o mercado de bens e serviços de saúde ressaltando os aspectos da seleção adversa. A Seção 4C.3 faz uma breve revisão da literatura empírica. A Seção 4C.4 apresenta a metodologia de estimativa e a Seção 4C.5 apresenta os principais resultados.

## **4C.2 O MERCADO DE BENS E SERVIÇOS DE SAÚDE**

O consumo de bens e serviços de saúde apresenta características fundamentais que o tornam diferente do consumo de grande parte dos demais bens e serviços. Podemos destacar três dessas características: o ambiente de incerteza, a presença de informação assimétrica e o ambiente de mercado nos quais os bens e serviços de saúde são consumidos.

A seleção adversa de consumidores pode ser entendida como a seleção, por parte da seguradora, de uma carteira de indivíduos que difere em determinados atributos de risco da média observada da população total. A seleção adversa ocorre quando a seleção é desfavorável à seguradora, ou seja, quando o grupo de indivíduos que adere ao seguro apresenta determinados atributos de risco que aumentam a chance de utilização do seguro contratado. Outra possibilidade é a ocorrência do *cream-skimming*, que consiste em uma seleção favorável de consumidores. No caso de um seguro de saúde, é possível qualificar características como sendo, na média, características que alteram a utilização dos serviços de saúde. Entre as usuais, podemos citar idade, sexo e estado de saúde.

Existem dois sistemas usuais de tarifação no cálculo dos prêmios de risco das seguradoras. O primeiro considera o risco individual, ou seja, o cálculo do prêmio é realizado com base no gasto esperado de cada indivíduo. No segundo sistema, o prêmio é calculado com base no risco médio do grupo de indivíduos que compõem a carteira da seguradora.

O primeiro sistema de tarifação é denominado *experience rating* e tem como conseqüência que quanto maior o grau de risco individual, mais elevado é o prêmio do seguro. Uma das dificuldades de se adotar esse sistema reside em seus efeitos sobre a distribuição de renda. Os indivíduos com maior risco e, portanto, mais vulneráveis, seriam penalizados pagando prêmios mais elevados. No extremo, esse sistema pode até determinar a exclusão de grupos sociais menos favorecidos do mercado de serviços de saúde. A partir da potencialidade desses problemas, alguns governos, na tentativa de proteger grupos sociais excluídos do mercado privado de atenção à saúde, impedem a discriminação dos indivíduos por atributos de risco através da proposição de leis específicas de regulamentação do mercado de bens e serviços de saúde.

O segundo sistema de tarifação é denominado *community rating* e consiste em um sistema de pagamento de tarifa única com base no risco médio da população da carteira do seguro. Na prática, a implementação desse sistema só é possível se a compra do seguro for mandatária para indivíduos de uma mesma região geográfica ou se o seguro for administrado por algum empregador. Nesses casos está se forçando um subsídio cruzado entre os diversos grupos de risco.

No sistema suplementar de saúde brasileiro, as empresas de autogestão são um exemplo de modalidade de seguradora cuja precificação do prêmio de risco é feita com base no risco médio da carteira de indivíduos. O mesmo acontece quando a compra do plano é realizada por intermédio do empregador. Esses tipos de plano tendem a ter um preço mais baixo que o observado para os planos com adesão individual. Existem dois fatores associados a essa diferença no preço. Primeiro, uma queda nos custos de transação que é resultado da negociação direta com o empregador. Segundo, o risco médio da população da carteira da seguradora tende a se aproximar do risco médio observado para a população.

A possibilidade da seleção adversa de consumidores no mercado de planos e seguros de saúde privado tem impactos diretos sobre a regulamentação do mercado. Entre as formas usuais de evitar o problema de seleção adversa podemos citar a imposição de cláusulas contratuais como cobertura de doenças preexistentes e tempo de carência para tratamentos específicos.

Nesse contexto de mercado, são dois os argumentos a favor do financiamento público dos bens e serviços de saúde: o primeiro diz respeito aos aspectos distributivos do acesso aos bens e serviços e o segundo, à não-observância do problema de seleção adversa. Em sistemas públicos de financiamento universal dos serviços de saúde, o mercado funciona com apenas uma seguradora, nesse caso, o governo.

### 4C.3 REVISÃO DA LITERATURA

O objetivo desta seção é descrever alguns estudos teóricos e empíricos que discutem a seleção adversa de consumidores no mercado de seguros. A primeira subseção apresenta alguns estudos teóricos que analisam os efeitos da estrutura de contratos sobre o bem-estar social. Esses efeitos dependem, em grande parte, de como o mercado está organizado. A segunda parte resume alguns trabalhos empíricos sobre a seleção adversa no setor de saúde.

#### 4C.3.1 Aspectos teóricos

A seleção adversa é a seleção, por parte da seguradora, de um grupo de indivíduos com risco diferenciado do risco médio da população total. A seleção adversa ocorre porque, no estabelecimento de um contrato de seguro, os consumidores de alto risco têm incentivo a comprar mais cobertura que os consumidores de baixo risco, dentro do mesmo prêmio de risco médio. Entre os principais modelos teóricos que abordam a seleção adversa em vários mercados, podemos citar Akerlof (1970) e Rothschild e Stiglitz (1976).

O modelo seminal de equilíbrio com informação assimétrica foi desenvolvido por Rothschild e Stiglitz (1976). Em um mercado competitivo com dois tipos de consumidores: indivíduos de baixo risco e indivíduos de alto risco, onde consumidores conhecem melhor seu risco futuro que as seguradoras, ou se estas não podem usar toda a informação disponível para o cálculo do prêmio é possível apenas um equilíbrio,<sup>32</sup> o equilíbrio separador, onde os diferentes tipos compram diferentes contratos. O equilíbrio de *pooling*, com os dois tipos de consumidores comprando o mesmo contrato, não é possível nesse mercado. Esse resultado ocorre porque sempre haverá um contrato fora do *pooling* que é preferível para os indivíduos de baixo risco, dessa forma as condições de equilíbrio não estariam satisfeitas.

Cutler e Zeckhauser (2000) analisam os efeitos da seleção adversa partindo do modelo inicial de Rothschild e Stiglitz (1976) e propõem uma extensão com segmentações dos grupos de risco. São desenvolvidos três modelos.<sup>33</sup> O primeiro constitui o modelo básico inicial de equilíbrio com seleção. Suponha uma economia com dois tipos de indivíduos e dois tipos de planos. O plano generoso foi desenhado para indivíduos de alto risco e o plano moderado para os de baixo risco. Nessa economia, o equilíbrio de mercado pode ter, dependendo da proporção

32. O equilíbrio em um mercado de seguros competitivo é caracterizado pelo conjunto de contratos tal que os consumidores escolhem o contrato que maximiza sua utilidade esperada, nenhum contrato de equilíbrio tem lucro esperado negativo e não existe contrato fora do equilíbrio com lucro não-negativo.

33. A diferença do modelo de Rothschild e Stiglitz (1976) é que Cutler e Zeckhauser (2000) fixam o número de planos disponíveis no mercado. Essa hipótese, por sua vez, altera os resultados entre os modelos.

dos tipos de indivíduos, os dois tipos no plano do moderado (equilíbrio de *pooling*). Entretanto, como os indivíduos não pagam os custos de acordo com o seu próprio risco, e sim a média dos custos, esse equilíbrio não é eficiente.

Existem duas possibilidades de alocação mais eficiente sob essas condições que resultam em um equilíbrio separador: o subsídio cruzado e a manipulação de planos. O primeiro consiste numa taxação adicional sobre o prêmio do plano moderado que é usada para compensar o prêmio do plano generoso. Nesse caso, ambos os indivíduos estarão melhor com o subsídio do que sem ele, e esse resultado ocorre porque indivíduos de alto risco irão preferir adquirir o plano mais generoso saindo do *pool* de indivíduos de baixo risco que, por sua vez, terão seu prêmio reduzido com a saída dos de alto risco. A manipulação de planos<sup>34</sup> é o segundo mecanismo que pode induzir um equilíbrio separador numa economia sob essas condições. Suponha a mesma economia descrita anteriormente. Nesse contexto, um plano básico caracterizado como um plano menos abrangente que o plano moderado é também oferecido aos indivíduos. Se o plano básico e o plano generoso são os dois únicos tipos oferecidos nesse mercado, os indivíduos de baixo risco irão adquirir o plano básico, e os de alto risco, o plano generoso. Entretanto, mesmo que o plano moderado continue a ser oferecido no mercado, nenhum indivíduo de baixo risco irá adquiri-lo, uma vez que o benefício líquido privado obtido com a aquisição desse plano é inferior ao obtido com a aquisição do plano básico. A manipulação de planos é um mecanismo que reduz a eficiência, uma vez que diminui a possibilidade de divisão do risco.

O segundo modelo consiste no equilíbrio com indivíduos múltiplos em um grupo de risco. A hipótese é que existe um contínuo de tipos diferentes de indivíduos de baixo e alto risco, ou seja, existem vários consumidores com preferências similares sobre o seguro, e cada consumidor pertence a um grupo de baixo ou alto risco. Suponha que são oferecidos três tipos de plano: generoso, moderado e básico. Inicialmente, todos os indivíduos de baixo risco optam pelo plano moderado em vez do generoso (esse resultado deriva do modelo inicial de *pooling*). Os indivíduos do grupo de alto risco, por sua vez, se dividem na escolha entre o plano moderado e o plano generoso. Esse resultado ocorre porque, inicialmente, existe uma tendência de *pooling* no plano moderado, mas como os indivíduos de alto risco diferem em preferências, parte opta pelo plano generoso. O equilíbrio com todos os de baixo risco no plano moderado e parte dos de alto risco no plano generoso e parte no plano moderado é chamado de equilíbrio híbrido, e este equilíbrio é estável. Entretanto, como a identidade

34. A manipulação de planos consiste em um mecanismo no qual as seguradoras oferecem um rol de contratos que se diferenciam quanto à qualidade dos serviços. O desenho dos contratos faz com que os indivíduos revelem qual é o seu tipo, ou seja, qual o seu grau de risco.

dos consumidores do tipo de alto risco que optam pelo plano moderado afeta significativamente os custos do plano, os indivíduos de baixo risco adquirem o plano básico. Sob essa hipótese o equilíbrio é separador.

O terceiro modelo supõe que, em vez de apenas dois grupos de indivíduos, existe um contínuo de grupos de risco. São oferecidos dois tipos de planos: plano generoso e moderado. Como existe uma distribuição contínua de riscos na população, os indivíduos vão se dividir entre os dois planos em um primeiro momento. Essa escolha depende da expectativa de gastos individual. Os prêmios dos planos, por outro lado, dependem da identidade do segurado inscrito no plano. A diferença entre os prêmios dos dois planos pode ser decomposta em: diferença de custos média que o plano moderado oferece a sua carteira de consumidores e diferença no nível médio de ocorrência do evento (doença) entre os dois planos. O segundo termo é consequência da seleção adversa.

Na margem, as pessoas vão se deslocar do plano generoso para o plano moderado. Com a seleção adversa, o prêmio dos planos mais generosos aumenta mais que o do plano moderado, e esse resultado ocorre porque as despesas médicas são significativamente maiores quanto pior é o estado de saúde.

Dessa forma, pequenas variações nos preços podem resultar em grandes diferenças nas alocações, levando o plano generoso a desaparecer do mercado. Esse resultado é consequência de um processo dinâmico de seleção adversa chamado de “espiral”. Nessa situação, os de alto risco acabam adquirindo uma apólice menos generosa que o seu nível ótimo de cobertura, enquanto os de baixo risco continuam em sua apólice preferida.

Dois importantes resultados, a partir da análise teórica, são que os efeitos da seleção adversa podem ser perversos para os grupos de risco mais vulneráveis e podem resultar em impactos significativos sobre o bem-estar. No sistema de saúde brasileiro, ainda não conhecemos claramente os impactos da regulamentação do setor de saúde suplementar sobre a alocação dos indivíduos entre os diferentes tipos de contratos. Uma hipótese forte, mas ainda não testada empiricamente, é que a imposição de faixas etárias para ajuste do prêmio esteja gerando seleção adversa no setor suplementar na medida em que impede as seguradoras de discriminhar o consumidor segundo a idade. A discriminação nesse caso é dentro de um intervalo de idade.

#### **4C.3.2 Evidências empíricas**

A forma mais usual de analisar empiricamente a seleção adversa de consumidores no mercado de bens e serviços de saúde é através da utilização de bases

longitudinais. Essas bases acompanham o mesmo indivíduo em diferentes pontos do tempo, permitindo observar os efeitos de uma mudança na cobertura do seguro de saúde de acordo com as características de risco. No Brasil, a utilização desses dados em saúde é ainda quase inexistente devido à restrita disponibilidade de bases desse tipo. É importante chamar a atenção para o fato de que a ANS dispõe de informações sobre os indivíduos com plano de saúde. Essas informações possibilitam a montagem de uma base longitudinal, desde que é possível seguir o indivíduo durante um período de tempo e controlar para diversas características como grau de cobertura do plano, prêmio de risco, utilização dos serviços e atributos individuais.

Cutler e Zeckhauser (1997) investigam a presença de seleção advera utilizando dados longitudinais da cobertura por seguro de saúde dos empregados da Universidade de Harvard. Os trabalhos desenvolvidos a partir dessas informações constituem uma experiência importante na literatura, na medida em que foi possível analisar em diversos pontos do tempo o comportamento individual do consumidor diante de uma mudança na estrutura de incentivos para aquisição de diferentes tipos de planos de saúde. A análise dos dados aponta que a mudança de um sistema de subsídios generosos para um sistema de divisão de pagamento igual do prêmio de seguro de saúde entre empregados e empregadores resultou em uma significativa resposta por parte dos consumidores aos novos incentivos. Os empregados passaram a optar por planos menos generosos, o que levou a uma queda substancial dos prêmios para esses tipos de planos, enquanto os planos mais generosos tiveram aumentos sucessivos nos preços. Em três anos, os planos mais generosos deixaram de ser oferecidos nesse mercado, resultado que se explica pela presença de seleção advera nos planos com cobertura mais ampla: apenas indivíduos com maior taxa de risco permanecem no *pool* desses planos, enquanto os de baixo risco acabam transferindo suas apólices para planos menos generosos. Os resultados dessa análise reforçam o modelo teórico de seleção advera em espiral, explicado na seção anterior, no qual pequenas variações nos preços resultam em grandes alterações na alocação dos indivíduos entre os planos.

Utilizando os mesmos dados da reforma no mercado de seguro de saúde ocorrida na Universidade de Harvard, entre os anos de 1992 e 1996, Cutler e Reber (1998) estimam as perdas decorrentes do processo de seleção advera e os ganhos decorridos da competição nesse mercado. Esses dois processos constituem um *trade-off*: quando o tipo de consumidor afeta os custos, a competição entre as seguradoras pode ter resultados ambíguos. Permitir que os indivíduos escolham entre planos concorrentes pode alocá-los de forma apropriada e permitir, através de incentivos, que ocorra o provimento eficiente. Entretanto, a

competição traz problemas de seleção adversa, que induz os indivíduos que podem estar em um *pool* mais saudável a se associarem a planos menos generosos, e isso gera incentivos para que os planos distorçam suas ofertas de maneira a serem menos generosos com os doentes [Cutler e Zeckhauser (2000)].

Os principais resultados encontrados por Cutler e Reber (1998) mostram que a perda ocorrida devido à seleção adversa foi cerca de 2% a 4% da base de gastos inicial e, em contrapartida, a redução nos prêmios, decorrente da competição, foi de 4% a 8% da base de gastos inicial, resultando em um ganho líquido. O segundo resultado importante mostra que para a sociedade, o resultado líquido foi apenas de perda devida à seleção adversa, pois a redução nos lucros das seguradoras foi transferida para o empregador (no caso, Harvard), não existindo implicações de eficiência sobre essas transferências. Entretanto, é importante ressaltar que esse resultado foi verificado para um pequeno grupo de consumidores. No caso de um maior número de empregadores adotarem essa política, pode ser que a estratégia das seguradoras não seja apenas de redução do prêmio de risco com redução no lucro, como supostamente em Harvard, podendo ocorrer, por exemplo, manipulação de planos através da qualidade ou restrições à utilização.

Sapelli e Vial (2003) analisam a presença de auto-seleção no sistema de saúde do Chile. A base de dados utilizada é a Pesquisa Socioeconômica Nacional (CASEN Survey) de 1996. A seleção é mensurada por características de risco observáveis e não-observáveis. As características de risco são aquelas que se referem a um aumento na utilização do serviço de saúde. Para analisar a seleção não-observável (a seleção não-observável é sinalizada por uma correlação positiva entre a probabilidade de ter plano e a utilização dos serviços de saúde) são consideradas duas variáveis de utilização: consultas médicas e dias de hospitalização. A decisão de adquirir um plano de saúde é avaliada dentro de dois segmentos da população: trabalhadores dependentes, que optam entre adquirir plano público ou privado e trabalhadores independentes que decidem, voluntariamente, comprar ou não um seguro de saúde. O modelo é estimado por máxima verossimilhança com informação completa. Os principais resultados indicam a presença de seleção para atributos não-observáveis contra as seguradoras para trabalhadores independentes e contra o seguro público para dependentes. Esses resultados são observados no modelo que considera a utilização de dias de internação.

Como visto, devido à restrita disponibilidade de bases de dados, ainda são poucos os estudos empíricos brasileiros que analisam os incentivos entre os agentes no mercado de seguros de saúde, especificamente a seleção adversa.

Um estudo pioneiro foi desenvolvido por Alves (2003). O trabalho consiste em uma adaptação do teste para seleção advera proposto por Chiappori e Salanié (2000). O teste consiste na estimação de duas equações de probabilidade, em que a interdependência condicional entre os resíduos dos modelos indica a presença de seleção advera. A primeira equação é a probabilidade de escolha do contrato e, a segunda, a probabilidade de utilização.

A base de dados utilizada é a PNAD de 1998. Na equação de probabilidade de escolha, o indivíduo opta entre adquirir um plano de saúde do tipo amplo em contraposição a um plano do tipo mínimo<sup>35</sup> e na equação de probabilidade de utilização, ele escolhe entre utilizar ou não o serviço de atenção à saúde. A variável de utilização é construída a partir de um índice que incorpora diferentes dimensões de utilização do serviço. Assim, o autor estabelece um ponto que separa indivíduos que utilizaram o plano de saúde e os que não utilizaram. Os resultados do teste indicam ausência de seleção advera no mercado brasileiro de saúde suplementar.

#### **4C.4 METODOLOGIA DE ESTIMAÇÃO**

Nesta seção, apresentamos o modelo empírico, a base de dados utilizada e uma breve descrição das principais variáveis.

##### **4C.4.1 Modelo**

O valor observado dos serviços médicos utilizados pelos indivíduos resulta de duas decisões, a primeira refere-se à decisão de comprar ou não um plano de saúde e a segunda à decisão de freqüência de utilização de cuidados médicos.<sup>36</sup> A utilização dos serviços de saúde está sujeita aos problemas de agência presentes nesses processos de decisão, os quais podem ser caracterizados de duas formas: o problema de seleção advera presente na decisão de aquisição do plano de saúde e o problema de risco moral presente na decisão de freqüência de utilização. Entendemos como seleção advera a seleção de um grupo específico de indivíduos que difere em determinados atributos de risco<sup>37</sup> da média observada na população total. Quando a probabilidade de escolher um plano é maior para indivíduos com maiores taxas de risco baseada em características observáveis, dizemos que há seleção baseada em características observáveis.<sup>38</sup> Quando, porém, essa probabilidade é

35. O autor estabelece um ponto de corte para planos do tipo amplo e do tipo mínimo. Os primeiros se caracterizam como contratos que cobrem um maior número de procedimentos na prestação do serviço médico. A amostra inclui apenas indivíduos com plano de saúde.

36. O conjunto de variáveis de controle que afeta a decisão de comprar um plano e utilizar o cuidado de saúde são diferentes. Essa hipótese não invalida o modelo e é justificada pelo fato de que são diferentes processos de decisão.

37. Os atributos de risco são aqueles que aumentam a chance de utilização do serviço de saúde.

38. As características são observáveis para o pesquisador.

maior para aqueles com maiores taxas de risco não-observável, a seleção é com base em características não-observáveis.

Neste trabalho, consideramos distintos segmentos do sistema de saúde brasileiro, de modo a contemplar diferentes *status* de cobertura de plano de saúde. Uma primeira divisão importante no desenho do sistema brasileiro considera os indivíduos com plano de saúde e os sem plano de saúde. Uma segunda segmentação considera, dentro da população que possui plano de saúde, subgrupos, dependendo do tipo de adesão ao plano.

O primeiro subgrupo se refere aos indivíduos que possuem plano de saúde privado adquirido através do empregador (autogestões e contratações coletivas). Esse tipo de apólice possui uma característica particular em relação às demais, porque seu preço, geralmente, é taxado com base no risco médio da população coberta (*community rating*) e não com base em características próprias dos indivíduos (*experience rating*). Um segundo subgrupo é caracterizado por indivíduos que possuem plano de saúde de adesão voluntária, que consiste naqueles que compram o plano sem o intermédio do empregador. Em geral, esses planos são especificados com base em algumas características observáveis individuais, por exemplo, a idade. E, por fim, aqueles que possuem plano de saúde adquirido por intermédio do vínculo trabalhistico com o setor público e esse plano é gerido por uma instituição pública. O critério para classificação dos subgrupos segue a idéia de que diferentes dinâmicas de decisão para aquisição do plano de saúde estariam envolvidas em cada subgrupo.

Assim, o objetivo do trabalho é avaliar a presença de seleção adversa entre a população com plano e a população coberta apenas pelo SUS, e entre os três subgrupos definidos dentro da população que possui plano de saúde. Para analisar a presença de seleção adversa com base em características de risco não-observáveis, estimamos uma equação de utilização dos serviços considerando duas variáveis do cuidado de saúde: *a)* número de consultas médicas nos últimos 12 meses e *b)* número de dias internado na última internação nos últimos 12 meses.

Como a decisão de compra do plano de saúde e a utilização dos serviços são determinadas conjuntamente, a forma mais adequada de estimação utiliza a abordagem do método de equações simultâneas. Existem duas abordagens para estimá-las: os métodos de equação única com informação limitada e os métodos de sistemas de equações com informação completa. No primeiro método, cada equação no sistema de equações é estimada individualmente, considerando-se as restrições impostas sobre a equação, sem considerar as restrições impostas

sobre as outras equações. No segundo método, todas as equações são estimadas simultaneamente, considerando todas as restrições sobre tais equações.

Neste trabalho, utilizamos a estimação simultânea das equações por Máxima Verossimilhança com Informação Completa (FIML), segundo metodologia adotada em Greene (1997). Sapelli e Vial (2003) adotam o mesmo procedimento para estimar risco moral e seleção adversa no sistema de saúde do Chile.

A forma mais adequada de estimar a freqüência de utilização dos serviços de saúde utiliza modelos de contagem. Nesses modelos, a variável dependente consiste de valores inteiros e não-negativos. O modelo de Poisson que explica a quantidade de serviços consumido no segmento  $j$  do sistema tem a seguinte forma:

$$E[y_i | d_{ji}, \varepsilon_i] = \exp(\beta' x_i + \varepsilon_i) \quad (1)$$

onde,  $y_i$  é o número de serviços consumido pelo indivíduo  $i$  em uma unidade de tempo;  $d_{ji}$  é uma variável binária que assume valor 1 se o indivíduo pertence ao segmento  $j$  e 0 se não pertence;  $\varepsilon_i$  é o componente heterogêneo na equação de contagem; e  $x_i$  é o vetor de características individuais.

A equação de seleção, estimada por um modelo *probit* que analisa a probabilidade de participação em um segmento  $j$  de seguro, é caracterizada por:

$$d_{ji} = 1, \text{ se } \alpha' w_i + \mu_i > 0 \quad (2)$$

$$d_{ji} = 0, \text{ se não} \quad (3)$$

onde,  $w_i$  é um vetor de características que afeta a escolha de participação no segmento de seguro, e  $\mu_i$  é um termo randômico.

Seguindo Greene (1997),<sup>39</sup> temos que a função log-verossimilhança para o modelo completo é a densidade conjunta para os dados observados. Quando  $d_{ji}=1$ , ( $y_i, x_i, d_{ji}, w_i$ ) são observados. Estamos interessados em obter  $P[y_i, d_{ji}=1 | x_i, w_i]$ , onde  $P$  é a função com uma distribuição de Poisson ( $P$ ), condicionada a  $\varepsilon_i$ ,  $d_i$  e  $y_i$  serem independentes. Então:

$$P[y_i, d_{ji}=1 | x_i, w_i, \varepsilon_i] = P[y_i | x_i, \varepsilon_i] \text{Prob}[d_{ji}=1 | w_i, \varepsilon_i] \quad (4)$$

---

39. A operacionalização do modelo com informação completa é realizada pelo LIMDEP - versão 8.0.

Por normalidade conjunta,  $f(u_i | \varepsilon_i) = N\left[(\rho/\sigma)\varepsilon_i, (1-\rho^2)\right]$

onde:

$\sigma$  = desvio-padrão da heterogeneidade; e

$\rho$  = correlação entre os erros da equação de seleção e de contagem.

A probabilidade acumulada da distribuição normal condicional é:

$$\text{Prob}[d_i = 1 | w_i, \varepsilon_i] = \Phi\left(\frac{1}{\sqrt{1-\rho^2}}(\alpha'w_i + (\rho/\sigma)\varepsilon_i)\right) \quad (5)$$

A probabilidade não-condicional é:

$$\begin{aligned} P[y_i, d_i = 1 | x_i, w_i] &= \int_{-\infty}^{\infty} \frac{\exp[-\lambda_i(\varepsilon)] \lambda_i(\varepsilon)^{y_i}}{y_i!} \\ &\cdot \Phi\left(\frac{1}{\sqrt{1-\rho^2}}(\alpha'w_i + (\rho/\sigma)\varepsilon)\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\varepsilon^2/(2\sigma^2)\right] d\varepsilon \quad (6) \end{aligned}$$

Quando  $d_i = 0$ , somente  $(d_i, w_i)$  são observados. A contribuição para a função de verossimilhança é:

$$\text{Prob}[d_i = 0 | w_i, \varepsilon_i] = 1 - \Phi\left[\gamma'w_i + \tau\varepsilon_i / (\sqrt{2}\sigma)\right] \quad (7)$$

Onde  $\tau = \sqrt{2}\left[\rho/\sqrt{1-\rho^2}\right]$ ,  $\gamma = \left[1/\sqrt{1-\rho^2}\right]\alpha$  e  $\Phi$  = função de distribuição acumulada normal.

Os estimadores de máxima verossimilhança de  $[\beta, \gamma, \theta, \tau]$  são obtidos pela maximização:

$$\log - L = \sum_{z=0} \log \text{Prob}[d_i = 0 | w] + \sum_{z=1} \log P[y_i, d_i = 1 | x, w] \quad (8)$$

A presença de seleção adversa para características não-observáveis é sinalizada por uma estimativa positiva para  $r$ , que indica uma correlação positiva entre variáveis de risco não-observáveis e a probabilidade de ter em plano do tipo  $j$ .

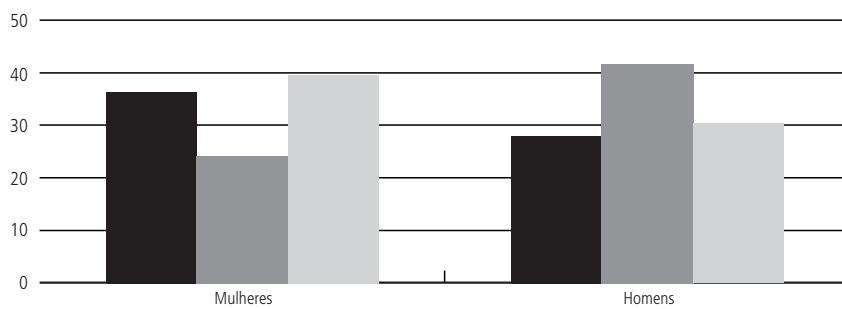
A seleção adversa para características observáveis pode, entretanto, estar presente. Para verificar a seleção observável, analisamos a probabilidade de o indivíduo ter um plano do tipo  $j$  e os atributos de risco observáveis.

#### **4C.4.2 Banco de dados e análise descritiva**

A base de dados usada é a da PNAD de 1998, que possui um suplemento com informações a respeito das características de saúde dos moradores. Neste trabalho, utilizamos dados referentes apenas aos chefes da família. Essa opção se justifica por serem estes os agentes diretamente responsáveis pela decisão de adquirir plano de saúde.

O Gráfico 4C.1 ilustra a proporção de indivíduos por tipo de adesão segundo o sexo. Observamos que entre as mulheres o maior número de adesões ocorre no plano de saúde do tipo voluntário, enquanto entre homens esse percentual é maior no plano via empregador privado. Esse padrão pode estar relacionado à expectativa de utilização do serviço de saúde, que pode ser maior entre as mulheres, fazendo com que aumente o número de adesões ao plano voluntário. Outra explicação possível é o diferente padrão de inserção no mercado de trabalho entre homens e mulheres — observa-se que a proporção de mulheres com plano público é relativamente maior que a de homens, enquanto essa relação se inverte na modalidade de plano via empregador privado.

**GRÁFICO 4C.1  
TIPO DE ADESÃO, SEGUNDO SEXO**



Fonte: Elaboração própria a partir de dados da PNAD de 1998.

■ Plano público ■ Plano via empregador privado ■ Plano voluntário

A Tabela 4C.2 sistematiza a idade média e a renda média segundo a cobertura por plano e a categoria de adesão ao plano. Entre os grupos de indivíduos considerados, a maior idade média é na amostra de indivíduos com plano de saúde de adesão voluntária, enquanto a menor média é na amostra de indivíduos com plano via trabalho privado. Observamos que a diferença é bastante expressiva entre as duas subamostras. Nos planos de adesão voluntária a idade média dos indivíduos é de 49 anos e nos planos com adesão via trabalho privado, de 39 anos. Esse resultado indica existirem diferenças importantes na composição da carteira da seguradora quando analisamos o tipo de adesão ao plano. Quando observamos a idade média segundo cobertura por plano de saúde, a diferença é bem menor que entre as categorias de adesão ao plano.

Com relação à renda média familiar *per capita*, a maior renda observada é a de indivíduos com plano de saúde de adesão voluntária. A renda média de indivíduos com plano chega a ser quase quatro vezes a renda média de indivíduos sem plano. Esse resultado sinaliza a importância da renda na determinação da escolha de adquirir um plano de saúde, enquanto na escolha da modalidade de plano, a renda parece ter um efeito importante, porém menos significativo.

Pela análise da proporção de indivíduos com plano segundo os decis de renda familiar *per capita* (Gráfico 4C.2), constatamos uma tendência de aumento na proporção de indivíduos com plano com o aumento da renda. O nono e o

TABELA 4C.1  
NÚMERO DE OBSERVAÇÕES E PROPORÇÃO DE INDIVÍDUOS, SEGUNDO COBERTURA E CATEGORIA DE ADESÃO AO PLANO

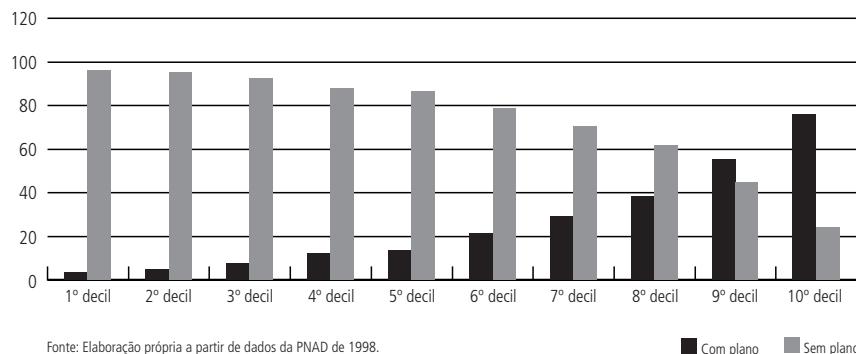
	Total da amostra			Amostra com plano			Total
	Com plano	Sem plano	Total	Público	Via trabalho privado	Voluntário	
Número de observações	25.026	70.646	95.672	7.104	8.784	7.775	23.663
Proporção	26,16	73,84	100	30,02	37,12	32,86	100

TABELA 4C.2  
IDADE MÉDIA (EM ANOS) E RENDA PER CAPITA MÉDIA (EM R\$), SEGUNDO COBERTURA E CATEGORIA DE ADESÃO AO PLANO

	Total da amostra		Com plano de saúde		
	Com plano	Sem plano	Público	Via trabalho privado	Voluntário
Idade média	45,5	44,0	46,8	39,9	49,0
Renda média	667	177	601	588	849

Fonte: Elaboração própria a partir de dados da PNAD de 1998.

**GRÁFICO 4C.2  
PROPORÇÃO DE INDIVÍDUOS COM PLANO, SEGUNDO DECIS DE RENDA**



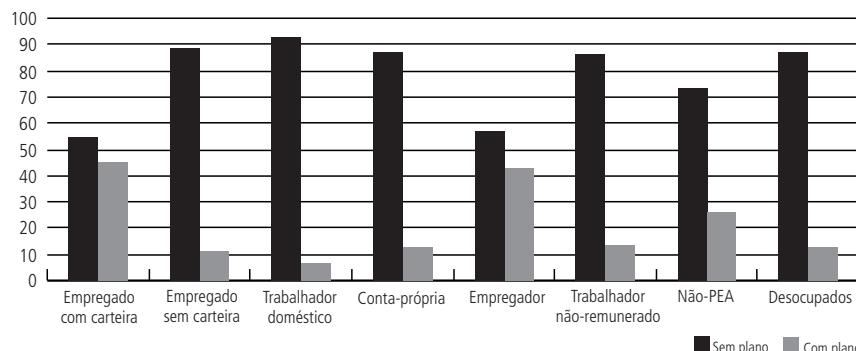
Fonte: Elaboração própria a partir de dados da PNAD de 1998.

■ Com plano ■ Sem plano

décimo decil apresentam uma proporção maior de indivíduos com plano de saúde relativamente aos sem plano, sendo que no décimo decil de renda, quase 80% dos indivíduos possuem algum tipo de plano ou seguro de saúde.

Já pela análise da proporção de indivíduos com plano segundo a ocupação (Gráfico 4C.3), observa-se que empregados com carteira e empregadores têm um percentual acima de 40% de indivíduos com plano. Entre os indivíduos não economicamente ativos a proporção com plano chega a aproximadamente 30%. Esses resultados permitem formular duas hipóteses: primeiro os postos de trabalho mais qualificados, no sentido de propiciarem melhores condições de trabalho, têm maior possibilidade de oferecer benefícios como o acesso a um plano de saúde via empregador; segundo, os diferentes grupos de indivíduos por ocupação refletem diferentes tipos de indivíduos, tanto nos aspectos referentes aos atributos de risco, quanto a renda e características não-observáveis como grau de aversão ao risco.

**GRÁFICO 4C.3  
PROPORÇÃO DE INDIVÍDUOS COM PLANO, SEGUNDO CATEGORIA DE OCUPAÇÃO**



## 4C.5 RESULTADOS

Nesta seção, discutimos os resultados encontrados para a probabilidade de ter plano e a probabilidade de ter uma das categorias de plano. Como explicado na seção metodológica, analisamos a presença de seleção adversa para características de risco observáveis e não-observáveis.

A seleção adversa para características não-observáveis é sinalizada por uma estimativa positiva para  $\rho$ . O coeficiente indica uma correlação positiva entre atributos de risco não-observáveis e a probabilidade de ter plano de saúde. A seleção adversa para características observáveis, por outro lado, é avaliada a partir dos coeficientes estimados pelo modelo *probit* que analisa a probabilidade de participação em um segmento  $j$  de seguro.

### 4C.5.1 A seleção no sistema de saúde suplementar

No primeiro modelo, estamos interessados em analisar a presença de seleção adversa no sistema de saúde suplementar. A equação de seleção corresponde à probabilidade de ter plano. A equação de utilização, numa primeira estimativa, considera a freqüência de consultas médicas e, na segunda, o número de dias internado.

Para analisar a presença de seleção observável, estimamos o efeito marginal das variáveis explicativas do modelo *probit* que analisa a probabilidade de ter plano (Tabela 4C.3). O efeito marginal expressa a taxa de mudança de uma quantidade relativa a outra, ou seja, é a mudança na variável dependente por unidade de mudança na variável independente.

Considerando-se o grupo das principais variáveis de risco, podemos destacar o sexo, a idade, a renda e a escolaridade como variáveis de risco que alteram a chance de adquirir plano.

Estima-se que homens têm probabilidade 3,4% menor de ter plano de saúde que mulheres. A idade afeta a chance de ter plano, isto é, quanto mais velho o indivíduo, maior a probabilidade de adquirir um plano de saúde. A renda familiar *per capita* também é importante quando se trata de adquirir plano de saúde, ou seja, quanto maior a renda, maior a probabilidade de aquisição. Com relação à escolaridade, observamos que quanto mais baixo o nível de escolaridade, menor é a probabilidade de o indivíduo ter um plano. Esse resultado pode ser explicado pela seguinte hipótese: quanto mais bem informados são os indivíduos, maior é o entendimento sobre o funcionamento do mercado de seguros e mais baixos são os custos de informação.

Para analisar com mais clareza os efeitos dos atributos de risco sobre a probabilidade de ter plano, plotamos a probabilidade estimada de ter plano

TABELA 4.C3  
**EFETO MARGINAL ESTIMADO PARA PROBABILIDADE DE TER PLANO DE SAÚDE**

Variáveis	Efeito marginal	Desvio-padrão
Norte	-0,0644537	0,0053231
Nordeste	-0,0735632	0,0042697
Sudeste (exceto São Paulo)	-0,0518382	0,0040549
Sul	-0,0606241	0,0039748
Centro-Oeste	-0,0545574	0,0046917
Sexo	-0,0349803	0,0036498
Idade	0,0053871	0,0005704
Idade <sup>2</sup>	-0,0000271	5,59e-060
Raça	0,0360936	0,0030629
Metropolitana	0,1537754	0,0062231
Urbana	0,1038386	0,0058460
Log renda familiar <i>per capita</i>	0,0524722	0,0022746
Até 3 anos de estudo	-0,3438319	0,0057867
De 4 a 7	-0,2504653	0,0050229
De 8 a 10	-0,1596723	0,0038256
De 11 a 14	-0,1049217	0,0046306
Filhos de 0 e 1 ano	0,0118594	0,0046433
Filhos de 2 a 7	-0,0178389	0,0027290
Filhos de 8 a 15	-0,0135567	0,0021729
Filhos de 16 a 25	-0,0100473	0,0023607
Saúde auto-avaliada	0,0090331	0,0033046
Número de componentes na família	0,0117840	0,0015594
Empregado sem carteira	-0,1593823	0,0026547
Trabalhador doméstico	-0,1566943	0,0033084
Conta-própria	-0,1739232	0,0025804
Empregador	-0,0956783	0,0039427
Não-remunerado	-0,0929025	0,0071126
Não-PEA	-0,0838917	0,0039136
Desocupados	-0,1066975	0,0053058

sobre um dos atributos de risco considerando um determinado grupo de indivíduos. Inicialmente, escolhemos dois grupos: Grupo 1, residente da região Sudeste, branco, área metropolitana, de oito a dez anos de estudo, empregado com carteira e saúde boa; Grupo 2, residente da região Sudeste, branco, área metropolitana, de oito a dez anos de estudo, empregado com carteira e saúde ruim.

O Gráfico 4C.4 mostra como a probabilidade varia de acordo com a renda (grupo de controle 1). Observamos que quanto maior a renda, maior a probabilidade de ter plano. Além disso, nos níveis iniciais de renda a probabilidade de ter plano aumenta mais rapidamente com aumentos na renda que nos níveis seqüentes. Ao longo de todos os níveis de renda plotados a probabilidade de ter um plano é maior para mulheres do que para homens. Essa diferença é de aproximadamente 4% e se mantém relativamente constante ao longo dos níveis de renda considerados.

O Gráfico 4C.5 ilustra a variação na probabilidade por variações na idade. Os aumentos na probabilidade de ter plano são suavemente maiores nos níveis iniciais da idade, e para mulheres a probabilidade é sempre maior do que para homens. Entre 20 e 60 anos, a probabilidade varia cerca de 20% independentemente do sexo.

No Gráfico 4C.6 observamos a variação na probabilidade de acordo com a renda no grupo de controle 2. O primeiro aspecto que chama a atenção é quase não existirem diferenças entre homens e mulheres. Em segundo lugar, os acréscimos na probabilidade são maiores nos níveis iniciais de renda. No Gráfico 4C.7 repetimos o mesmo exercício, porém, considerando a variável idade. Os aumentos

**GRÁFICO 4C.4  
PROBABILIDADE DE TER PLANO E RENDA, SEGUNDO SEXO (GRUPO DE CONTROLE 1)**

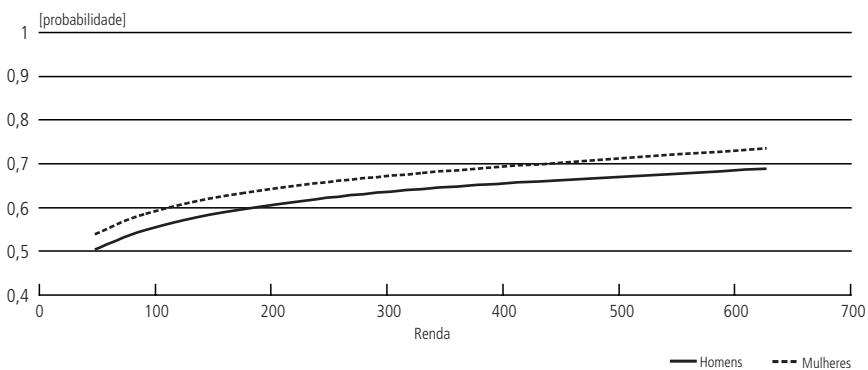


GRÁFICO 4C.5

**PROBABILIDADE DE TER PLANO E IDADE, SEGUNDO SEXO (GRUPO DE CONTROLE 1)**

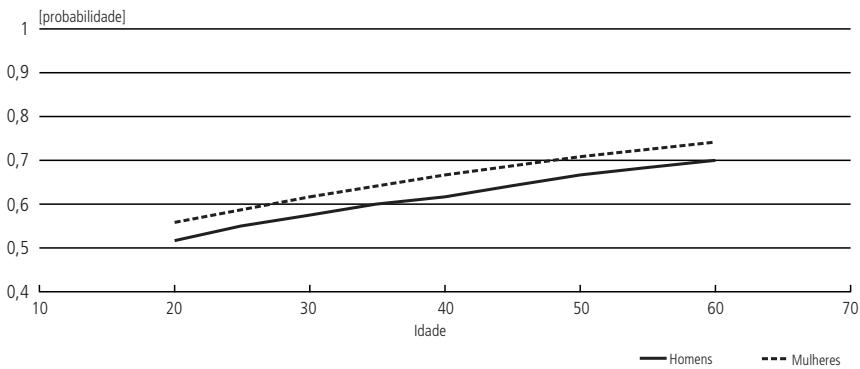


GRÁFICO 4C.6

**PROBABILIDADE DE TER PLANO E RENDA, SEGUNDO SEXO (GRUPO DE CONTROLE 2)**

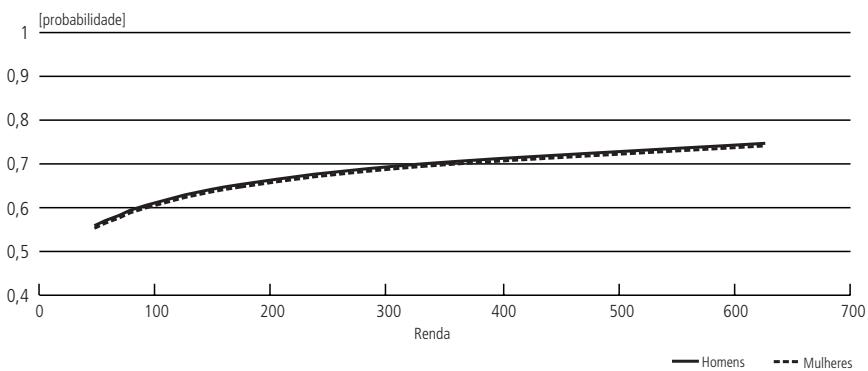
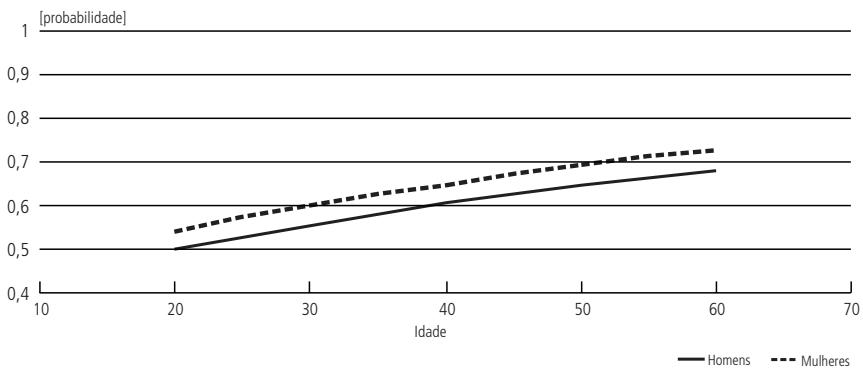


GRÁFICO 4C.7

**PROBABILIDADE DE TER PLANO E IDADE (GRUPO DE CONTROLE 2)**



são relativamente constantes e a diferença na probabilidade entre homens e mulheres é de aproximadamente 4%.

O Gráfico 4C.8 mostra as variações na probabilidade de acordo com a renda entre homens dos grupos 1 e 2 de indivíduos. A diferença entre os dois grupos é o estado de saúde auto-avaliado. Entre os indivíduos com pior estado de saúde a chance de ter plano é maior do que entre os que possuem saúde melhor, independentemente do nível de renda: a diferença na probabilidade é de aproximadamente 5%. Realizando-se o mesmo exercício de acordo com a idade (Gráfico 4C.9), o resultado se inverte, ou seja, os indivíduos em pior estado de saúde têm chance menor de ter plano do que aqueles em melhor estado de saúde. A diferença entre as probabilidades se mantém em cerca de 1,5% ao longo da idade.

GRÁFICO 4C.8

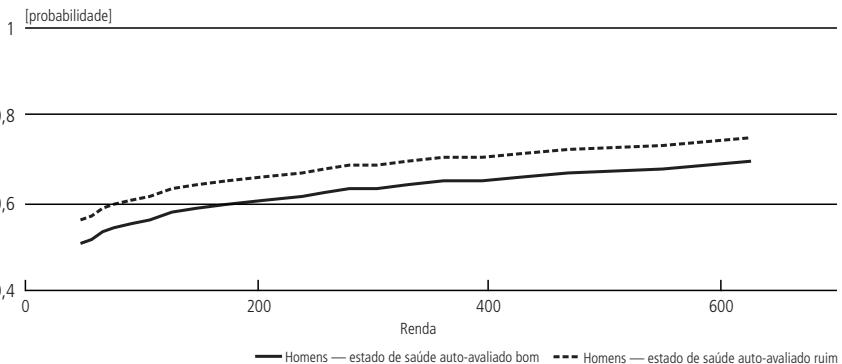
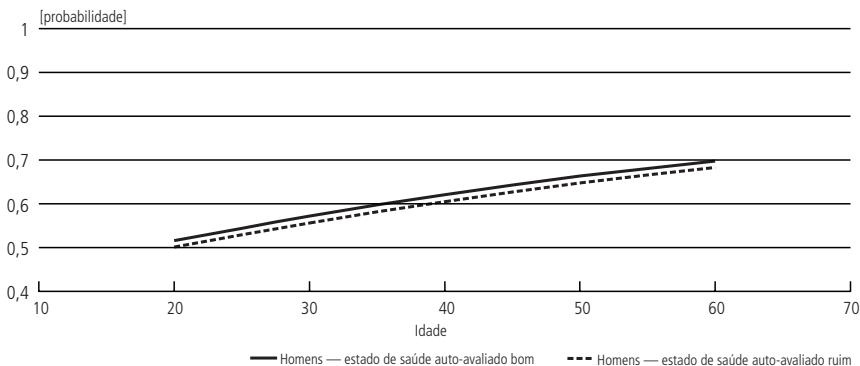
**PROBABILIDADE DE TER PLANO E RENDA (HOMENS — GRUPO 1 E GRUPO 2)**

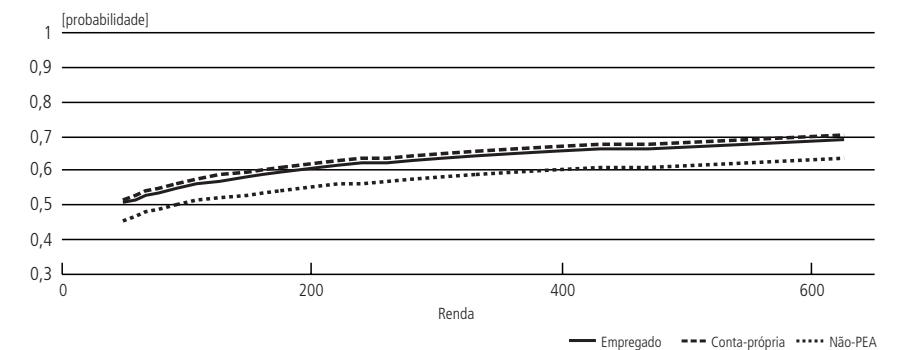
GRÁFICO 4C.9

**PROBABILIDADE DE TER PLANO E IDADE: HOMENS (GRUPO 1 E GRUPO 2)**

O Gráfico 4C.10 mostra o efeito da renda sobre a probabilidade de ter plano para os homens do primeiro grupo de indivíduos e de acordo com três categorias de ocupação — empregado conta-própria e não-PEA. A maior probabilidade de ter plano é entre os conta-própria, sendo essa diferença muito pequena para os empregados com carteira.

Até o momento, reportamo-nos aos resultados da seleção observável. A seleção não-observável, por sua vez, foi encontrada não-significativa, independentemente do tipo de cuidado considerado. As estimativas para o coeficiente Rho que indica a correlação entre os modelos encontram-se no Anexo.

**GRÁFICO 4C.10  
PROBABILIDADE DE TER PLANO E RENDA DE ACORDO COM A OCUPAÇÃO: HOMENS (GRUPO 1)**



#### **4C.5.2 A seleção adversa entre as modalidades de adesão ao plano**

Nesta subseção, discutimos os resultados da seleção entre as modalidades de adesão ao plano de saúde. A Tabela 4C.4 mostra o efeito marginal estimado para a probabilidade de ter uma determinada categoria de plano, dado que se possua um plano. Na discussão dos resultados, analisamos as principais variáveis de risco associadas a essa probabilidade. Com relação à variável sexo, observamos que a probabilidade de se ter um plano via empregador público e a probabilidade de se ter um plano de adesão voluntária é menor para homens, enquanto no plano via empregador privado, a probabilidade é maior para homens. Esses resultados mostram que, considerando-se sexo uma variável de risco, os planos via empregador público e os planos de adesão voluntária selecionam adversamente seus consumidores.<sup>40</sup>

40. Como o número de observações é mais reduzido que no modelo anterior, optamos por não plotar o efeito das mudanças nos atributos de risco sobre a probabilidade de ter plano em um determinado grupo de indivíduos.

TABELA 4C.4

**EFETO MARGINAL ESTIMADO PARA PROBABILIDADE DE TER PLANO DE SAÚDE PÚBLICO, TER PLANO VIA EMPREGADOR PRIVADO E PLANO DE ADESÃO VOLUNTÁRIA, DADO QUE SE POSSUA UM PLANO**

Variáveis	Plano via empregador público		Plano via empregador privado		Plano de adesão voluntária	
	Efeito marginal	Desvio-padrão	Efeito marginal	Desvio-padrão	Efeito marginal	Desvio-padrão
Norte	0,4179	0,0158817	-0,2516	0,0099084	-0,0808	0,0139967
Nordeste	0,2759	0,0119579	-0,1600	0,0092343	-0,0663	0,0096194
Sudeste (exceto São Paulo)	0,1602	0,0113816	-0,0862	0,0093258	-0,0352	0,0092068
Sul	0,2197	0,0118657	-0,0504	0,0097247	-0,1262	0,0085514
Centro-Oeste	0,3925	0,0132239	-0,1783	0,0101123	-0,1429	0,0098337
Sexo	-0,0351	0,007774	0,1118	0,0081365	-0,0733	0,0082909
Idade	0,0048	0,0013943	-0,0079	0,0016199	0,0073	0,0014371
Idade <sup>2</sup>	0,0000 <sup>+</sup>	0,00000138	0,0000 <sup>+</sup>	0,00000169	-0,0001	0,0000142
Raça	-0,0550	0,0071794	-0,0028 <sup>+</sup>	0,0078264	0,0640	0,0073103
Metropolitana	-0,0252 <sup>+</sup>	0,0182252	0,0454	0,0186924	-0,0207 <sup>+</sup>	0,0191311
Urbana	0,0219 <sup>+</sup>	0,0182665	0,0097 <sup>+</sup>	0,0189323	-0,0317	0,0188814
Log renda familiar <i>per capita</i>	-0,0128	0,002568	0,0112	0,0031531	0,0031 <sup>+</sup>	0,002642
Até 3 anos de estudo	-0,0832	0,0112225	0,1430	0,015817	-0,0502	0,0125368
De 4 a 7	-0,0804	0,00951	0,1341	0,0120679	-0,0529	0,010273
De 8 a 10	-0,0351	0,0103604	0,0757	0,0126268	-0,0472	0,0107631
De 11 a 14	0,0108 <sup>+</sup>	0,0089951	0,0429	0,0102942	-0,0603	0,0091958
Filhos de 0 e 1 ano	0,0066 <sup>+</sup>	0,011209	-0,0069 <sup>+</sup>	0,0117523	-0,0101 <sup>+</sup>	0,0125228
Filhos de 2 a 7	0,0271	0,0066357	-0,0197	0,0072057	-0,0207	0,0073604
Filhos de 8 a 15	0,0216	0,0054708	-0,0062 <sup>+</sup>	0,0060783	-0,0278	0,0060754
Filhos de 16 a 25	0,0245	0,0056111	-0,0062 <sup>+</sup>	0,006376	-0,0227	0,0060111
Saúde auto-avaliada	-0,0183	0,0078307	0,0313	0,0086838	-0,0083 <sup>+</sup>	0,0081733
Número de componentes na família	-0,0123	0,0040182	0,0234	0,0045773	-0,0054 <sup>+</sup>	0,004238
Empregado sem carteira	-0,0100 <sup>+</sup>	0,0136812	-0,2043	0,0104027	0,3096	0,0153846
Trabalhador doméstico	-0,1226	0,0265587	-0,1564	0,0274816	0,3828	0,0332624
Conta-própria	-0,1036	0,0087416	-0,2919	0,0064897	0,4897	0,0096254
Empregador	-0,2022	0,0079742	-0,2825	0,0071666	0,5676	0,009547
Não-remunerado	-0,0309 <sup>+</sup>	0,0276522	-0,2329	0,0208771	0,3696	0,0308633
Não-PEA	0,0251	0,0108211	-0,2369	0,0092329	0,2695	0,0118394
Desocupados	-0,0327 <sup>+</sup>	0,0218247	-0,2220	0,0152434	0,3597	0,0236521

+ Coeficiente estimado não-significativo.

Com relação à idade, o efeito é positivo sobre a chance de ter plano via empregador público e plano voluntário, e negativo sobre a chance de ter plano via empregador privado. Como a idade é um fator de risco importante na composição da carteira das seguradoras, podemos inferir que os planos via empregador privado selecionam indivíduos com uma taxa de risco menor que as de outras modalidades de adesão.

A renda, por sua vez, teve efeito negativo sobre a probabilidade de ter plano público. Esse resultado pode ser explicado pelos menores níveis de renda entre os que possuem essa categoria de plano. No plano via empregador privado o efeito da renda foi positivo e no plano com adesão voluntária o coeficiente foi não-significativo.

A seleção adversa com base em atributos de risco não-observáveis foi não-significativa nos três modelos estimados, independentemente do cuidado de saúde. As estimativas do coeficiente Rho encontram-se no Anexo.

#### **4C.5.3 Considerações finais**

Este estudo procurou investigar a presença da seleção adversa no sistema de saúde suplementar brasileiro. Analisamos a seleção com base em atributos de risco observáveis e não-observáveis. A seleção observável é aquela na qual foi possível identificar a característica de risco do indivíduo — essa característica é aquela que aumenta a chance de utilização do seguro contratado.

A seleção adversa é analisada em quatro momentos distintos: probabilidade de ter plano de saúde, de ter plano público, privado ou voluntário, dado que se possua plano de saúde. Os principais resultados revelam que atributos de risco observáveis, como sexo e idade, são importantes tanto na decisão de adquirir plano de saúde, quanto na categoria do plano a adquirir. As características de risco não-observáveis foram não-significativas.

Esses resultados são importantes na medida em que indicam as diferenças de risco médio entre os grupos de indivíduos no sistema de saúde brasileiro. Se as seguradoras podem, por sua vez, discriminar os consumidores por seus atributos de risco, a seleção é totalmente observável e é possível ajustar o prêmio do seguro com base no risco da população que adquire o plano. Se, por outro lado, as seguradoras não discriminam seus consumidores, ou devido a dispositivos legais que impedem a discriminação ou porque não identificam essas características, a seleção adversa pode gerar impactos sobre o bem-estar dos indivíduos e sobre a eficiência na alocação dos recursos.

No sistema de saúde brasileiro, a regra de apreçamento para reajuste do prêmio dos planos, segundo a faixa etária dos consumidores, pode resultar em seleção adversa no mercado de saúde suplementar. Esses efeitos só poderão ser mensurados com a utilização de dados mais recentes que os deste trabalho. Nesse sentido, este estudo constitui uma primeira investigação sobre em que medida o setor de saúde suplementar apresenta particularidades no contexto de regulação do setor.

## 4D. ESTUDO DE CASO DA DEMANDA

### 4D.1 INTRODUÇÃO

Este estudo apresenta os resultados finais da estimativa da demanda por serviços de saúde considerando a população beneficiária do plano de saúde oferecido pela Fundação Sabesp de Seguridade Social (Sabesp). A Sabesp é gestora do Fundo de Assistência e Previdência dos Funcionários da Companhia de Água e Esgoto do Estado de São Paulo (Sabesp). A estimativa foi feita a partir dos gastos anuais individuais. Os gastos com saúde foram classificados em ambulatoriais e de internação. Os resultados mostram que os atributos de risco, idade, sexo e estado de saúde passado são bastante relevantes na determinação dos gastos com serviços de saúde. Os coeficientes estimados, tanto para a probabilidade de gastar quanto para a equação do valor esperado dos gastos ambulatoriais e de internação, mostram que existem diferenças importantes para os grupos etários e entre os sexos. Esses resultados são fundamentais uma vez que a regulamentação vigente no Brasil não permite discriminação por sexo e apresenta uma divisão de grupos etários pouco justificada na literatura. Os resultados da análise controlada e da estatística descritiva apontam para outra divisão etária que parece ser mais relevante do ponto de vista da construção de grupos de risco mais homogêneos.

O trabalho está dividido em mais quatro seções além desta breve introdução: a Seção 4D.2 descreve o banco de dados e a construção das variáveis, a Seção 4D.3 apresenta a metodologia e a Seção 4D.4 os modelos estimados e resultados encontrados. Por fim a Seção 4D.5 expõe as conclusões.

### 4D.2 BASE DE DADOS

A base de dados utilizada neste trabalho consiste dos registros administrativos dos beneficiários da Sabesp,<sup>1</sup> para os anos de 2001 a 2003. O plano de

saúde oferecido se insere na modalidade de autogestão sendo ofertados aos beneficiários (empregados e dependentes diretos e indiretos) três tipos de contratos: o primeiro, denominado plano pleno, contempla os funcionários da ativa da Sabesp e os seus dependentes legais (cônjuges e filhos), os demais, denominados plano básico e plano especial, têm como público-alvo os aposentados e ex-funcionários da Sabesp, seus pensionistas (viúvos e viúvas) e parentes em vários graus (dependentes e agregados). A única diferença entre esses dois contratos é a cobertura de serviços: a do especial é mais abrangente que a do básico.<sup>41</sup> O plano da SabespPrev opera com rede credenciada e permite a cobrança de co-participações e franquias. Para este trabalho não tivemos acesso às informações de pagamento de co-participações.

Este banco difere dos bancos usualmente utilizados em pesquisas acadêmicas, uma vez que é um banco administrativo, cuja unidade de análise são os procedimentos utilizados por paciente. As informações originais estavam organizadas em cinco bancos separados, conforme diagrama apresentado no Anexo III, tendo sido necessário utilizar variáveis-chave que fizessem a ligação desses bancos. As variáveis disponibilizadas no banco original também são apresentadas no mesmo anexo.

No Brasil as informações disponíveis de gastos com saúde são, em geral, de dois tipos: informações provenientes de pesquisas domiciliares nas quais os indivíduos declaram quanto gastam com saúde e os respectivos componentes [por exemplo, Pesquisa Nacional por Amostra de Domicílios (PNAD) e Pesquisa de Orçamento Familiar (POF)] e, informações obtidas de prestadores públicos sobre gastos hospitalares e ambulatoriais de alta complexidade [Sistema de Informação Hospitalar (SIH) e Sistema de Informação Ambulatorial (SIA)]. Os problemas relativos a cada um deles referem-se, em primeiro lugar, à confiabilidade dos dados, uma vez que são autodeclarados e suscetíveis a juízos de valor por parte de quem presta a informação e, em segundo, à sujeição a distorções nos preços relativos, principalmente nos procedimentos realizados na área ambulatorial. As críticas ao SIA são relacionadas à elaboração dos registros de cobrança, que permitem ao médico preencher a guia de um procedimento e realizar outro. Tal distorção é menor na área hospitalar, uma vez que o setor apresenta uma estrutura de mercado mais oligopolizada. O banco de dados neste trabalho é inédito, e sua principal vantagem é permitir acesso a informações sobre as quais não ocorrem as distorções já descritas, uma vez que a informação é preenchida pelo próprio gestor que realiza os pagamentos e tende a coibir as

---

41. Os três tipos de contrato estão disponíveis em [www.sabesp.com.br](http://www.sabesp.com.br).

práticas de registro citadas, que ocorrem principalmente no setor público, devendo à própria necessidade de manter a sustentabilidade da carteira.

Para realizar a análise da demanda dos serviços de saúde, o primeiro passo foi a reorganização do banco de dados original considerando-se como unidade de análise os indivíduos. Desse modo, procedeu-se à vetorização do banco, procedimento que consiste em reconstruir o banco de forma que cada linha representasse um indivíduo com todos os procedimentos realizados pelo mesmo.

A base de dados construída neste trabalho contém informações sobre a utilização e os gastos de serviços de saúde e sobre dados socioeconômicos e demográficos de seus beneficiários para o período 2001-2003, sendo composta de, aproximadamente, 66 mil indivíduos — funcionários, ex-funcionários da empresa e seus dependentes, agregados e designados. Além disso, como o banco possuía mais informações sobre os titulares — funcionários da empresa — foi criada uma amostra apenas desses indivíduos, que permitiu a inclusão de alguns atributos de risco não-disponíveis para os beneficiários. A amostra dos titulares é composta de cerca de 20 mil indivíduos. As estimativas da demanda por serviços de saúde serão realizadas considerando-se o conjunto de toda a população beneficiária dos planos de saúde da Sabesperv e o conjunto dos titulares.

#### **4D.2.1 Construção das variáveis socioeconômicas e demográficas**

O banco de dados original disponibilizado pela Sabesperv é um banco administrativo dos registros de procedimentos médicos realizados com a carteira de beneficiários. Para utilizá-lo foi necessário organizar as informações em microdados sendo a unidade de análise cada usuário/indivíduo. Além disso, os dados de titulares e beneficiários estavam armazenados em bases separadas, existindo mais informações para os titulares. O banco original continha uma série de variáveis que não foram utilizadas, seja por não serem relevantes para a análise, seja por não apresentarem consistência no preenchimento. As informações extraídas para os usuários foram de dois tipos: atributos pessoais e registros de utilização dos serviços médicos. O primeiro grupo contém as seguintes informações: identificação no plano (código Sabesperv), sexo, data de nascimento, estado civil, tipo de plano, data de inclusão no plano, descrição e data da situação atual no plano, descrição e data da alteração do plano, renda e escolaridade.<sup>42</sup> As informações de utilização foram o valor e a descrição de cada

---

42. As informações de renda e escolaridade só são disponíveis para os titulares por serem os empregados da empresa. Além dessas informações estava disponível ainda a data de admissão na empresa. A informação do plano não foi utilizada para os titulares, uma vez que todos deveriam estar, obrigatoriamente inscritos no plano pleno.

procedimento.<sup>43</sup> No Anexo III encontram-se o formato no qual o banco foi disponibilizado e a forma de ligação entre as diversas bases.

#### **4D.2.2 Construção e tipologia dos gastos**

Neste trabalho a demanda por serviços de saúde é estimada utilizando-se como medida os gastos anuais individuais com esses serviços. Outra forma usual de se estimar a demanda por serviços de saúde é através da sua utilização. A literatura empírica utiliza indiscriminadamente uma forma ou outra, dependendo da disponibilidade de informações. No caso do banco da Sabesperv as informações de gastos são obtidas diretamente, daí a nossa escolha por essa unidade de medida.

A tipologia utilizada neste trabalho classifica os gastos em ambulatoriais e com internação hospitalar. Essa classificação é importante porque a natureza desses dispêndios é totalmente distinta. Os gastos ambulatoriais incluem, de maneira genérica, gastos com consultas, exames e tratamento preventivo (*ex ante* a realização da incerteza). Os gastos com internação hospitalar se referem a dispêndios usualmente realizados quando o paciente está doente, ou seja, *ex post* a realização da incerteza. Desse modo, do ponto de vista da elasticidade-renda e elasticidade dos atributos de risco, o comportamento deve ser muito diferente segundo a categoria de gasto, justificando sua separação.

Os gastos com internação foram identificados utilizando-se dois critérios dada a disponibilidade de informações contidas no banco: o Código da Associação Médica Brasileira (AMB)<sup>44</sup> e o Porte Anestésico (PA). Inicialmente foram identificados nas tabelas da AMB os procedimentos utilizados exclusivamente em internações.<sup>45</sup> Os gastos com esses procedimentos referentes a qualquer usuário foram separados mensalmente. Com relação ao segundo critério, todos os atos anestésicos foram classificados em portes de 0 a 8, conforme a complexidade do procedimento. Segundo especialistas, quando o porte anestésico é igual ou superior a 3, o procedimento deve ser realizado dentro dos hospitais. Assim, os procedimentos com porte anestésico nessa categoria também foram identificados e separados mensalmente como procedimentos de internação hospitalar. Dentro do mês, sempre que um usuário fosse identificado como submetido a algum procedimento de internação, todos os seus gastos foram

43. Os gastos individuais e a forma como foram construídos e categorizados serão descritos a seguir.

44. No banco original os procedimentos estavam registrados segundo diversas tabelas da AMB. Desse modo, a informação sobre o tipo de tabela também foi utilizada.

45. O Anexo V apresenta os procedimentos utilizados na decomposição dos gastos de internação.

classificados nessa categoria e somados para a vетorização do banco, inclusive os ambulatoriais. O banco foi vетorizado para que o usuário passasse a ser a unidade de análise, de modo que cada linha do banco contivesse todos os custos de procedimentos realizados com aquele usuário em cada mês e ano. A utilização do período de tempo mensal justifica-se como sendo o prazo suficiente para o usuário realizar as consultas e exames necessários e, junto com seu médico, decidir pela internação.<sup>46</sup> Os gastos ambulatoriais foram identificados por exclusão, ou seja, todos os gastos que não atenderam aos critérios anteriormente descritos foram classificados como ambulatoriais. Após a vетorização, os gastos mensais foram somados para se encontrar os gastos anuais ambulatoriais e com internação hospitalar de cada usuário.

#### **4D.2.3 Estatística descritiva**

Esta subseção apresenta a estatística descritiva das principais informações contidas no banco de dados. As estatísticas são apresentadas separadamente para beneficiários e titulares. Foi realizada a análise descritiva das variáveis socioeconômicas e demográficas para o ano de 2003.<sup>47</sup> As estatísticas dos gastos são apresentadas para os anos de 2001 a 2003. No final é apresentada uma subseção relacionando os gastos aos principais atributos de risco.

##### **4D.2.3.1 Análise descritiva das variáveis socioeconômicas e demográficas**

A Tabela 4D.1 apresenta a idade e a renda média de titulares, segundo gênero.<sup>48</sup> Observa-se que predominam homens na amostra, representando cerca de

TABELA 4D.1  
**IDADE E RENDA MÉDIA DOS TITULARES SEGUNDO O SEXO — 2003**

	Número	%	Idade média	Renda média
Homens	15.551	81,49	44,80 (10,21)	2.306,11 (1.909,12)
Mulheres	3.531	18,51	42,53 (10,45)	2.512,49 (1.701,82)
Total	19.082	100,00	44,34 (10,30)	2.344,30 (1.874,15)

Fonte: Elaboração própria a partir dos dados da Sabesperv.

46. Evidentemente, pode-se incorrer em um erro do tipo II neste caso (de aceitar a hipótese nula, com sendo falsa), pois um usuário com problema renal, por exemplo, pode, eventualmente, ir ao oculista no mesmo mês em que se interna. Tais eventos, entretanto, são classificados como raros e, dada a forma como o banco estava construído, ter-se-ia um custo elevado para isolar esses casos. Além disso, qualquer periodicidade escolhida seria arbitrária. A fragilidade do período de um mês decorre da possibilidade de haver procedimentos que foram realizados em determinado mês, mas o pagamento ter sido executado somente no mês subsequente.

47. A análise para os demais anos é bastante similar dado que em um período de três anos a alteração da população usuária não é tão grande. Segundo dados da Tabela VI.4 do Anexo VI, no período de 2001 a 2003 foram excluídos da carteira de titulares perto de 1.500 empregados, o que corresponde a cerca de 7,5% dos titulares.

48. A análise realizada para os beneficiários inclui informações de gastos, idade e sexo, devido a disponibilidade dos dados.

81% da população. Isso pode ser explicado pelo tipo de serviço operacional da empresa, que é ainda predominantemente ocupado por indivíduos do sexo masculino. A idade média dos homens é ligeiramente superior à das mulheres — 44,8 *versus* 42,5 anos. A renda média, no entanto, é maior entre as mulheres em cerca de 9%. Esse resultado sugere que as mulheres ocupam postos de trabalho com remuneração maior, uma vez que apresentam distribuição mais concentrada nos graus de instrução mais elevados.

No caso dos beneficiários a distribuição da idade por gênero é bastante diferente da observada para os titulares: as mulheres representam 51,28%, aproximando-se da distribuição verificada na população brasileira e a idade média é bastante inferior — 34 anos *versus* 44,34 para os titulares. Os resultados são apresentados na Tabela 4D.2.

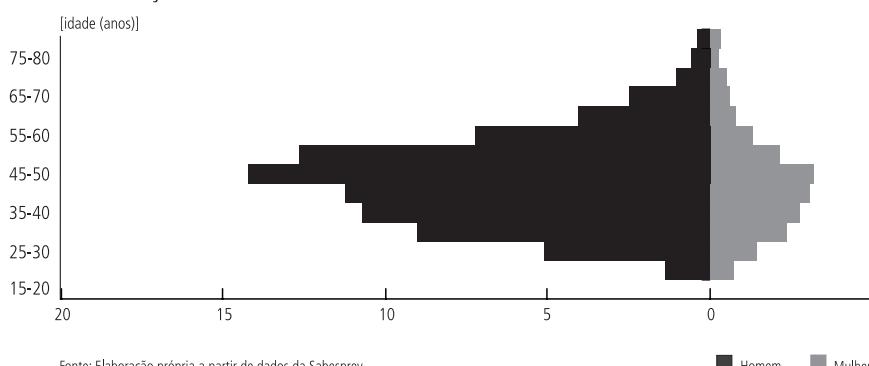
TABELA 4D.2  
IDADE MÉDIA DOS BENEFICIÁRIOS SEGUNDO O SEXO — 2003

	Número	%	Idade média
Homens	32.042	48,71	33,05 (18,61)
Mulheres	33.727	51,28	35,15 (20,09)
Total	65.769	100,00	34,13 (19,42)

Fonte: Elaboração própria a partir dos dados da Sabesperv.

O Gráfico 4D.1 e a Tabela VI.1 (Anexo VI) apresentam a distribuição etária dos titulares do plano segundo o sexo. Nota-se que o grupo etário de maior participação na amostra é o de 45 a 49 anos, tanto para os homens quanto para as mulheres titulares do plano.

GRÁFICO 4D.1  
DISTRIBUIÇÃO ETÁRIA DOS TITULARES DA SABESP



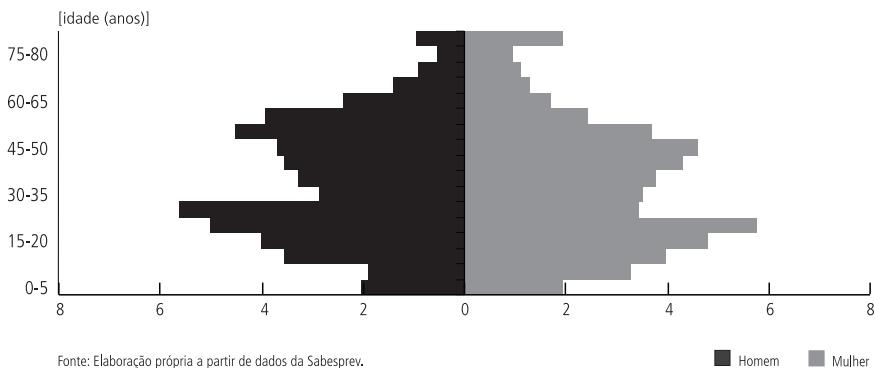
Para os beneficiários, a distribuição etária segundo o sexo é apresentada no Gráfico 4D.2 e na Tabela VI.2 do Anexo VI. O grupo com maior participação é o de 20 a 25 anos e a participação das mulheres é mais elevada em todas as faixas etárias.

As informações atinentes ao grau de instrução estão disponíveis apenas para os titulares do plano e apresentam as seguintes categorias: analfabeto, primário incompleto, primário completo, primeiro grau incompleto, primeiro grau completo, segundo grau incompleto, segundo grau completo, terceiro grau incompleto e terceiro grau completo. Essas categorias foram codificadas de maneira a corresponder a ciclos completos de instrução: indivíduos com menos de 4 anos de estudo, indivíduos com 4 a 7 anos de estudo, indivíduos com 8 a 10 anos, indivíduos com 11 a 14 e indivíduos com 15 anos ou mais.

A distribuição dos titulares segundo o grau de instrução evidencia um elevado nível de escolaridade para os empregados da Sabesp — mais de 50% tanto de homens quanto de mulheres apresentam pelo menos segundo grau completo, ou seja, acima de 11 anos de estudo. Além disso, o percentual de indivíduos com menos de 4 anos de estudo é bem inferior ao observado para a população brasileira — 9,32% dos homens têm apenas primário completo e/ou primeiro grau incompleto, contra 2,21% das mulheres. Os resultados são apresentados na Tabela 4D.3.<sup>49</sup>

Ao se comparar esses resultados com o Censo Demográfico de 2000 do IBGE, observa-se que 15,36% da população entrevistada têm de 8 a 10 anos de

**GRÁFICO 4D.2  
DISTRIBUIÇÃO ETÁRIA DOS BENEFICIÁRIOS DA SABESP**



49. Das 19.082 observações, 2.896 não apresentaram declaração de escolaridade.

**TABELA 4D.3  
DISTRIBUIÇÃO DOS TITULARES SEGUNDO A ESCOLARIDADE — 2003**

Escolaridade	Total		Homens		Mulheres	
	Número	%	Número	%	Número	%
0 (analfabeto/primário incompleto)	1.309	8,09	1.246	9,32	63	2,23
1 (primário completo/1º grau incompleto)	3.051	18,86	2.916	21,81	135	4,79
2 (1º grau completo/2º grau incompleto)	3.024	18,68	2.704	20,22	320	11,35
3 (2º grau completo/3º grau incompleto)	5.266	32,53	4.167	31,17	1.099	38,99
4 (3º grau completo)	3.536	21,84	2.335	17,47	1.201	42,62
Total	16.186	100,00	13.368	100,00	2.818	100,00

Fonte: Elaboração própria a partir dos dados da Sabesperv.

estudo (correspondente aos indivíduos com segundo grau completo/terceiro grau completo do plano Sabesperv). Ou seja, um resultado inferior comparado ao da Sabesperv. As pessoas com menos de 4 anos de estudo equivalem a 29,83% do total dos entrevistados pelo censo, ou seja, representam quase três vezes mais o valor percentual encontrado para a Sabesperv. Isso se deve ao fato de que são considerados como titulares da Sabesperv apenas indivíduos com 15 ou mais anos de idade.

Quanto ao estado civil, os titulares — em sua maioria (67%) — pertencem ao grupo 1 (casados), como pode ser visto na Tabela 4D.4. Esse grupo representa 72,29% dos homens e 45,9% das mulheres.

Em relação ao estado civil dos beneficiários, os solteiros são o grupo mais representativo, com o percentual de homens superior ao de mulheres. Por outro lado, os casados e os outros grupos apresentam um percentual de pessoas do sexo feminino superior ao de pessoas do sexo masculino.

**TABELA 4D.4  
DISTRIBUIÇÃO DOS TITULARES SEGUNDO O ESTADO CIVIL — 2003**

Estado civil	Total		Homens		Mulheres	
	Número	%	Número	%	Número	%
0 (solteiro)	3.611	18,92	2.340	15,05	1.271	35,99
1 (casado)	12.833	67,25	11.241	72,29	1.592	45,09
2 (outros)	2.638	13,83	1.970	12,67	668	18,92
Total	19.082	100,00	15.551	100,00	3.531	100,00

Fonte: Elaboração própria a partir dos dados da Sabesperv.

**TABELA 4D.5  
DISTRIBUIÇÃO DOS BENEFICIÁRIOS SEGUNDO O ESTADO CIVIL — 2003**

Estado civil	Total		Homens		Mulheres	
	Número	%	Número	%	Número	%
0 (solteiro)	31.214	47,46	16.014	49,98	15.200	45,07
1 (casado)	28.708	43,65	13.648	42,60	15.060	44,65
2 (outros)	5.846	8,89	2.379	7,42	3.467	10,28
Total	65.768	100,00	32.041	100,00	33.727	100,00

Fonte: Elaboração própria a partir dos dados da Sabesperv.

Com relação à situação atual dos titulares na empresa tem-se que cerca de 87,79% dos indivíduos estão ativos no período 2001-2003 e 3,40% foram desligados sem justa causa. Os resultados encontram-se na Tabela VI.3 do Anexo VI.

#### 4D.2.3.2 Análise descritiva dos gastos

A Tabela 4D.6 apresenta os valores médios dos gastos dos titulares decompostos segundo a metodologia proposta para os anos de 2001, 2002 e 2003. Todos os valores estão apresentados a preços constantes de 2001.<sup>50</sup> Como se observa, os comportamentos dos gastos médios ambulatoriais e de internação são distintos ao longo dos três anos: enquanto os gastos médios ambulatoriais declinam no período (cerca de 18,30%), os gastos com internação se elevam (cerca de 54,02%). Essa diferença pode estar associada a dois motivos básicos: o primeiro, a inflação dos serviços de saúde, que foi de 9,01% em 2002, em comparação aos 4,11% de 2001. Como a inflação apresentou grande variação em 2002 em relação a 2001 os preços podem não ter sido reajustados imediatamente, causando queda dos gastos reais. Por outro lado, como pode ser visto

**TABELA 4D.6  
GASTO MÉDIO COM SERVIÇOS DE SAÚDE DOS TITULARES — 2001 A 2003**

	2001				2002				2003			
	Média	Desvio- padrão	Mínimo	Máximo	Média	Desvio- padrão	Mínimo	Máximo	Média	Desvio- padrão	Mínimo	Máximo
Ambulatorial	202	480,8	-138	44.594	176	412,6	0	24.311	165	353,3	0	17.935
Internação	228	2.024	-47,8	2E+05	282	1.638	0	70.017	352	2.002	0	68.571

Fonte: Elaboração própria a partir dos dados da Sabesperv.

50. Os gastos foram deflacionados pelo IPC saúde da Fipe disponível em: <[www.fipe.com/indices](http://www.fipe.com/indices)>.

TABELA 4D.7  
GASTO MÉDIO COM SERVIÇOS DE SAÚDE DOS BENEFICIÁRIOS — 2001 A 2003

Gastos	2001				2002				2003			
	Média	Desvio-padrão	Mínimo	Máximo	Média	Desvio-padrão	Mínimo	Máximo	Média	Desvio-padrão	Mínimo	Máximo
Ambulatorial	165,84	517,6	-138,38	44.593,5	141,73	403,99	-4,03	38.770,42	135,90	520,90	0	60.030,65
Internação	228,20	2.145,3	-47,80	182.469,2	285,65	2.412,56	-108,3	228.362,22	329,49	2.821,89	0	264.727,61

Fonte: Elaboração própria a partir dos dados da Sabesperv.

na Tabela 4D.8 ocorreu, a partir de 2002, uma elevação na utilização dos serviços hospitalares (na divisão por quantis passaram de cerca de 10% para 25% dos usuários), o que também contribui para a elevação do gasto médio. O segundo é a alteração do perfil de utilização dos usuários. Como a carteira é relativamente estável no período — 7,5% dos titulares foram excluídos nos três anos — é pouco razoável associar toda essa variação a uma alteração desse perfil. Além da diferença dos valores médios entre os gastos ambulatoriais e de internação, a dispersão dos gastos com internação é bem superior, apresentando desvio-padrão cerca de quatro vezes superior ao observado para os gastos médios ambulatoriais. Os gastos com internação apresentam ainda valores máximos elevados, reforçando a importância do seguro para esse tipo de gasto.

Os valores negativos presentes na tabela são decorrentes dos valores glosados. Os maiores valores glosados referem-se aos gastos ambulatoriais, mas também não representam um montante tão significativo.

A queda nos gastos médios ambulatoriais dos beneficiários é similar à dos titulares: cerca de 18%. Por outro lado, a elevação dos gastos hospitalares é menor: 44,39%.<sup>51</sup> Os resultados são apresentados na Tabela 4D.7. As razões para as variações nos gastos são similares àquelas apresentadas para os titulares.

TABELA 4D.8  
DISTRIBUIÇÃO DOS GASTOS MÉDIOS DOS TITULARES POR QUANTIS — 2001 A 2003

Quantil (%)	2001		2002		2003	
	Ambulatorial	Internação	Ambulatorial	Internação	Ambulatorial	Internação
100 max	44.593,5	152.402	24.311,53	70.016,54	17.935,44	68.570,51
99	1.133,43	4.049,8	1.027,97	4.794,29	974,89	5.569,77
95	661,64	979,61	583,64	1.167,08	544,59	1.374,36
90	485	267,62	424,7	451,03	386,88	567,52
75	253,32	0	216,85	90,64	201,25	136,24
50 mediana	110,92	0	92,34	0	91,2	0
25	40,13	0	31,69	0	33,32	0
10	9,71	0	6,09	0	7,69	0
5	0	0	0	0	0	0
1	0	0	0	0	0	0
0 min	-138,39	-47,8	0	0	0	0

Fonte: Elaboração própria a partir dos dados da Sabesperv.

51. De maneira similar à amostra dos titulares, o teste de médias implicou rejeitar a hipótese nula de diferença igual a 0 para a média de gastos entre um ano e o imediatamente subsequente, tanto para gastos ambulatoriais, quanto para gastos de internação na população dos beneficiários.

A Tabela 4D.8 apresenta a distribuição dos gastos por quantil dos titulares. Pode-se perceber que os gastos hospitalares são mais concentrados do que os ambulatoriais. Cerca de 90% dos usuários realizaram algum gasto ambulatorial no período (exceto em 2002, quando cerca de 75% realizaram gasto). Por outro lado, verifica-se uma elevação na utilização dos serviços hospitalares: em 2001 apenas 10% dos usuários realizaram gastos nessa categoria, enquanto em 2002 e 2003 o percentual elevou-se para aproximadamente 25%. Cerca de 99% dos gastos ambulatoriais ficaram abaixo de R\$ 1.100 em 2001 e de R\$ 1 mil em 2002 e 2003. Nos casos de internação, 95% dos gastos ficaram entre R\$ 1 mil e R\$ 1.400 nos três anos.

Ao se separarem os gastos de todos os beneficiários por quantis verifica-se que, se por um lado o padrão de utilização dos gastos ambulatoriais difere dos titulares, por outro os gastos hospitalares apresentam um padrão similar. O percentual de beneficiários que não utilizam qualquer serviço ambulatorial é mais elevado que o de titulares: aqui ao menos 10% da população não realizam qualquer gasto, contra 5% dos titulares nos anos analisados. Os resultados são apresentados na Tabela 4D.9.

A Tabela 4D.10 apresenta os valores médios para os titulares que realizaram gastos positivos para o ano de 2003. Enquanto o gasto médio ambulatorial

**TABELA 4D.9  
DISTRIBUIÇÃO DOS GASTOS MÉDIOS DOS BENEFICIÁRIOS POR QUANTIS — 2001 A 2003**

Quantil (%)	2001		2002		2003	
	Ambulatorial	Internação	Ambulatorial	Internação	Ambulatorial	Internação
100 max	44.593,5	182.469,18	38.770,42	228.362,22	60.030,65	264.727,61
99	1.047,22	3.911,61	936,14	4.515,37	887,03	4.971,19
95	568,54	889,22	501,96	1.037,87	462,08	1.122,46
90	401,56	224,3	350,40	409,09	320,4	454,31
75	197,91	0	168,21	66,15	158,58	91
50 mediana	76,99	0	63,64	0	61,51	0
25	20	0	14,51	0	14,29	0
10	0	0	0	0	0	0
5	0	0	0	0	0	0
1	0	0	0	0	0	0
0 min	-138,38	-47,8	-4,03	-108,03	0	0

Fonte: Elaboração própria a partir dos dados da Sabesperv.

TABELA 4D.10  
GASTO MÉDIO DOS TITULARES QUE REALIZARAM GASTOS — 2003

Gastos	Número de observações	Média	Desvio-padrão	Mínimo	Máximo
Ambulatorial	17.573	179,28	364,68	1,08	17.935,44
Internação	8.779	764,69	2.897,2	2,05	68.570,51

Fonte: Elaboração própria a partir dos dados da Sabesperv.

não possui uma elevação significativa, o contrário acontece com o gasto médio hospitalar. A menor elevação nos gastos ambulatoriais se deve ao fato de esse tipo de gasto ser mais comum na população, enquanto os gastos com internação são mais raros. Os valores médios condicionados com internações são mais de duas vezes superiores ao gasto médio, reforçando mais uma vez a importância do seguro para esse tipo de componente e assimetria da distribuição desses gastos na população.

A Tabela 4D.11 apresenta os gastos médios dos beneficiários que realizaram gastos. Em um padrão similar ao dos titulares, os gastos médios condicionados apresentaram elevação: a dos gastos hospitalares foi mais acentuada, cerca de 144%, enquanto a dos gastos ambulatoriais foi de cerca de 20%.

TABELA 4D.11  
GASTO MÉDIO DOS BENEFICIÁRIOS QUE REALIZARAM GASTOS — 2003

Gastos	Número de observações	Média	Desvio-padrão	Mínimo	Máximo
Ambulatorial	54.907	162,78	566,24	0,37	60.030,65
Internação	26.903	805,48	4.368,51	0,07	264.727,61

Fonte: Elaboração própria a partir dos dados da Sabesperv.

#### 4D.2.3.3 Análise descritiva dos gastos segundo atributos de riscos individuais

Esta subseção realiza a análise descritiva dos gastos considerando os atributos de risco individuais. Os mais usuais são idade e sexo. Idade e sexo são atributos importantes, na medida em que são, na maior parte das vezes, observáveis pelas seguradoras sendo passíveis de incorporação nos contratos. A evidência empírica indica diferenciais nos gastos entre indivíduos, segundo essas duas variáveis demográficas, as que se distinguem entre subpopulações e países [ver Van de Ven e Ellis (2000)]. Para o caso brasileiro existem dados disponíveis que permitem inferir diferenças nos gastos e utilização segundo diferentes grupos de idade e sexo, corroborando a importância desses atributos para a análise do risco. A despeito dessas evidências não existem trabalhos que procurem mensurar a elasticida-

de dos gastos em relação a esses atributos. A análise dos gastos é realizada também segundo estado civil e escolaridade, neste último caso apenas para os titulares, dada a disponibilidade de informações no banco de dados.

Para analisar esses gastos foram construídos grupos de idade que apresentam perfis de gasto mais típico. Ribeiro (2005) utiliza o critério de taxas de utilização para definir os grupos etários. Idades com taxas de utilização estatisticamente iguais são incluídas no mesmo grupo possibilitando a identificação de pontos de corte que indiquem mudanças no perfil de utilização. A autora encontra 29 grupos etários para utilização de serviços ambulatoriais e 23 grupos para serviços hospitalares. De maneira geral, os indivíduos apresentam padrão de utilização diferente ano a ano na infância e na adolescência. A partir de uma certa idade, normalmente na fase adulta, os indivíduos podem ser incluídos no mesmo grupo por um período de tempo de cerca de cinco anos e esse padrão segue até a velhice. Com isso, procurou-se construir neste trabalho grupos de idade menores para as pessoas mais jovens, aumento da faixa de idade dos grupos a partir da adolescência: 0-2 (recém-nascidos, devido a maior freqüência nos gastos); 2-5 (infância); 5-15 (período de pré-adolescência); 15-25 (jovens); 25-40 (idade reprodutiva da mulher/idade ativa); 40-50; 50-60; 60-70; 70-80 e 80 ou mais (e a partir dos 40 anos, os grupos etários representam a evolução dos gastos na medida em que a população envelhece).<sup>52</sup> Neste trabalho, essas variáveis são modeladas através de *dummies* para cada grupo.

As Tabelas 4D.12 e 4D.13 e os Gráficos 4D.3 e 4D.4 apresentam o gasto médio ambulatorial para homens e mulheres beneficiários do plano de saúde da Sabesp, por grupos de idade, para o ano de 2003. O formato da curva em U tanto para homens como para mulheres corrobora os fatos estilizados já presentes na literatura empírica. Os gastos são mais elevados para crianças até dois anos de idade, seguido de uma tendência decrescente e voltam a subir para as mulheres após os 15 anos e para os homens após os 25 anos. A diferença mais relevante entre homens e mulheres ocorre exatamente no período reprodutivo da mulher. Isso pode ser justificado pelo acompanhamento preventivo mais rigoroso durante o período fértil feminino. No caso dos homens, os gastos começam a crescer no período de 25 a 40 anos, caracterizando o fato de indivíduos do sexo masculino iniciarem medidas preventivas em idade mais avançada do que as mulheres. Os grupos etários que apresentam os maiores valores para gastos ambulatoriais são os de menores de dois anos e os com idade mais avançada, comprovando os estudos que evidenciam esses grupos como os que mais utilizam os serviços de saúde. Para

52. No banco de dados dos titulares a análise é feita a partir do grupo de idade de 15 a 25 anos devido à população ativa na Sabesp.

ambos os sexos, os gastos ambulatoriais diminuem após os 80 anos, sugerindo que nessa faixa etária os indivíduos realizam mais gastos de internação.

**TABELA 4D.12  
GASTO MÉDIO AMBULATORIAL DAS MULHERES BENEFICIÁRIAS DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
0-1	82	180,03	161,09	0	888,37
2-4	1.055	98,97	105,65	0	858,18
5-14	5.161	73,03	136,49	0	3.867,09
15-24	6.484	94,49	168,61	0	4.462,54
25-39	6.751	154,81	245,93	0	8.775,39
40-49	6.396	206,32	810,26	0	41.846,05
50-59	4.071	218,88	677,11	0	27.561,04
60-69	1.757	283,57	1.361,16	0	36.507,38
70-79	1.287	224,13	474,91	0	13.582,39
80 +	683	204,17	557,78	0	9.098,78

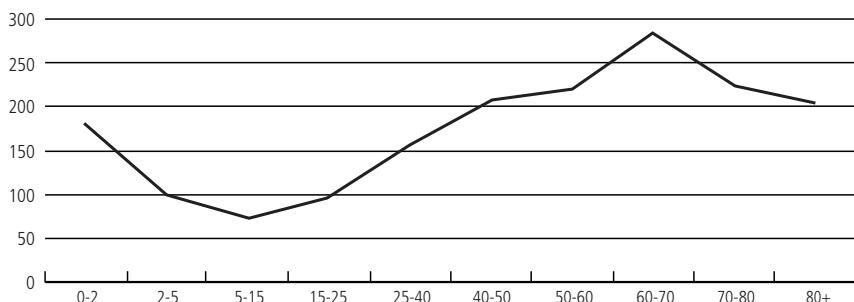
Fonte: Elaboração própria a partir dos dados da Sabesperv.

**TABELA 4D.13  
GASTO MÉDIO AMBULATORIAL DOS HOMENS BENEFICIÁRIOS DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
0-1	82	153,11	136,74	0	589,12
2-4	1.076	107,91	140,62	0	2.316,19
5-14	5.423	79,36	167,24	0	5.405,06
15-24	6.486	58,41	131,11	0	4.746,40
25-39	6.349	114,74	780,85	0	60.031,66
40-49	6.127	135,70	236,25	0	6.436,67
50-59	4.377	168,68	385,70	0	13.656,65
60-69	1.300	193,12	728,26	0	17.936,44
70-79	559	313,50	1.479,53	0	27.924,37
80 +	262	192,50	328,96	0	2.750,00

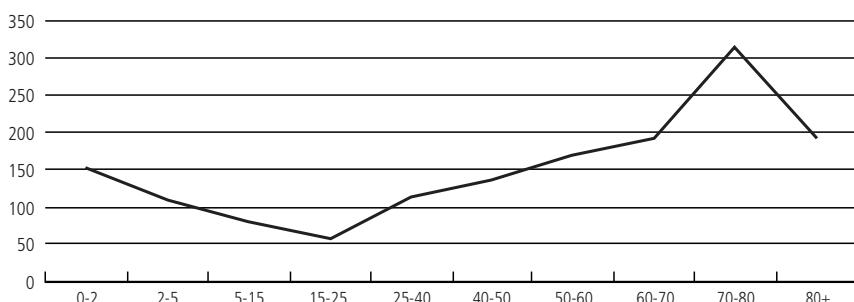
Fonte: Elaboração própria a partir dos dados da Sabesperv.

GRÁFICO 4D.3

**GASTO AMBULATORIAL MÉDIO DAS MULHERES BENEFICIÁRIAS DE 2003**

Fonte: Elaboração própria a partir de dados da Sabesperv.

GRÁFICO 4D.4

**GASTO AMBULATORIAL MÉDIO DOS HOMENS BENEFICIÁRIOS DE 2003**

Fonte: Elaboração própria a partir de dados da Sabesperv.

Os gastos médios com internação por grupos etários para o ano de 2003 são apresentados nas Tabelas 4D.14 e 4D.15 e nos Gráficos 4D.5 e 4D.6. Verifica-se que esses gastos são elevados para crianças menores de 2 anos, caindo até o grupo de idade de 5 a 15 anos, e apresentando um comportamento crescente após essa idade, mas com níveis baixos. Os gastos com internação, até o grupo de idade de 40 a 50 anos, são maiores para o sexo feminino o que pode estar associado às despesas com internações obstétricas. Após esse grupo etário, os gastos com internação crescem significativamente para ambos os sexos.

Ao serem analisados os gastos segundo atributos de risco para a carteira de titulares do plano de saúde da Sabesperv, verifica-se que para esse tipo de popu-

**TABELA 4D.14  
GASTO MÉDIO COM INTERNAÇÃO DAS MULHERES BENEFICIÁRIAS DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
0-1	82	3.438,13	29.220,96	0	264.728,60
2-4	1.055	193,82	2.201,92	0	67.467,38
5-14	5.161	77,76	931,72	0	59.675,93
15-24	6.484	142,56	830,73	0	35.735,36
25-39	6.751	310,23	1.305,13	0	71.757,16
40-49	6.396	365,43	2.031,20	0	93.700,13
50-59	4.071	398,44	1.669,47	0	45.864,73
60-69	1.757	802,43	4.278,67	0	109.973,90
70-79	1.287	1.039,69	4.610,31	0	88.736,15
80 +	683	1.867,33	7.265,41	0	105.449,80

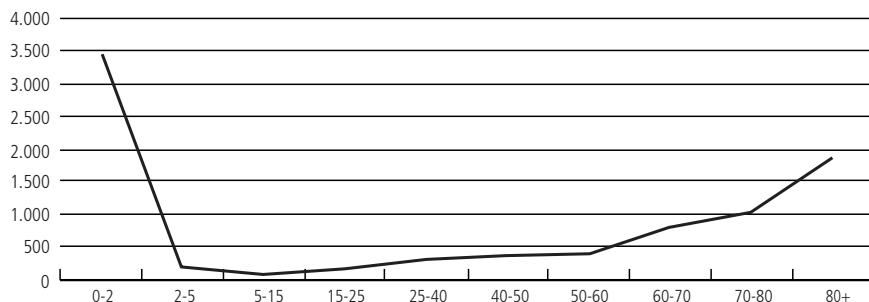
Fonte: Elaboração própria a partir dos dados da Sabesperv.

**TABELA 4D.15  
GASTO MÉDIO COM INTERNAÇÃO DOS HOMENS, BENEFICIÁRIOS DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
0-1	82	140,12	317,44	0	2.237,92
2-4	1.076	189,61	1.371,27	0	40.514,28
5-14	5.423	97,54	1.151,38	0	59.797,59
15-24	6.486	101,82	799,73	0	37.572,75
25-39	6.349	187,17	1.203,73	0	57.374,50
40-49	6.127	273,69	1.679,46	0	68.571,51
50-59	4.377	459,78	4.272,39	0	242.740,50
60-69	1.300	840,57	4.204,16	0	71.022,99
70-79	559	2.631,54	10.696,60	0	131.714,90
80 +	262	3.817,12	17.269,12	0	204.869,20

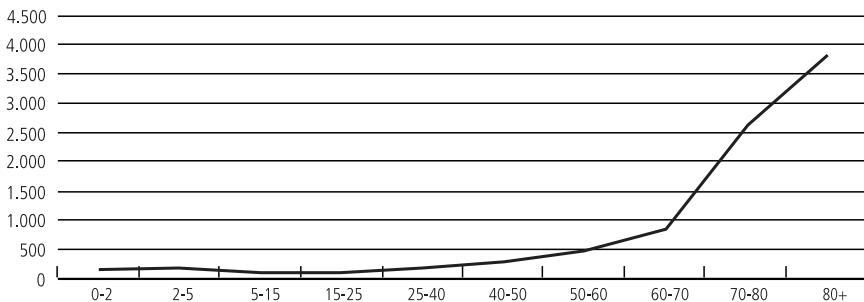
Fonte: Elaboração própria a partir dos dados da Sabesperv.

**GRÁFICO 4D.5  
GASTO MÉDIO COM INTERNAÇÃO DAS MULHERES BENEFICIÁRIOS DE 2003**



Fonte: Elaboração própria a partir de dados da Sabesperv.

**GRÁFICO 4D.6  
GASTO MÉDIO COM INTERNAÇÃO DOS HOMENS BENEFICIÁRIOS DE 2003**



Fonte: Elaboração própria a partir de dados da Sabesperv.

lação o perfil da curva de gasto médio é bastante distinto. O gasto médio ambulatorial é crescente até o grupo de idade de 50 a 60 anos e decresce no último grupo etário. Esse resultado reflete o aumento nos gastos de internação nas idades mais avançadas. Constatase a superioridade dos gastos femininos em todos os grupos etários, refletindo a maior preocupação das mulheres com cuidados preventivos (ver Tabelas 4D.16 e 4D.17 e Gráficos 4D.7 e 4D.8).

Em relação ao gasto médio com internação (ver Tabelas 4D.18 e 4D.19 e Gráficos 4D.9 e 4D.10) observa-se que esses gastos são crescentes com a idade para ambos os sexos, e à medida que a idade aumenta, essa tendência de elevação fica mais acentuada. Os gastos médios das mulheres são mais elevados que os dos homens, o que pode estar associado a internações por parto e gravidez.

**TABELA 4D.16  
GASTO MÉDIO AMBULATORIAL DAS MULHERES TITULARES DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
15-24	115	182,87	179,00	0	880,79
25-39	1.317	239,03	257,30	0	3.648,65
40-49	1.276	293,47	627,03	0	14.401,04
50-59	461	335,78	678,99	0	12.635,56
60 +	43	198,35	210,78	0	772,90

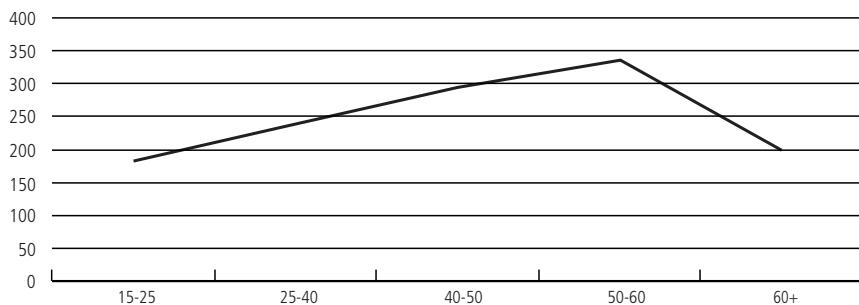
Fonte: Elaboração própria a partir dos dados da Sábesprev.

**TABELA 4D.17  
GASTO MÉDIO AMBULATORIAL DOS HOMENS TITULARES DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
15-24	200	97,43	122,30	0	836,70
25-39	5.059	113,14	162,14	0	2.749,56
40-49	5.382	137,26	211,32	0	5.483,08
50-59	3.554	175,51	406,04	0	13.656,65
60 +	538	171,80	434,09	0	8.479,95

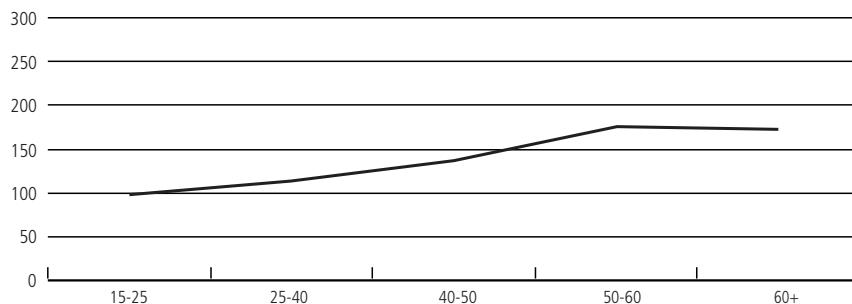
Fonte: Elaboração própria a partir dos dados da Sábesprev.

**GRÁFICO 4D.7  
GASTO AMBULATORIAL MÉDIO DAS MULHERES TITULARES DE 2003**



Fonte: Elaboração própria a partir de dados da Sábesprev.

**GRÁFICO 4D.8  
GASTO AMBULATORIAL MÉDIO DOS HOMENS TITULARES DE 2003**



Fonte: Elaboração própria a partir de dados da Sabesperv.

**TABELA 4D.18  
GASTO MÉDIO COM INTERNAÇÃO DAS MULHERES TITULARES DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
15-24	115	159,56	505,87	0	4.283,23
25-39	1.317	350,90	880,70	0	13.011,92
40-49	1.276	470,57	1.761,97	0	28.824,23
50-59	461	611,53	1.572,115	0	14.213,61
60 +	43	1.107,919	3.837,90	0	24.989,28

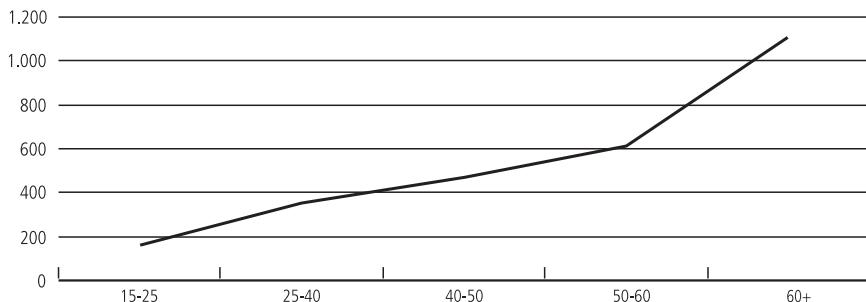
Fonte: Elaboração própria a partir dos dados da Sabesperv.

**TABELA 4D.19  
GASTO MÉDIO COM INTERNAÇÃO DOS HOMENS TITULARES DE 2003**

Grupo de idade	Observação	Média	Desvio-padrão	Mínimo	Máximo
15-24	200	158,71	694,20	0	7.077,23
25-39	5.059	199,00	1.301,05	0	57.374,5
40-49	5.382	265,86	1.582,66	0	68.571,51
50-59	3.554	394,10	2.075,484	0	61.896,1
60 +	538	1.066,27	5.097,85	0	64.342,27

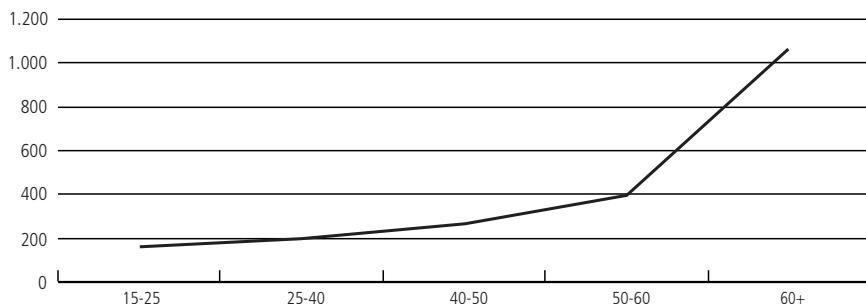
Fonte: Elaboração própria a partir dos dados da Sabesperv.

**GRÁFICO 4D.9  
GASTO COM INTERNAÇÃO MÉDIO DAS MULHERES TITULARES DE 2003**



Fonte: Elaboração própria a partir de dados da Sabesperv.

**GRÁFICO 4D.10  
GASTO COM INTERNAÇÃO MÉDIO DOS HOMENS TITULARES DE 2003**



Fonte: Elaboração própria a partir de dados da Sabesperv.

As Tabelas 4D.20 e 4D.21 apresentam os valores médios dos gastos para homens e mulheres titulares do plano de saúde segundo o grau de escolaridade. O perfil dos gastos médios é bastante distinto quando se consideram as categorias de gasto entre os sexos: de maneira geral, para os homens em todos os três anos os gastos com internação são decrescentes conforme aumenta o grau de escolaridade, enquanto os gastos ambulatoriais são relativamente estáveis. Gastos ambulatoriais médios similares entre os grupos de escolaridade justificam o valor da informação:

é razoável supor que pessoas convivendo em um mesmo ambiente tenham acesso ao mesmo tipo de informação, ao menos quando esta parte de uma política institucional da empresa no sentido de prevenir problemas de saúde. Porém, a relação decrescente entre escolaridade e gastos com internação sugere que os homens mais escolarizados têm um melhor estado de saúde seja devido ao cuidado preventivo em sua vida pregressa, ou porque sabem realizar uma melhor prevenção, por valorizarem as informações a que têm acesso. Para as mulheres o padrão de gastos ambulatoriais é próximo ao dos homens, apresentando valores médios relativamente estáveis entre os grupos de escolaridade. Os gastos com internação também apresentam um comportamento decrescente de acordo com a escolaridade, ocorrendo um aumento no grupo das mais escolarizadas.

**TABELA 4D.20  
GASTO MÉDIO POR ESCOLARIDADE DOS TITULARES DO SEXO MASCULINO — 2001 A 2003**

	2001		2002		2003	
	Ambula-torial	Intenção	Ambula-torial	Intenção	Ambula-torial	Intenção
Analfabeto/primário incompleto	201,36	385,39	174,45	500,14	145,54	571,24
Primário completo/1º grau incompleto	176,51	307,47	128,05	360,98	134,20	489,75
1º grau completo/2º grau incompleto	154,94	235,09	131,72	240,52	127,93	269,20
2º grau completo/ 3º grau incompleto	164,43	169,22	146,69	184,99	137,26	239,94
3º grau completo	202,13	157,86	180,33	215,93	177,19	278,59

Fonte: Elaboração própria a partir dos dados da Sabesperv.

**TABELA 4D.21  
GASTO MÉDIO POR ESCOLARIDADE DOS TITULARES DE SEXO FEMININO — 2001 A 2003**

	2001		2002		2003	
	Ambula-torial	Internação	Ambula-torial	Internação	Ambula-torial	Internação
Analfabeto/primário incompleto	322,54	108,64	327,97	332,32	240,04	1.353,21
Primário completo/1º grau incompleto	360,30	384,88	234,31	481,63	231,18	631,30
1º grau completo/2º grau incompleto	314,88	155,42	271,66	324,34	241,13	744,50
2º grau completo/3º grau incompleto	300,80	237,38	282,41	329,89	243,07	425,57
3º grau completo	335,33	243,06	333,66	413,05	316,25	357,60

Fonte: Elaboração própria a partir dos dados da Sabesperv.

Quanto ao estado civil tanto para homens como para mulheres, ser casado eleva os gastos ambulatoriais e de internação (Tabelas 4D.22 a 4D.25). Isso provavelmente estará refletindo uma preocupação maior com o estado de saúde, dado que os laços familiares alteram o valor da saúde e o controle sobre a mesma. Além disso, os indivíduos solteiros apresentam média de idade inferior aos casados.

TABELA 4D.22

**GASTOS MÉDIOS SEGUNDO O ESTADO CIVIL DOS TITULARES DE SEXO MASCULINO — 2001 A 2003**

Gastos	2001			2002			2003		
	Solteiros	Casados	Outros	Solteiros	Casados	Outros	Solteiros	Casados	Outros
Ambulatorial	154,67	179,91	175,34	128,16	151,82	145,28	122,99	145,83	141,64
Internação	257,26	212,58	275,60	165,53	283,38	257,46	191,67	357,74	326,08

Fonte: Elaboração própria a partir dos dados da Sabesperv.

TABELA 4D.23

**GASTOS MÉDIOS SEGUNDO O ESTADO CIVIL DOS TITULARES DO SEXO FEMININO — 2001 A 2003**

Gastos	2001			2002			2003		
	Solteiros	Casados	Outros	Solteiros	Casados	Outros	Solteiros	Casados	Outros
Ambulatorial	315,04	318,71	343,63	281,57	317,96	301,28	261,16	269,26	274,78
Internação	151,11	293,07	249,88	262,74	459,96	324,22	352,44	499,21	535,53

Fonte: Elaboração própria a partir dos dados da Sabesperv.

TABELA 4D.24

**GASTOS MÉDIOS SEGUNDO O ESTADO CIVIL DOS BENEFICIÁRIOS DO SEXO MASCULINO — 2001 A 2003**

Gastos	2001			2002			2003		
	Solteiros	Casados	Outros	Solteiros	Casados	Outros	Solteiros	Casados	Outros
Ambulatorial	91,78	181,63	199,19	79,26	152,71	151,62	82,00	146,15	156,74
Internação	129,01	284,13	455,96	144,00	389,26	376,54	142,11	476,57	434,68

Fonte: Elaboração própria a partir dos dados da Sabesperv.

TABELA 4D.25  
**GASTOS MÉDIOS SEGUNDO O ESTADO CIVIL DOS BENEFICIÁRIOS DO SEXO FEMININO — 2001 A 2003**

Gastos	2001			2002			2003		
	Solteiros	Casados	Outros	Solteiros	Casados	Outros	Solteiros	Casados	Outros
Ambulatorial	120,68	240,37	286,58	108,11	207,30	231,83	105,95	192,50	215,58
Internação	121,97	294,35	472,00	130,03	396,23	637,27	184,16	442,62	689,42

Fonte: Elaboração própria a partir dos dados da Sabesperv.

### 4D.3 METODOLOGIA

O banco de dados da Sabesperv possui informações dos usuários para três anos: 2001, 2002 e 2003. O modelo utilizado neste trabalho é devido a Duan *et alii* (1983): o Modelo de Quatro Equações (MQE). Os gastos com serviços de saúde são estimados através de duas metodologias distintas: uma análise em *cross-section* para o ano de 2003 e uma análise em painel para os três anos disponíveis. Uma das principais vantagens da análise em painel é, sob certas condições, eliminar o problema das variáveis omitidas. A principal heterogeneidade não-observável dos gastos em saúde é o estado de saúde passado dos indivíduos. Por outro lado, dada a importância dessa variável na literatura empírica [ver Wagstaff (1986) e Van de Ven e Ellis (2000), por exemplo], buscaram-se estimativas com os dados em *cross-section* que incorporassem indicadores dessa variável. Sendo assim, a estimação da *cross-section* para o ano de 2003 justifica-se na medida em que nela procurou-se incorporar variáveis indicadoras do estado de saúde passada dos indivíduos. Como apresentado na tipologia dos gastos, as estimativas serão efetuadas considerando-se os gastos decompostos em ambulatoriais e de internação. No restante desta seção são apresentados o modelo utilizado, a interpretação dos coeficientes e os testes realizados neste trabalho.

#### 4D.3.1 O modelo de quatro equações

Neste trabalho estimamos a demanda por serviços de saúde a partir de um MQE, seguindo a proposta de Duan *et alii* (1983). O comportamento dos indivíduos pode ser modelado por meio de três decisões distintas: uma é, a de procurar pelo serviço de saúde pelo próprio usuário e as outras são decisões do tipo de gasto a ser realizado e de quanto gastar, tomadas junto ao médico.

A primeira equação do modelo é uma equação *probit* para a probabilidade de utilização de qualquer serviço médico durante o ano, ou seja, realização de gastos positivos:<sup>53</sup>

$$\Pr(MED_i > 0) = \Phi(x_i\beta_1) \quad (1)$$

A segunda equação é uma equação *probit* para a probabilidade condicional de os indivíduos que utilizaram algum serviço médico durante o ano terem ao menos uma internação:

$$\Pr(INP_i | MED_i > 0) = \Phi(x_i\beta_2) \quad (2)$$

A terceira equação é a regressão linear logarítmica para os gastos médicos somente para usuários de serviços ambulatoriais:

$$\ln(MED_i | MED_i > 0, INP_i = 0) = x_i\beta_3 + \varepsilon_{3i} \quad (3)$$

A quarta equação é a regressão linear logarítmica dos gastos médicos para os usuários de serviços de internação:

$$\ln(MED_i | INP_i > 0) = x_i\beta_4 + \varepsilon_{4i} \quad (4)$$

Essa abordagem possui duas vantagens principais. Em primeiro lugar permite separar os gastos dos usuários em relação a duas características de não normalidade da distribuição: a de que uma parcela significativa da população não utiliza nenhum serviço médico durante um período determinado e a da heterogeneidade dos gastos médicos dos usuários de consultas e internações. Em segundo lugar, possibilita analisar os gastos ambulatoriais e de internação separadamente. Essa divisão é importante devido à assimetria existente na distribuição de gastos com a saúde, em que grande parcela de indivíduos usuários de serviços ambulatoriais realiza gastos pouco significativos em determinado ano e alguns poucos usuários representam grande parcela dos gastos anuais com internação.

---

53. Para uma descrição mais detalhada do modelo *probit* e do modelo de OLS aplicados a uma *cross-section* e a dados em painel, ver Greene (2003) e Wooldridge (2002).

#### 4D.3.1.1 Interpretação dos coeficientes no modelo *probit*

Os parâmetros do modelo *probit* não podem ser interpretados diretamente. A interpretação do modelo é retirada do cálculo dos efeitos marginais das variáveis explicativas sobre a variável dependente. Para o modelo *probit* o efeito marginal é dado por:

$$\frac{\partial E[y|x]}{\partial x} = \phi(x'\beta)\beta \quad (5)$$

Em (5) a derivada parcial do valor esperado de  $y$  (condicionado a  $x$ ), em relação a um atributo individual  $x$ , gera o efeito marginal do atributo sobre a probabilidade de realização de gastos positivos. Ou seja,  $\phi(x'\beta)\beta$  dá o efeito do atributo  $x$  na probabilidade de realização de gasto positivo, tudo o mais constante. O sinal da derivada parcial dá o sentido do efeito (se aumento ou queda na probabilidade da realização de gastos).

Para o cálculo dos efeitos marginais existem duas opções. Ou as expressões são avaliadas na média dos dados amostrais ou o efeito marginal é avaliado para todas as observações e utiliza-se a média dos efeitos marginais individuais. Conforme Greene (2003), os resultados são similares em grandes amostras. Outra complicação com o cálculo dos efeitos marginais são as variáveis *dummies* que, por assumirem valores discretos, não se podem aplicar às derivadas. Nesse caso, o autor mostra que o correto seria calcular o efeito marginal da seguinte forma:

$$\text{Prob}\left[Y = 1 \mid \bar{x}_{(d)}, d = 1\right] - \text{Prob}\left[Y = 1 \mid \bar{x}_{(d)}, d = 0\right] \quad (6)$$

onde  $\bar{x}_{(d)}$  denota a média de todas as outras variáveis explicativas. A equação (6) dá a diferença entre a probabilidade de a variável dependente assumir o valor 1, quando avaliada na média dos dados amostrais, com a variável *dummy* assumindo o valor 1 ou 0.

Na prática, conforme destacado por Greene (2003), tomar a derivada de uma variável binária como se fosse contínua é uma aproximação freqüentemente utilizada e que conduz a resultados similares. Neste trabalho os resultados são apresentados para o indivíduo médio aplicando-se a média dos dados amostrais. Além disso, toma-se a derivada das variáveis *dummies*.

#### 4D.3.1.2 Testes realizados para o modelo *probit*

##### a) Significância dos parâmetros

Para o teste de significância conjunta dos parâmetros,<sup>54</sup> realizamos o teste de Wald no modelo robusto, que consiste em rejeitar a hipótese nula de que todos os parâmetros são estatisticamente iguais a 0:

$$\begin{cases} H_0 : R\beta_i = 0 \\ H_1 : R\beta_i \neq 0 \end{cases}$$

O teste segue uma distribuição qui-quadrado com  $K$  graus de liberdade, onde  $K$  é o número de variáveis independentes.

##### b) Testes de especificação

A ocorrência de um problema de erro de especificação no modelo se deve à existência de variáveis relevantes omitidas. O problema nesse caso é que, mesmo a variável omitida não sendo correlacionada com a variável incluída no modelo, o coeficiente desta última será inconsistente. Um teste para a presença de variáveis omitidas é baseado na idéia de que, se a regressão é corretamente especificada nenhuma variável independente significativa pode ser adicionada.<sup>55</sup>

A presença de heterocedasticidade torna os parâmetros estimados pela função de verossimilhança inconsistentes e a matriz de covariância assintótica inapropriada. Uma tentativa de correção do problema consiste em estimar o modelo heterocedástico, indicando qual variável (ou conjunto delas) suspeita-se que está causando a heterocedasticidade, e verificar a significância dos parâmetros no teste LR (a hipótese nula é de variância constante). O teste consiste em aceitar ou rejeitar a hipótese nula no nível de significância desejado. Outra tentativa de correção, adotada na prática devido à dificuldade de se encontrar qual variável ou combinação de variáveis está causando a heterocedasticidade, é estimar o modelo robusto. Neste trabalho a análise foi realizada a partir do modelo robusto.

54. O teste de significância dos coeficientes individuais no modelo *probit* utiliza a tabela normal padrão e consiste na seguinte hipótese:

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

Rejeito  $H_0$  quando a área acima (ou abaixo) do valor calculado, que indica a probabilidade  $P > |\mathbf{z}|$ , for inferior ao valor tabelado.

55. No *Stata* (o pacote econômétrico utilizado) esse teste é denominado *linktest*. O teste cria duas variáveis *\_hat* e *\_hatsq* e reajusta o modelo reutilizando-as. A variável *\_hat* deve ser significativa dado que ela é o valor predito dos gastos. Por outro lado, se o modelo está corretamente especificado a variável *\_hatsq* não pode ser significativa, porque o quadrado do valor predito não pode, então, ter poder explicatório. Caso tivesse poder de explicar a probabilidade dos gastos, significaria que algo que não está incluído em nossa previsão, em nosso valor predito, justifica a probabilidade de realização de gastos. Com isso devemos verificar o valor-p de *\_hatsq*.

### c) Qualidade do ajuste

Uma medida similar ao  $R^2$  (da regressão de mínimos quadrados ordinários) é o índice da razão de verossimilhança ou pseudo  $R^2$ , cuja fórmula é dada por:

$$PseudoR^2 = 1 - \frac{\ln \hat{L}_U}{\ln \hat{L}_R} \quad (7)$$

onde  $\hat{L}_R$  é a função de log - verossimilhança para o modelo restrito e  $\hat{L}_U$  para o modelo estimado. A medida varia de 0 a 1, com o valor 1 indicando um ajuste perfeito.

Outra medida da qualidade do ajuste é o percentual dos valores corretamente preditos. A medida consiste em calcular a probabilidade de ocorrência de resultado positivo (igual a 1) para cada indivíduo. Se a probabilidade é maior que 0,5 a previsão para o indivíduo é 1 e se for menor que 0,5 a probabilidade é 0. Essa medida, entretanto, pode dar uma falsa impressão de um bom ajuste, uma vez que se algum resultado é muito mais provável que outro, a medida pode reportar um alto percentual de acertos (basta que ocorram muitos acertos para o resultado provável), mesmo que não ocorram acertos para o resultado cuja freqüência é menor.

#### 4D.3.1.3 Interpretação dos coeficientes no modelo de OLS

O valor do coeficiente  $b_i$  representa a derivada parcial da variável dependente em relação a determinado atributo, tudo o mais constante. Ou seja, medimos quanto o logaritmo dos gastos varia quando o atributo se modifica, mantendo todos os demais atributos constantes:

$$\frac{\partial \ln y}{\partial x_i} = \beta_i \quad (8)$$

Os gastos com serviços ambulatoriais e hospitalares representam, em certa medida, a demanda dos serviços em saúde.<sup>56</sup> Assim, buscamos interpretar os

---

56. Os gastos representam a demanda atendida e, em certo aspecto, é uma simplificação para a demanda por serviços de saúde. Quanto mais universalizada for a oferta dos serviços por parte dos gestores, melhor os gastos representarão a demanda real por parte dos usuários do plano.

coeficientes das regressões de OLS como sendo a elasticidade da demanda por serviços de saúde em relação ao atributo analisado. O parâmetro estimado seria uma representação direta da elasticidade da demanda apenas se tomássemos os atributos em logaritmos.<sup>57</sup>

Como esse não é o caso, em modelos log-lineares a variável dependente é analisada em valores percentuais e a variável independente contínua em valores absolutos. Assim, o coeficiente  $\beta_i$  reflete a variação percentual no logaritmo dos gastos ambulatoriais ou hospitalares, quando o atributo em referência varia em uma unidade. Para as variáveis independentes binárias o coeficiente reflete a variação percentual no logaritmo dos gastos em saúde, na presença do efeito positivo captado pela *dummy*. O sinal do coeficiente indica se o efeito analisado aumenta ou diminui os gastos.

#### 4D.3.1.4 Testes realizados no modelo OLS

Além dos tradicionais testes t e F para significância dos parâmetros e a análise do  $R^2$  e do  $R^2$  ajustado para a qualidade do ajuste, neste trabalho é realizado o teste de Cook-Weisberg para detectar a presença de heterocedasticidade. O teste consta em aceitar ou rejeitar a hipótese nula de variância constante nos parâmetros e segue uma estatística qui-quadrado com um grau de liberdade.

Para analisar o problema das variáveis omitidas além do teste de adição de variáveis omitidas igual ao realizado para o *probit*, aqui é efetuado o RESET (Regression Specification Error Test). A hipótese nula do teste é a ausência de variáveis omitidas. O teste segue uma estatística-F com 3 graus de liberdade no numerador e  $n-K-3$  no denominador.

A presença de colinearidade perfeita entre as variáveis explicativas torna a variância do coeficiente estimado infinita. Uma alta colinearidade entre as variáveis leva a vários erros na estimação dos coeficientes como: pequenas mudanças nos dados gerando grandes variações nos coeficientes estimados; coeficientes com desvios-padrão elevados e baixo nível de significância; e magnitudes e sinais encontrados implausíveis com a teoria subjacente ao modelo. Sendo assim, foi realizado um teste para detectar a presença de multicolinearidade entre as variáveis, reportado no apêndice após as regressões. A variável independente pode estar trazendo o problema para o modelo sempre que o valor de 1/VIF for menor ou igual a 0,10.

---

57. O que é inviável com variáveis binárias. Nesse caso teríamos então que:  $\frac{\partial \ln y}{\partial \ln x_i} = \beta_i$ . Para essa discussão, ver Varian (1994).

### 4D.3.2 Análise em painel

Na economia da saúde a principal heterogeneidade não-observável a afetar o comportamento dos indivíduos em relação à demanda pelos serviços é o estado de saúde passado dos mesmos. Este seria o principal efeito (ou variável omitida) não captado diretamente em nossa análise a afetar a demanda pelos serviços.<sup>58</sup> A vantagem principal da análise de painel é eliminar, sob certas hipóteses, o problema das variáveis omitidas<sup>59</sup> gerando estimadores consistentes a partir das regressões realizadas.

#### 4D.3.2.1 Interpretação dos coeficientes no *probit* para dados longitudinais

Para os modelos em painel a interpretação dos coeficientes é semelhante à de uma *cross-section*. O efeito marginal é dado por:

$$\frac{\partial E[y|x]}{\partial x} = \Phi(x'\beta)\beta \quad (9)$$

Em (9) novamente  $\Phi(x'\beta)\beta$  representa o efeito do atributo  $x$  na probabilidade de realização de gasto positivo, tudo o mais constante, com o sinal indicando o sentido do efeito.

#### 4D.3.2.2 Testes realizados no *probit*

Para o modelo *probit* em painel é apresentado o teste de significância individual dos parâmetros e o teste de Wald para a significância conjunta, que segue uma estatística qui-quadrado com graus de liberdade igual ao número de parâmetros a serem testados. Outra questão relevante seria a verificação de qual modelo se ajusta melhor aos nossos dados: o modelo de efeitos fixos ou o modelo de efeitos aleatórios. A questão nesse caso é que os efeitos fixos não podem ser tratados como parâmetros incidentais sem viesar os outros coeficientes do modelo, desde que o número de observações seja superior ao número de períodos. O modelo de efeitos fixos necessita, com isso, de um número elevado de períodos para que os parâmetros estimados sejam consistentes. Esse é o problema dos parâmetros incidentais. Quão sério seria esse viés permanece uma ques-

58. Apesar de criarmos variáveis que tentam captar o estado de saúde passado dos indivíduos, estas são apenas uma aproximação (ou *proxy*).

59. A principal simplificação neste caso é considerar a variável omitida constante no tempo. Apesar dessa limitação, podemos considerar que o estado de saúde das pessoas não terá mudanças bruscas no período analisado de três anos (exceto devido a choques exógenos).

tão em aberto na literatura.<sup>60</sup> Neste trabalho, portanto, estimamos diretamente o modelo de efeitos aleatórios.

#### 4D.3.2.3 Interpretação dos coeficientes no OLS para dados longitudinais

Novamente os coeficientes são interpretados de maneira similar à *cross-section*:

$$\frac{\partial \ln y}{\partial x_i} = \beta_i \quad (10)$$

Em (10)  $\beta_i$  indica a variação percentual no logaritmo dos gastos quando o atributo em referência varia em uma unidade e, no caso das variáveis *dummies*, indica a variação percentual no logaritmo dos gastos na presença do efeito positivo. O sinal do coeficiente indica o sentido do efeito da variável sobre os gastos.

#### 4D.3.2.4 Testes realizados no OLS

Neste trabalho é analisado o teste t de significância individual, o teste de Wald de significância conjunta dos parâmetros e o  $R^2$  ajustado para a qualidade do ajuste. Seguindo as estimativas das regressões *probit*, estimamos diretamente o modelo de efeitos aleatórios.

Na próxima seção descrevemos a especificação das variáveis explicativas utilizadas nas regressões em *cross-section* e painel e apresentamos os resultados do modelo.

### 4D.4 ESPECIFICAÇÃO DO MODELO E RESULTADOS

Nesta seção são apresentados os resultados do MQE. Foram utilizados três grupos de variáveis para buscar explicar as decisões de gastos e de quanto gastar: variáveis socioeconômicas, demográficas e do estado de saúde. O efeito esperado dessas variáveis do modelo é apresentado a seguir.

A maior parte dessas informações foi modelada sob a forma de variáveis *dummies*: uma *dummy* para sexo, tendo como categoria de referência as mulheres; uma *dummy* para tipo de plano, tendo como categoria de referência o plano

---

60. Essa discussão está colocada em Greene (2003). O problema dos parâmetros incidentais aparece nas regressões de variáveis dependentes discretas porque utilizamos nesse caso os dados originais. Nos modelos de OLS estimamos as regressões em painel pelos desvios dos resultados em relação ao valor médio das variáveis, o que resulta em  $\bar{y}_i$  ser uma estatística mínima suficiente. Essa estatística não está disponível para o modelo *probit*. Chamberlain (1980), conforme destacado por Greene (2003), deriva uma função de verossimilhança condicional para o modelo *logit* livre de parâmetros incidentais, permitindo a regressão do modelo *logit* para efeitos fixos. Tal análise, entretanto, foge ao escopo deste trabalho.

pleno; duas *dummies* de estado civil separando os usuários em três grupos: solteiros, casados e um terceiro grupo de referência que engloba viúvos, desquitados e outros; quatro *dummies* de escolaridades obtidas a partir da descrição do grau de instrução, dividindo-se os titulares em cinco grupos: analfabetos e primário incompleto (categoria de referência); primeiro grau completo e segundo grau incompleto; segundo grau completo e terceiro grau incompleto; e terceiro grau completo.

Variáveis explicativas usadas nas equações:

*a)* Gênero: segundo a literatura, o padrão de consumo entre homens e mulheres é distinto, por isso é necessário verificar se as mulheres utilizam mais o serviço de saúde que os homens. Utiliza-se a variável *dummy sex1* = 1 (se for homem) e 0 caso contrário.

*b)* Idade: espera-se que à medida que a idade aumenta, aumente também a utilização dos serviços de saúde. As idades são representadas pelos grupos etários de 1 a 10 mencionados anteriormente. Utilizamos as interações entre sexo e grupos de idade, dadas pelas seguintes *dummies*: *idh1* a *idh10* (sexo masculino e grupos de idade de 1 a 10) e *idm1* a *idm10* (sexo feminino e grupos de idade de 1 a 10).

*c)* Renda do titular do plano de saúde: variável incluída apenas no modelo dos titulares e representada pela variável *ren*. A princípio o efeito da renda sobre os gastos é ambíguo, uma vez que a variável influencia de forma mais dispersa a decisão de cuidado com a saúde. Por um lado, quanto maior a renda, maior o nível de informação que as pessoas possuem e maior a utilização de serviços de saúde de caráter preventivo. Além disso, os indivíduos são não-saciados em serviços de saúde, existindo indução da demanda por parte do paciente.<sup>61</sup> Por outro lado, o custo de oportunidade dos grupos socioeconômicos mais elevados é alto, de forma que o sacrifício que esses indivíduos têm de fazer para a aquisição desses bens e serviços de saúde é cada vez maior. Além disso, esse custo depende do vínculo empregatício desses agentes. No caso dos titulares da Sabesperv todos possuem carteira assinada, então, a princípio, faltas por eventos em saúde são justificadas. Dependendo do cargo, entretanto, torna-se mais difícil a ausência na empresa seja por motivos relacionados a responsabilidades assumidas, seja por motivos morais, éticos etc.

*d)* Anos de estudo: tenta captar qual o efeito que o grau de instrução tem em relação a quanto gastar e o valor médio esperado desse gasto. O aumento do nível educacional pode levar a uma utilização maior dos serviços de saúde, princi-

---

61. Serviços ofertados dentro da restrição orçamentária do indivíduo sempre serão demandados quando o mesmo se depara com algum evento em saúde.

palmente se forem de caráter preventivo. O efeito esperado é diferente, dependendo do tipo de cuidado. Para o modelo hospitalar, pode ser que o aumento do nível de instrução leve à redução dos gastos, mas para o ambulatorial, pode haver um aumento nos gastos, dado o caráter preventivo.

e) Tipo de plano de saúde: a inclusão da variável *plano* pretende verificar se a maior cobertura leva a um aumento na utilização dos serviços de saúde. Essa variável é incluída apenas para os beneficiários, pois estes têm a possibilidade de obter o plano básico ou o especial, diferentemente dos titulares da Sabesperv que possuem o plano pleno e não têm possibilidade de escolha.

f) Estado civil: espera-se que a utilização dos serviços de saúde seja menor entre os solteiros (*sol*) e maior entre os casados (*so2*). Isso pode estar relacionado à utilização preventiva mais acentuada entre os casados. A preocupação com a saúde é maior entre esses indivíduos uma vez que, de forma geral, existe mais de uma pessoa influenciando no seu cuidado preventivo.

g) Situação atual no plano de saúde: as variáveis desistiram do plano (*desist*), ficaram inadimplentes (*inadimp*), adquiriram a maioridade (*maiorid*), faleceram naquele ano (*falec*), foram transferidos de unidade (*transf*) ou foram demitidas da empresa (*demit*), foram construídas com o objetivo de verificar se a situação atual dos indivíduos poderia explicar a influência sobre os resíduos. Para os indivíduos que faleceram naquele ano, existe um fato empírico de que os indivíduos gastam mais quando estão perto da morte. Esse comportamento dos gastos se reflete de forma diferente dependendo do tipo de gasto relacionado. Caso o cuidado seja de internação, por exemplo, haverá um aumento na utilização desses serviços se o indivíduo faleceu naquele ano. Dessa forma, o modelo quer testar como a situação desses indivíduos diante da empresa afeta o tipo de gasto. Se a pessoas foram transferidas, mudaram de plano ou, atingiram a maioridade, por exemplo, elas podem gerar um problema de risco moral no plano, ou seja, a sobreutilização dos serviços de saúde.

h) Gasto e internação no período anterior: essas variáveis foram construídas para serem *proxies* do estado de saúde passado. A variável *Gst1* se refere ao gasto no período anterior que excedeu a média mais de uma vez e meia o desvio-padrão. A variável *Int1* se refere à internação no período anterior à análise. É esperado que, se os indivíduos tiveram gastos com qualquer serviço de saúde no período anterior ou foram internados anteriormente, maior será a utilização dos serviços de saúde no período corrente. Isso pode estar relacionado com a continuidade de tratamento, dependendo da gravidade da doença. Essas variáveis são indicadores de estado de saúde precário dos indivíduos.

i) *Dummies* de período: na análise em painel também foram geradas *dummies* de período (*d01* à *d03*) para os anos analisados. Nesse caso, o período de referência foi 2001 (*d01*). Essas variáveis procuram captar os efeitos sazonais que podem ter afetado os gastos em saúde num determinado período. Esses efeitos sazonais podem afetar os gastos negativa ou positivamente em um determinado ano: a ocorrência de uma epidemia de dengue pode afetar os gastos negativamente, enquanto uma campanha preventiva de vacinação contra a febre amarela pode afetar os gastos positivamente, por exemplo. As variáveis, entretanto, não conseguem captar qual fator está afetando os gastos.

A Tabela 4D.26 apresenta um resumo das variáveis utilizadas e o sentido esperado dos efeitos sobre os gastos ambulatoriais e de internação.

TABELA 4D.26  
DESCRIÇÃO DAS VARIÁVEIS E SENTIDO DOS EFEITOS MARGINAIS

Variável	Descrição	Gasto <sup>c</sup>	Internação <sup>d</sup>
Sex1	Homem = 1, mulher = 0	-	-
<i>Idh1</i> a <i>Idh10</i>	Interação entre sexo e idade (homens)	+. <sup>a</sup>	+. <sup>a</sup>
<i>Idm1</i> a <i>Idm10</i>	Interação entre sexo e idade (mulher)	+. <sup>a</sup>	+. <sup>a</sup>
<i>So1</i>	Solteiro = 1, caso contrário = 0	-	-
<i>So2</i>	Casado = 1, caso contrário = 0	+	+
<i>Plano</i>	Plano pleno = 1, caso contrário = 0	+	+
<i>Gst1</i>	Gasto ano anterior = $\mu + 1,5\sigma$ = 1, caso contrário = 0	+	+
<i>Int1</i>	Teve internação no ano anterior = 1, caso contrário = 0	+	+
<i>Ren</i> <sup>b</sup>	Renda	.-/+	.-/+
<i>Esc2</i> <sup>b</sup>	Primário completo/1º grau incompleto = 1, caso contrário = 0	-	-
<i>Esc3</i> <sup>b</sup>	1º grau completo/2º grau incompleto = 1, caso contrário = 0	-	-
<i>Esc4</i> <sup>b</sup>	2º grau completo/3º grau incompleto = 1, caso contrário = 0	-	-
<i>Esc5</i> <sup>b</sup>	3º grau completo = 1, caso contrário = 0	-	-
<i>desist</i>	Desistiram do plano = 1, caso contrário = 0	-	.-/+
<i>inadimp</i>	Ficaram inadimplentes = 1, caso contrário = 0	-	.-/+
<i>falec</i>	Faleceram = 1, caso contrário = 0	-	+
<i>transf</i>	Foram transferidos de unidade = 1, caso contrário = 0	-	.-/+
<i>demit</i>	Foram demitidos da empresa = 1, caso contrário = 0	-	.-/+
<i>maiorid</i>	Atingiram a maioridade = 1, caso contrário = 0		
<i>d02</i>	2002 = 1, caso contrário = 0	.-/+	.-/+
<i>d03</i>	2003 = 1, caso contrário = 0	.-/+	.-/+

<sup>a</sup> Efeito marginal conjunto da forma funcional da idade.

<sup>b</sup> Informação disponível apenas para titulares.

<sup>c</sup> Sentido do efeito marginal esperado para qualquer tipo de gasto.

<sup>d</sup> Sentido do efeito marginal esperado para gastos de internação.

#### 4D.4.1 Resultados do modelo em *cross-section*

Esta subseção descreve os resultados do MQE estimado para o ano de 2003. Os resultados são apresentados separadamente para os beneficiários e para os titulares. A Tabela 4D.27 apresenta os resultados dos coeficientes marginais estimados para

TABELA 4D.27  
MODELO DE QUATRO EQUAÇÕES: *CROSS-SECTION* DOS BENEFICIÁRIOS — 2003

	<i>Probit</i> gasto <sup>a</sup>	OLS gasto ambulatorial	<i>Probit</i> internação <sup>a</sup>	OLS gasto internação
<i>sex01</i>	-0,0582	-0,0875	(0,0145)**	0,1968
<i>idh1</i>	0,0995	0,9746	0,2804	0,4396
<i>idh2</i>	0,0906	0,4896	0,0685	0,2287
<i>idh3</i>	0,0484	0,1267	(-0,0019)**	-0,1154
<i>idh5</i>	0,0500	0,4648	0,0648	0,2826
<i>idh6</i>	0,0679	0,6925	0,0580	0,4595
<i>idh7</i>	0,0676	0,8589	0,1012	0,6539
<i>idh8</i>	0,0481	0,8457	0,0929	0,9773
<i>idh9</i>	0,0484	1,0587	0,2006	1,4014
<i>idh10</i>	0,0415	0,8735	0,1350	1,6077
<i>idm1</i>	0,1005	0,9640	0,2146	(0,3619)**
<i>idm2</i>	0,0766	0,3742	0,1130	0,2592
<i>idm4</i>	(0,0006)**	0,2961	0,0623	0,4505
<i>idm5</i>	(0,0074)**	0,7546	0,1460	0,9016
<i>idm6</i>	(0,0034)**	0,9428	0,1539	0,9414
<i>idm7</i>	(-0,0041)**	1,0325	0,1753	1,1030
<i>idm8</i>	-0,0255	1,0532	0,1801	1,2084
<i>idm9</i>	(0,0084)**	1,0446	0,1739	1,3545
<i>idm10</i>	-0,0706	0,8424	0,1740	1,5953
<i>so1</i>	-0,0461	(0,0403)***	(-0,0134)**	-0,1692
<i>so2</i>	(-0,0077)**	(-0,0248)**	-0,0313	(-0,0280)**
<i>plano</i>	0,0983	-0,0893	-0,0727	-0,1056
<i>gst1</i>	0,0346	0,5666	(0,0020)**	0,7816

(continua)

(continuação)

	<i>Probit</i> gasto <sup>a</sup>	OLS gasto ambulatorial	<i>Probit</i> internação <sup>a</sup>	OLS gasto internação
<i>int1</i>	0,1181	0,0404	0,3541	0,0455
<i>desist</i>	-0,2339	-0,5523	-0,1877	-0,3476
<i>inadimp</i>	-0,1715	-0,5912	-0,0840	(0,1010)**
<i>maiorid</i>	-0,2325	-0,4104	-0,0594	-0,1397
<i>falec</i>	-0,1000	-0,3132	0,3395	2,5192
<i>transf</i>	-0,1467	-0,5357	-0,1447	(-0,3183)**
<i>demit</i>	-0,2152	-0,5325	-0,1709	(-0,1725)**
<i>mudanc</i>	0,0750	(0,0157)**	(-0,0161)**	(0,0977)**
<i>out</i>		1,9581		7,9205
<i>const</i>		3,9163		4,4789
<i>R</i> <sup>2</sup>	0,1111	0,1230	0,1085	0,1461
Probabilidade predita	0,8944		0,4707	

<sup>a</sup> Efeito marginal.

\*\* Não Significativo.

\*\*\* Significativo a 10%.

as quatro equações considerando a população de beneficiários. A primeira coluna apresenta os resultados para a probabilidade de o beneficiário realizar qualquer tipo de gasto com saúde; a segunda coluna os resultados para o modelo de mínimos quadrados estimado para os gastos ambulatoriais, a terceira coluna os resultados para a probabilidade de o beneficiário ter gastos com internação condicionado a ter tido algum tipo de gasto; e a quarta coluna apresenta os resultados para o modelo de gastos com internação.

#### 4D.4.1.1 Resultados do modelo em *cross-section* para os beneficiários

Podemos destacar nas regressões realizadas que, de forma geral, a grande maioria das variáveis é significativa. As variáveis consideradas de risco (idade, sexo e estado de saúde passada) são significativas em praticamente todos os modelos. O modelo dos beneficiários possui melhor ajuste aos dados, condizente com o encontrado na literatura [ver Van de Ven (2000)], dada a maior variância das variáveis de risco: o *R*<sup>2</sup> encontrado variou entre 0,11 e 0,14, nas regressões realizadas.

Os resultados dos beneficiários para *cross-section* de 2003 são apresentados na Tabela 4D.27. Para o MQE dos beneficiários os principais resultados que

podemos destacar são: os efeitos da idade, do sexo masculino, do estado de saúde passado e do tipo de plano.

Um primeiro resultado que chama a atenção é a probabilidade predita do beneficiário médio de realizar qualquer tipo de gasto, 0,89, sugerindo que durante o período de um ano todos os indivíduos necessitam receber algum tipo de serviço de saúde. Esse valor é bem inferior quando consideramos a probabilidade de haver gasto com internação condicionado a ter havido algum gasto, 0,47. Apesar de inferior, esse resultado é surpreendente, sendo provavelmente superior para a população da Sabesperv, que possui plano de saúde.

Com relação à idade, a probabilidade de realização de qualquer tipo de gasto é elevada nos anos iniciais da vida, diminui na infância e adolescência para subir em seguida, corroborando os gráficos apresentados anteriormente. Após os 60 anos essa probabilidade volta a diminuir, influenciada pela queda na realização de gastos ambulatoriais. Ser do sexo masculino reduz a probabilidade de realização de gasto: tudo o mais constante, ela é inferior em 5,82%. O estado de saúde passado, aproximado pelas variáveis de realização do gasto e/ou internação no ano anterior, eleva a probabilidade de realização de gastos assim como possuir plano pleno eleva a probabilidade de realização de qualquer tipo de gasto em 9,83%.

No modelo OLS, o efeito da idade sobre o logaritmo dos gastos ambulatoriais reflete o fato estilizado na literatura dos gastos elevados de recém-nascidos, decrescentes na infância e adolescência e crescentes na fase adulta: indivíduos recém-nascidos contribuem o dobro para a elevação dos gastos ambulatoriais em relação ao grupo de idade de 2 a 5 anos. Por outro lado, os indivíduos do grupo de idade de 60 a 70 anos e de 70 a 80 anos contribuem para a elevação dos gastos ambulatoriais em proporções próximas às dos recém-nascidos. Indivíduos do sexo masculino contribuem para a queda dos gastos ambulatoriais em 8,75%. Indivíduos com estado de saúde precário também possuem elasticidade positiva contribuindo para a elevação dos gastos ambulatoriais e os indivíduos listados no plano pleno reduzem os gastos ambulatoriais em cerca de 9%.

No caso da probabilidade de realização de internações, podemos destacar dois fatos principais: primeiramente o sexo não é uma variável significativa na realização de internações, sugerindo que esse fator de risco não é determinante nos eventos exógenos que levam os indivíduos a procurar um serviço mais complexo. Por outro lado, a idade de indivíduos do sexo feminino é um fator determinante na probabilidade de realização de gastos de internação, diferente da probabilidade de realização de qualquer outro tipo de gasto. Na regressão de OLS para a internação, os resultados de idade, estado de saúde precário e plano de

saúde seguem tendências similares ao caso da regressão de gastos ambulatoriais. A diferença principal aqui é o sexo, com os homens contribuindo para o aumento dos gastos ambulatoriais.

#### 4D.4.1.2 Resultados do modelo em *cross-section* para os titulares

A Tabela 4D.28 apresenta os resultados do MQE para os titulares da *cross-section* de 2003. Nesse caso, o ajuste do modelo foi inferior ao dos beneficiários com o  $R^2$  variando entre 0,07 e 0,12.

TABELA 4D.28  
MODELO DE QUATRO EQUAÇÕES: *CROSS-SECTION* DE TITULARES — 2003

	Probit gasto <sup>a</sup>	OLS gasto Ambulatorial	Probit internação <sup>a</sup>	OLS gasto internação
sex1	(-0,0044)**	(-0,1836)**	(-0,1369)**	(-0,0597)**
idh4		(-0,2621)***	-0,1663	(0,0558)**
idh5	(-0,0016)**	-0,3188	(-0,0149)**	-0,7180
idh6	(0,0048)**	-0,1328	-0,0214	-0,5205
idh7	(0,0063)**	(0,0217)**	(0,0051)**	-0,3607
idm4	(-0,0504)**	(0,1110)**	(-0,0139)**	-0,8027
idm5	(0,0253)**	(0,1603)**	(0,0159)**	(-0,3282)**
idm6	(0,0232)**	(0,2843)**	(-0,0009)**	(-0,2397)**
idm7	(0,0191)**	(0,3658)**	(-0,0186)**	(0,2186)**
ren	-0,00000319	0,0000	0,00000147	(-0,0000011)**
so1	-0,0175	(0,0096)**	(-0,0261)**	(-0,0343)**
so1	(0,0006)**	(-0,0089)**	-0,0526	(0,0096)**
Gst1	0,0203	0,5700	(0,0167)**	0,8317
Int1	0,0373	0,0858	0,3774	(0,0409)**
esc2	(0,0002)**	(-0,0339)**	-0,0430	(0,0657)**
esc3	(0,0059)**	(0,0079)**	(-0,0322)**	(-0,0147)**
esc4	(0,0011)**	(0,0685)***	-0,0714	(-0,0528)**
esc5	(-0,0006)**	0,1759	-0,1215	(0,0490)**
desist	-0,3547	(-0,2916)**	-0,3336	(-1,2123)**
inadimp	-0,5629	(-0,2892)**	-0,2859	(-1,1691)**
falec	(-0,0348)**	(-0,4323)**	0,3219	2,2707
transf	-0,1604	-0,6286	-0,2423	(-0,3150)**
demit	-0,1577	-0,5138	-0,1681	(0,1675)**
const		4,6177		5,55583
$R^2$	0,0716	0,0951	0,1215	0,0721
Probabilidade predita	0,9620		0,4811	

<sup>a</sup> Efeito marginal.

\*\* Não-significativo.

\*\*\* Significativa a 10%.

Podemos destacar quatro resultados principais no modelo: a renda, o estado de saúde passado, a transferência e a demissão dos indivíduos a influenciar os gastos. Além desses, os resultados de internações dos indivíduos que falecem são relevantes. Por outro lado, as principais variáveis de risco destacadas na literatura, idade e sexo, não foram determinantes na realização de gastos. O resultado da idade pode ser considerado como esperado, uma vez que os gastos são maiores nos extremos da vida, e esses grupos não são considerados aqui. Por outro lado, o resultado para o sexo, apesar de inesperado uma vez que esse grupo inclui as mulheres em estado fértil, pode ser atribuído à menor representatividade das mulheres nessa população — apenas 15% dos indivíduos — e por estas provavelmente apresentarem um perfil específico, diferente da média da população.

Indivíduos com renda mais elevada contribuem para a queda da probabilidade de realização de qualquer tipo de gasto, enquanto o estado de saúde precário contribui para a elevação de qualquer tipo de gasto entre os titulares, e indivíduos transferidos e demitidos diminuem a probabilidade de realização de gastos.

Os indivíduos com a renda mais elevada e com o estado de saúde precário elevam os gastos ambulatoriais, de acordo com o resultado encontrado na regressão de OLS: ter realizado internação no ano anterior, por exemplo, eleva os gastos ambulatoriais em 8,58%, tudo o mais constante. Por outro lado, indivíduos transferidos e demitidos contribuem para a redução dos gastos ambulatoriais.

Indivíduos com rendas mais elevadas, que realizaram internação no ano anterior e que faleceram tiveram maior probabilidade de realização de gastos de internação: tiveram probabilidade de internação 32,19% superior tudo o mais constante. Novamente indivíduos que foram transferidos ou demitidos tiveram menor probabilidade de realização de gastos: tiveram probabilidade de realização de gastos de internação 16,81% inferior, tudo o mais constante. No caso dos gastos com internação, o resultado da regressão de OLS mostra que os principais determinantes são os indivíduos que faleceram e realizaram gastos elevados no ano anterior, respectivamente.

#### **4D.4.1.3 Resultados do modelo para dados em painel para os beneficiários**

De maneira geral os resultados do modelo em painel ratificam os resultados encontrados na *cross-section*. O efeito da idade, do sexo, do estado de saúde passado e do plano pleno ocorre no mesmo sentido da *cross-section* na probabilidade de realização de qualquer tipo de gasto: tudo mais constante, indivíduos de sexo masculino têm probabilidade de realização de qualquer tipo de gasto 58% inferior, por exemplo. Por outro lado, indivíduos que tiveram internação

no ano anterior têm probabilidade 39% maior de realizar qualquer tipo de gasto (Tabela 4D.29).

TABELA 4D.29  
MODELO DE QUATRO EQUAÇÕES: PAINEL DOS BENEFICIÁRIOS — 2003

	Probit gasto <sup>a</sup>	OLS gasto ambulatorial	Probit internação <sup>a</sup>	OLS gasto internação
<i>sex1</i>	-0,5807	-0,3702	-0,1265	-0,2523
<i>idh1</i>	2,0701	0,9631	0,5930	0,4294
<i>idh2</i>	1,4389	0,4337	0,3067	0,1769
<i>idh3</i>	0,5462	0,1002	(0,0064)**	-0,1182
<i>idh5</i>	0,6162	0,4355	0,1233	0,1987
<i>idh6</i>	0,8136	0,6664	0,1127	0,4279
<i>idh7</i>	0,8518	0,8225	0,2561	0,6501
<i>idh8</i>	0,7596	0,8443	0,3304	0,9632
<i>idh9</i>	0,8272	1,0755	0,6467	1,5300
<i>idh10</i>	0,4475	0,9835	0,6665	1,7359
<i>idm1</i>	1,5876	0,6119	0,2807	(-0,0273)**
<i>idm2</i>	0,8976	0,0594	0,1401	-0,2692
<i>idm3</i>	(0,0028)**	-0,2976	-0,1923	-0,5105
<i>idm5</i>	0,2141	0,4953	0,2230	0,4405
<i>idm6</i>	0,2096	0,6835	0,2148	0,4202
<i>idm7</i>	0,1520	0,7876	0,3066	0,5526
<i>idm8</i>	0,1365	0,8482	0,3836	0,7064
<i>idm9</i>	0,2442	0,8298	0,4031	0,9247
<i>idm10</i>	-0,2749	0,5907	0,5341	1,2025
<i>so1</i>	-0,4107	(-0,0178)**	-0,0666	-0,1996
<i>so2</i>	-0,0848	-0,0336	-0,0698	(-0,0332)**
<i>d02</i>	-0,0364	-0,1514	0,5341	-0,4180
<i>d03</i>	-0,0358	-0,2049	0,5463	-0,4754
<i>plano</i>	0,5598	-0,1418	-0,0836	-0,1218
<i>gst1</i>	0,1591	0,2044	-0,0432	0,1842
<i>int1</i>	0,3954	-0,0340	0,5542	0,1878
<i>desist</i>	-1,3253	-0,6066	-0,4652	(-0,0356)**
<i>inadimp</i>	-1,1131	-0,5144	-0,1258	(0,1270)**
<i>maiorid</i>	-1,2467	-0,3665	-0,2761	(-0,0117)**
<i>falec</i>	-0,9092	(-0,0375)**	1,1367	2,5199
<i>transf</i>	0,7288	-0,5498	-0,4278	(0,0319)**
<i>demit</i>	-1,2826	-0,5189	-0,4999	(-0,0021)**
<i>mudanc</i>	0,8481	(0,0598)**	(0,0371)**	(0,0848)**
<i>_const</i>	1,5963	4,5009	-0,7866	5,4348

<sup>a</sup> Efeito marginal.

\*\* Não-significativo.

No modelo de OLS para gastos ambulatoriais podemos destacar a diferença na elasticidade dos indivíduos que tiveram internação no ano anterior: tinha valor positivo na *cross-section* e negativo no painel. Esse resultado pode estar refletindo a elevação do risco da população com o decorrer dos anos.

Com relação à probabilidade de realizar gastos de internação, podemos destacar a importância do sexo como determinante do resultado. Diferentemente da *cross-section* de 2003, indivíduos do sexo masculino têm probabilidade de realização de gastos cerca de 12% inferior. No caso do modelo OLS para gasto de internação, o atributo que mais colabora com a elevação nos gastos é o falecimento dos indivíduos, resultado semelhante ao da *cross-section* de 2003.

#### 4D.4.1.4 Resultados do modelo para dados em painel para os titulares

No painel de titulares (Tabela 4D.30), uma diferença importante em relação à *cross-section* é o resultado para o estado de saúde passado da população: indivíduos que realizaram gastos no ano anterior não influenciaram a probabilidade de realização de qualquer tipo de gasto no painel. Por outro lado, os indivíduos que tiveram internação no ano anterior influenciaram significativamente a probabilidade de realização de gastos. Para o modelo de OLS de gastos ambulatoriais, os indivíduos que realizaram internação no ano anterior não contribuíram com os gastos ambulatoriais, contrariando o resultado da *cross-section*.

TABELA 4D.30  
MODELO DE QUATRO EQUAÇÕES: PAINEL DOS TITULARES — 2003

	Probit gasto <sup>a</sup>	OLS gasto ambulatorial	Probit internação <sup>a</sup>	OLS gasto internação
sex1	(-0,1018)**	-0,3846	(-0,1144)**	(0,0551)**
idh4	(0,1054)**	-0,5325	(-0,0959)**	-0,5709
idh5	(0,0605)**	-0,4363	-0,1257	-0,7257
idh6	0,1453	-0,2523	-0,1602	-0,5074
idh7	(0,1233)***	-0,1197	(-0,0301)**	-0,3264
idm4	(0,1506)**	-0,3067	(0,0093)**	(-0,3812)**
idm5	0,5263	(-0,0392)**	0,1872	(-0,2224)**
idm6	0,6404	(0,0510)**	(0,0827)**	(-0,1400)**
idm7	0,4946	(0,1010)**	(0,1591)**	(0,1447)**
ren	0,0000	0,0000	0,0000	(0,0000114)**
so1	-0,1165	(-0,0141)**	(-0,0560)***	-0,1188
so2	(0,0510)**	(-0,0260)**	-0,0985	(-0,0388)**
d02	-0,0685	-0,1658	0,5893	-0,3461
d03	(0,0176)**	-0,2102	0,6716	-0,3810
gst1	(0,1938)**	0,2241	(-0,0023)**	0,1411
int1	0,3271	(0,0132)**	0,4266	0,1820

(continua)

(continuação)

	<i>Probit</i> gasto <sup>a</sup>	OLS gasto ambulatorial	<i>Probit</i> internação <sup>a</sup>	OLS gasto internação
<i>esc2</i>	(−0,0331)**	(−0,0198)**	−0,0888	(0,0200)**
<i>esc3</i>	(0,0262)**	(0,0028)**	−0,1263	(−0,0574)**
<i>esc4</i>	(0,0348)**	0,0616	−0,2528	(−0,0773)**
<i>esc5</i>	(−0,0264)**	0,1210	−0,4199	(0,0325)**
<i>desist</i>	−1,4061	−0,4007	−0,6910	(−0,2032)**
<i>inadimp</i>	−2,1067	−0,6361	(−0,3539)**	(−0,0091)**
<i>falec</i>	−0,5002	(−0,0531)**	1,2144	2,8944
<i>transf</i>	−0,8022	−0,6662	−0,5591	(−0,4279)**
<i>demit</i>	−1,0856	−0,4986	−0,5879	0,2568
<i>mudanc</i>	(0,1117)**	−0,4320	(0,1207)**	(0,6373)***
<i>_const</i>	2,1099	5,1637	−0,4744	5,8536

<sup>a</sup> Efeito marginal.

\*\* Não-significativo.

\*\*\* Significativo a 10%.

Os resultados da probabilidade de realizar internação são similares aos da *cross-section*. As variáveis renda, estado de saúde passado, falecimento, transferência e demissão possuem efeito no mesmo sentido do encontrado para o ano de 2003. Similar ao modelo de OLS de internação dos beneficiários, o atributo que mais afeta os gastos é o falecimento dos indivíduos. Uma diferença do modelo da *cross-section* é que os indivíduos demitidos afetam positivamente os gastos de internação.

#### 4D.5 CONSIDERAÇÕES FINAIS

A principal contribuição deste trabalho é procurar entender os determinantes dos gastos e, conseqüentemente, da demanda por serviços de saúde no Brasil através da ótica de uma população segurada. Com esse fim, utilizou os dados do plano de autogestão da Sabesperv. Por formarem um banco de dados administrativo, utilizado para controlar e efetuar os pagamentos da rede provedora, existe grande confiabilidade nos valores dos pagamentos reportados. Por outro lado, a principal limitação do banco é que por não ter sido elaborado buscando a análise científica, algumas informações relevantes não foram diretamente reportadas: a principal delas é o estado de saúde passado dos indivíduos, o que nos levou a construir variáveis *proxies* que tentassem colher alguma informação do *status* de saúde dos segurados. Além disso, não foi possível efetuar a ligação entre os membros da mesma família, o que não permitiu verificar e, eventual-

mente corrigir, a correlação intrafamiliar<sup>62</sup> existente nem utilizar a informação de renda para todos os beneficiários.

Os principais resultados encontrados indicam grande importância dos atributos de sexo e idade na determinação dos gastos ambulatoriais. Além dessas, outras variáveis de grande importância para explicar os gastos foram as *proxies* utilizadas para o estado de saúde passada dos indivíduos. A renda também foi significativa, muito embora o efeito marginal encontrado seja pequeno em relação às outras variáveis. Existe importância também nos indivíduos que se encontram em fase de transição na empresa ou na vida particular: algumas das variáveis da situação atual dos indivíduos foram significativas na determinação dos gastos ambulatoriais. Por outro lado, nem o estado civil nem a escolaridade foram importantes na determinação desse tipo de gasto. A última informação, disponível apenas para os titulares, reflete certa homogeneidade nas informações sobre prevenção uma vez que todos usufruem o mesmo ambiente de trabalho.

No que diz respeito aos gastos hospitalares, os modelos captaram basicamente duas variáveis como determinantes desse tipo de gastos: a idade e os indivíduos que faleceram. Esse resultado segue o encontrado na literatura que indica gastos mais complexos, e consequentemente mais elevados, na população mais idosa e nos momentos finais da vida. É de grande relevância para o agente regulador, responsável pela proposição e implementação de políticas públicas, entender as causas de morte da população segurada. Na *cross-section* de 2003 o sexo também apresentou significância na determinação dos gastos. Esses resultados, entretanto, não são conclusivos.

Uma extensão deste trabalho deve incorporar informações precisas sobre o estado de saúde passada dos indivíduos e mais informações socioeconômicas e demográficas sobre os beneficiários: basicamente, renda e escolaridade devem influir na população como um todo, a qual é mais heterogênea do que quando são considerados apenas os funcionários da empresa e ela possui acesso a informação de maneira mais dispersa e menos direcionada do que os últimos. Por fim, relacionar titulares e dependentes seria um importante exercício para o entendimento da forma como as variáveis constantes ao nível das famílias — o tipo de plano e a renda, e a correlação intrafamiliar — afetariam os gastos e a demanda por serviços de saúde.

---

62. Os erros nas regressões do MQE podem exibir elevada correlação entre os membros da mesma família. Existe correlação na decisão de receber cuidados e nos gastos entre os membros da mesma família. Ao não se considerar essa correlação, pode-se gerar estimadores menos eficientes e subestimar o desvio-padrão para variáveis que são constantes no nível familiar, tais como o plano de saúde e a renda.

## BIBLIOGRAFIA

- AKERLOF, G. A. The market for “lemons”: quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, p. 488-500, Cambridge - Mass, 1970.
- ALVES, S. L. Eficiência nos contratos do mercado brasileiro de saúde suplementar: um teste econométrico para seleção adversa. *I Jornada de Economia da Saúde*, São Leopoldo - RS, 2003.
- ANDRADE, M. V. Ensaios em economia da saúde. *Ensaio 6: A regulamentação dos contratos de seguro saúde*. Escola de Pós-Graduação em Economia, Fundação Getulio Vargas, Rio de Janeiro, 2000 (Tese de Doutorado).
- ANDRADE, M. V., LISBOA, M. B. Velhos dilemas no provimento de bens e serviços de saúde: uma comparação dos casos canadense, inglês e americano. *Nova Economia*, Belo Horizonte, v. 10, n. 2, p. 73-115, dez. 2000.
- \_\_\_\_\_. A economia da saúde no Brasil. In: LISBOA, M. de B., MENEZES-FILHO, N. A. (orgs.). *Microeconomia e sociedade no Brasil*. Rio de Janeiro: Contra Capa, 2001.
- ANDRADE, M. V., MAIA, A. C., NORONHA, K. V. M. S. *Financiamento do setor de saúde suplementar no Brasil: uma investigação empírica a partir dos dados da PNAD/98*. Artigo apresentado ao Fórum de Saúde Suplementar promovido pela Agência de Saúde Suplementar, 2003.
- ARROW, K. Uncertainty and the welfare economics of medical care. *American Economic Review*, Nashville (TEEN), v. 55, n. 5, p. 942-973, Dec. 1963.
- BANCO MUNDIAL. Disponível em:<<http://www.worldbank.org/>>. Dados citados do Brasil acessíveis em: <<http://devdata.worldbank.org/hnpstatsHnpAtaGlance.asp?sCtry=BRA,Brazil>>.
- BERENSTEIN, C. K. *O perfil etário dos custos de internação na saúde pública no Brasil: uma análise para as capitais das regiões metropolitanas do Brasil em 2000*. Belo Horizonte: Cedeplar/UFMG, 2005, 91p.
- BRASIL. *O sistema público de saúde brasileiro*. Seminário Internacional Tendências e Desafios dos Sistemas de Saúde nas Américas, São Paulo: Ministério da Saúde, 2002a.
- \_\_\_\_\_. *Agência Nacional de Saúde Suplementar. Regulação e saúde: estrutura, evolução e perspectivas da assistência médica suplementar*. Rio de Janeiro: 2002b. 264p. (Série C. Projetos, Programas e Relatórios, 76). Acessível em: <[http://www.ans.gov.br/portal/upload/aans/publicacoes/livro\\_regulacao\\_e\\_saude.pdf](http://www.ans.gov.br/portal/upload/aans/publicacoes/livro_regulacao_e_saude.pdf)>. Acesso em 28/06/2004.
- CAMERON, A. C., TRIVEDI, P. K. *Regression analysis of count data*. Cambridge, UK: Cambridge University Press, 1998.
- CAMERON, A. C. *et alii*. A microeconomic model of the demand for health care and health insurance in Australia. *Review of Economic Studies*, v. 55, n. 1, p. 85-106, 1988.
- CHAMBERLAIN, G. Analysis of covariance with qualitative data. *Review of Economic Studies*, v. 47, p. 225-238, 1980.
- CHIAPPORI, P. A., DURAND, F., GEOFFARD, P. Y. Moral hazard and the demand for physician services: first lessons from a french natural experiment. *European Economic Review*, n. 42, p. 499-511, 1998.

- CHIAPPORI, P. A., SALANIÉ, B. Empirical contract theory: the case of insurance data. *European Economic Review*, p. 943-950, 1997.
- CUTLER, D. M., REBER, S. J. Paying for health insurance: the trade-off between competition and adverse selection. *The Quarterly Journal of Economics*, p. 433-465, 1998.
- CUTLER, D. M., ZECKHAUSER, R. J. *Adverse selection in health insurance*. Cambridge-Mass., 1997 (NBER Working Paper, 6.107).
- \_\_\_\_\_. The anatomy of health insurance. In: CULYER, A. J., NEWHOUSE, J. P. (orgs.). *Handbook of Health Economics*. Amsterdam: North-Holland Press, p. 563-643, 2000.
- DUAN, N. et alii. A comparison of alternative models of the demand for medical care. *Journal of Business and Economic Statistics*, v. 1, n. 2, p.115-126, Apr. 1983.
- GREENE, W.H. *Estimation of sample selection models for count data*. Stern School of Business, New York University, 1997.
- \_\_\_\_\_. *Econometric analysis*. 5<sup>th</sup> ed. Upper Saddle River: Prentice hall 2003, 1.026 p.
- JACK, W. Public intervention in health insurance markets: theory and four examples from Latin America. *The World Bank Research Observer*, v. 17, n. 1, p. 67-88, Spring 2002.
- LISBOA, M. B., MOREIRA, H. *Should we be afraid of managed care? A theoretical assessment*. Rio de Janeiro, 2000 (Ensaios Econômicos da EPGE, 365).
- LISBOA, M. B. et alii. *Política governamental e regulação do mercado de medicamentos*. Brasília, 2001. Acessível em: <[www.fazenda.gov.br](http://www.fazenda.gov.br)>.
- MAIA, A. C. *Seleção adversa e risco moral no sistema de saúde brasileiro*. Belo Horizonte: Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, 2004 (Dissertação de Mestrado).
- MANNING, W. G. et alii. Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review*, v. 77, n. 3, p. 251-277, 1987.
- MANSFIELD, E. *Patents and innovation: an empirical study*. Management Science, 1986.
- MESQUITA, M. A. F. de. A regulação da assistência suplementar à saúde: legislação e contexto institucional. *Regulação & Saúde: estrutura, evolução e perspectivas da assistência médica suplementar*. Rio de Janeiro: Ministério da Saúde/Agência Nacional de Saúde Suplementar, 2002.
- NEWDECK, W., PODCZECK, K. Adverse selection and regulation in health insurance markets. *Journal of Health Economics*, Amsterdam (NL), v. 15, p. 387-408, 1996.
- NEWHOUSE, J. P. Symposium on health care reform. *Journal of Economic Perspectives*, Nashville (TEEN), v. 8, n. 3, Summer 1994.
- \_\_\_\_\_. *Free for all? Lessons from the RAND health insurance experiment*. Harvard University Press, 1996a. 504p.
- \_\_\_\_\_. Reimbursing health plans and health providers: efficiency in production versus selection. *Journal of Economic Literature*, v. 33, p. 144-154, 1996b.
- REIS, C. O. O. O estado e os planos de saúde no Brasil. *Revista do Serviço Público*, n. 1, Ano 51, p. 123-146, 2000.

- RIBEIRO, M. M. *Utilização de serviços de saúde no Brasil: uma investigação do padrão etário por sexo e cobertura por plano de saúde*. Cedeplar/UFGM, 2005 (Dissertação de Mestrado).
- ROTHSCHILD, M., STIGLITZ, J. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics*, Cambridge-Mass., p. 629-650, 1976.
- SAPELLI, C., VIAL, B. Self-selection and moral hazard in Chilean health insurance. *Journal of Health Economics*, n. 830, p. 1-18, 2003.
- SAVAGE, E., WRIGHT, D. J. Moral hazard and adverse selection in Australian private hospitals. *Journal of Health Economics*, n. 826, p. 1-29, 2002.
- STANCIOLI, A. E., ZYLBERSTAJN, H. *Incentivos e risco moral nos planos de saúde no Brasil*. São Paulo: Faculdade de Economia, Administração e Contabilidade/USP, 2002 (Dissertação de Mestrado).
- VAN DE VEN, W. P. M. M., ELLIS, R. P. T. Risk adjustment in competitive health plan markets. In: CULYER, A. J., NEWHOUSE, J. P. (eds). *Handbook of Health Economics*. New York: Elsevier, v.1A, p. 755-845, 2000.
- VARIAN, H. R. *Microeconomia: princípios básicos*. 2<sup>a</sup> ed. Rio de Janeiro: Campus, 1994, 710 p.
- WAGSTAFF, A. The demand for health: some new empirical evidence. *Journal of Health Economics*, v. 5, n. 3, 1986.
- WHO – World Health Organization (Organização Mundial de Saúde). Disponível em: <<http://www.who.int/en/>>.
- WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. Cambridge, Mass.: The MIT Press, 2002.

## ANEXO I

### DEFINIÇÃO DAS VARIÁVEIS EXPLICATIVAS

<b>Características socioeconômicas</b>	
Renda	Logaritmo da renda familiar <i>per capita</i> .
Grupos de idade	Nove grupos de idade.
Gênero	<i>Dummy</i> , sendo 0 para mulheres e 1 para homens.
Área de residência	Conjunto de <i>dummys</i> , área de residência metropolitana, urbana e rural, padronizado pela região rural.
Unidades da federação	Conjunto de variáveis <i>dummies</i> para as unidades da federação do Brasil. A referência é o Estado de São Paulo.
Raça	<i>Dummy</i> , sendo 0 para branco e 1 para não-branco.
Anos de estudo	Conjunto de <i>dummies</i> para os seguintes níveis de escolaridade: 0 a 3 anos de estudo, 4 a 7, 8 a 10, 11 a 14 e 15 ou mais. A categoria de referência usada é de 15 anos ou mais de estudo.
Número de filhos	Conjunto de quatro variáveis discretas referente ao número de filhos na família com as seguintes faixas de idade: 0 e 1 ano, 2 a 7, 8 a 15, 16 a 25.
Posição na ocupação	Conjunto de oito variáveis <i>dummies</i> , empregado com carteira, empregados sem carteira, trabalhador doméstico, conta-própria, empregador, trabalhador não-remunerado ou na produção para o próprio uso/consumo, não-PEA e desocupados. Padronizado pelos empregados com carteira.
Número de componentes na família	Variável discreta.
Qualidade	Variável categórica igual a 0 se o indivíduo não utilizou nenhum serviço de saúde, igual a 1 se utilizou e considerou esse atendimento ruim, regular ou muito ruim, e igual a 2, se considerou esse atendimento muito bom ou bom (Questão V1367 da PNAD: Considera que esse atendimento de saúde foi: muito bom, bom, regular, ruim ou muito ruim).
Acesso	Variável <i>dummies</i> , igual a 1 se o indivíduo teve problema de acesso e 0, do contrário. A construção dessa variável consiste em uma combinação de três questões referentes à utilização dos serviços de saúde. Variável V1350: nas duas últimas semanas, procurou algum lugar, serviço ou profissional de saúde para atendimento relacionado à própria saúde? Caso a resposta seja afirmativa: variável V1354: nessa primeira vez que procurou atendimento de saúde nas duas últimas semanas foi atendido? Sim (não houve problema de acesso). Não (houve problema de acesso). Caso a resposta a V1350 seja negativa, ir para variável V1368: nas duas últimas semanas, por que motivo não procurou atendimento de saúde? Não houve necessidade (não houve problema de acesso). Categorias de resposta de 02 a 10, por exemplo: não tinha dinheiro, horário incompatível (houve problema de acesso).
Probabilidade de ter plano	Valor predito da variável dependente do modelo <i>probit</i> .
<b>Características do estado de saúde</b>	
Saúde auto-avaliada	<i>Dummy</i> igual a 1 se o indivíduo avalia o seu estado de saúde como bom ou muito bom e 0 se avalia como regular, ruim e muito ruim.
Presença de doença crônica	A PNAD questiona sobre a presença de 12 tipos diferentes de doenças crônicas: coluna, artrite ou reumatismo, câncer, diabetes, bronquite ou asma, hipertensão, doença do coração, doença renal crônica, depressão, tuberculose, tendinite e cirrose. Foi incluída uma variável <i>dummy</i> para cada doença. Caso o indivíduo responda que possui uma doença, a categoria de resposta é igual a 1, do contrário, 0.

TABELA I.1  
**CONSULTAS MÉDICAS MODELO DE UTILIZAÇÃO BINOMIAL NEGATIVO**

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Log renda	0.0146*	-0.0190
De 0 a 2 anos	0.5288	0.5449
De 3 a 5	0.2924	0.2799
De 11 a 25	-0.1857	-0.1453
De 26 a 40	-0.0831	-0.0700
De 41 a 50	-0.0837	-0.0548
De 51 a 60	-0.0875	-0.0666
De 60 a 70	-0.1161	-0.0796
71 ou mais	-0.0707	-0.0997
Sexo	-0.3182	-0.4163
Metropolitana	0.2596	0.3388
Urbana	0.0410*	0.1396
Rondônia	-0.2131	-0.3612
Acre	-0.0334*	-0.3519
Amazonas	-0.1370	-0.3704
Roraima	0.0415*	-0.0511*
Pará	-0.1737	-0.2686
Amapá	-0.1570*	-0.2672
Tocantins	-0.3236	-0.2209
Maranhão	-0.0796*	-0.4898
Piauí	-0.0313*	-0.2620
Ceará	-0.0280*	-0.1562
Rio Grande do Norte	0.1683	0.0013*
Paraíba	0.0668	-0.1369
Pernambuco	0.0541	-0.0500

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Alagoas	0.0694*	-0.2105
Sergipe	0.1647	-0.1764
Bahia	0.0415*	-0.1447
Minas Gerais	-0.0493	-0.0924
Espírito Santo	0.2389	0.0919
Rio de Janeiro	-0.0285*	-0.1032
Paraná	-0.0054*	-0.0643
Santa Catarina	-0.1203	-0.2652
Rio Grande do Sul	0.0456	-0.0543
Mato Grosso do Sul	0.0036*	-0.0279*
Mato Grosso	-0.1456	-0.3418
Goiás	-0.1369	-0.1542
Distrito Federal	0.0157*	-0.1234
Até 3 anos de estudo	-0.1341	-0.0788
De 4 a 7	-0.0929	0.0498*
De 8 a 10	-0.0533	0.0850
De 11 a 14	-0.0235*	0.0899
Filhos de 0 e 1 ano	0.2509	0.2217
Filhos de 2 a 7	-0.0404	-0.0285
Filhos de 8 a 15	-0.0893	-0.0796
Filhos de 16 a 25	-0.0582	-0.0730
Saúde auto-avaliada	-0.4586	-0.6756
Coluna	0.1683	0.1995
Artrite	0.0298*	0.0755
Câncer	0.5892	0.8152

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Diabete	0.2887	0.3860
Asma	0.3429	0.4445
Hipertensão	0.2186	0.3356
Coração	0.2006	0.2792
Renal	0.1974	0.1147
Depressão	0.2768	0.2540
Tuberculose	0.3763	0.5432
Tendinite	0.2034	0.1630
Cirrose	0.2968	0.2532
Acesso	0.1907	0.1612
Qualidade	0.3596	0.5206
Probabilidade estimada de possuir plano	0.1080	0.1668
Constante	1.1412	0.8851

\* Não-significativo.

TABELA I.2  
DIAS DE INTERNAÇÃO: MODELO DE UTILIZAÇÃO BINOMIAL NEGATIVO

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Log renda	-0.1025	-0.0898
De 0 a 2 anos	0.3233	0.7078
De 3 a 5	-0.0055*	0.3983
De 11 a 25	-0.0777*	0.3380
De 26 a 40	0.2109*	0.5016
De 41 a 50	0.3006	0.5064
De 51 a 60	0.2036*	0.3717
De 60 a 70	0.5214	0.7422
71 ou mais	0.9517	0.7676
Sexo	-0.0421*	-0.0236*
Metropolitana	0.0715*	0.1159*
Urbana	-0.0163*	0.1420
Rondônia	0.3449*	0.2116*
Acre	0.2996*	0.2449*
Amazonas	-0.4974	-0.4309
Roraima	-0.5396*	0.7186*
Pará	-0.0072*	-0.1207*
Amapá	0.3281*	0.2075*
Tocantins	-0.3578	0.0544*
Maranhão	-0.6408	-0.1281*
Piauí	-0.0245*	0.0141*
Ceará	-0.5784	0.2489
Rio Grande do Norte	-0.2832*	0.2541*
Paraíba	0.2328*	0.0269*
Pernambuco	-0.3342	0.2573

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Alagoas	–0.5549	0.0839*
Sergipe	–0.2327*	0.0613*
Bahia	–0.3248	–0.0347*
Minas Gerais	–0.2441	0.2552
Espírito Santo	0.0728*	0.0066*
Rio de Janeiro	–0.3199	0.0344*
Paraná	–0.2137*	0.2775
Santa Catarina	0.1169*	0.1540*
Rio Grande do Sul	0.0748*	0.3538
Mato Grosso do Sul	0.2475*	0.1925*
Mato Grosso	0.2249*	0.0165*
Goiás	0.0239*	0.2378
Distrito Federal	–0.1840*	–0.0330*
Até 3 anos de estudo	–0.3820	0.2124*
De 4 a 7	–0.2605*	0.2659*
De 8 a 10	–0.0528*	0.1295*
De 11 a 14	–0.2510	0.1962*
Filhos de 0 e 1 ano	0.9596	0.7952
Filhos de 2 a 7	–0.0947	–0.0650
Filhos de 8 a 15	–0.0656*	–0.1569
Filhos de 16 a 25	–0.0765	–0.1019
Saúde auto-avaliada	–0.9480	–1.2922
Coluna	0.0878*	–0.1451
Artrite	–0.1831	–0.1081
Câncer	1.4875	1.3690

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Diabete	0.3600	0.4859
Asma	0.5966	0.5146
Hipertensão	0.2390	0.0935
Coração	0.7437	0.5603
Renal	0.8492	0.5752
Depressão	0.4430	0.6941
Tuberculose	1.0268	1.2441
Tendinite	0.0194*	0.1858*
Cirrose	1.2753	1.3328
Acesso	0.4388	0.3734
Qualidade	0.4482	0.5545
Probabilidade estimada de possuir plano	-0.3714*	-0.4584
Constante	0.1067*	-0.9219

\* Não-significativo.

TABELA I.3  
NÚMERO DE INTERNAÇÕES: MODELO DE UTILIZAÇÃO BINOMIAL NEGATIVO

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Log renda	-0.0371	-0.0763
De 0 a 2 anos	0.0485*	0.4064
De 3 a 5	0.1071*	0.2962
De 11 a 25	0.2501	0.5436
De 26 a 40	0.4160	0.5846
De 41 a 50	0.4314	0.4560
De 51 a 60	0.4601	0.4409
De 60 a 70	0.5745	0.5924
71 ou mais	0.8781	0.7912
Sexo	-0.2803	-0.3346
Metropolitana	-0.3410	-0.2292
Urbana	-0.1106*	0.1122
Rondônia	0.2944*	0.1691*
Acre	0.1607*	-0.2061*
Amazonas	-0.3506	-0.4121
Roraima	0.1651*	-0.2253*
Pará	0.1871	0.0286*
Amapá	-0.5405*	-0.3302
Tocantins	0.1464*	0.4387
Maranhão	-0.0346*	0.0418*
Piauí	-0.0001*	0.0203*
Ceará	0.0203*	0.1256
Rio Grande do Norte	-0.0135*	0.1237
Paraíba	0.2961	0.0307*
Pernambuco	0.0183*	0.0529*

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Alagoas	-0.4983	-0.1948
Sergipe	0.0468*	-0.0711*
Bahia	0.0256*	-0.0286*
Minas Gerais	0.0053*	0.1013
Espírito Santo	0.1207*	0.0736*
Rio de Janeiro	-0.2797	-0.1889
Paraná	0.1913	0.1965
Santa Catarina	0.0147*	0.1057*
Rio Grande do Sul	0.0383*	0.1592
Mato Grosso do Sul	0.2000	0.1062*
Mato Grosso	0.3522	0.0993*
Goiás	0.2035	0.3119
Distrito Federal	0.0453*	0.1127*
Até 3 anos de estudo	-0.0761*	-0.1409*
De 4 a 7	-0.0462*	-0.0504*
De 8 a 10	0.0162*	-0.0734*
De 11 a 14	0.0338*	-0.0716*
Filhos de 0 e 1 ano	1.0265	0.8976
Filhos de 2 a 7	-0.0049*	-0.0385
Filhos de 8 a 15	-0.0794	-0.1151
Filhos de 16 a 25	-0.0593	-0.1030
Saúde auto-avaliada	-0.8211	-0.9882
Coluna	-0.0412*	-0.0169*
Artrite	-0.0892*	-0.0514*
Câncer	1.2434	1.3178

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Amostra com plano	Amostra sem plano
Diabete	0.2693	0.3412
Asma	0.5782	0.7319
Hipertensão	0.1501	0.2400
Coração	0.5093	0.5762
Renal	0.5044	0.4230
Depressão	0.2659	0.3116
Tuberculose	0.4553*	0.7191
Tendinite	-0.0185*	0.0038*
Cirrose	1.2509	1.0308
Acesso	0.2909	0.1610
Qualidade	0.3743	0.4811
Probabilidade estimada de possuir plano	-0.1085*	-0.4739
Constante	-1.9285	-2.1484

\* Não-significativo.

TABELA I.4

**RESULTADOS DOS COEFICIENTES ESTIMADOS PELO MODELO *PROBIT* PARA PROBABILIDADE DE TER PLANO DE SAÚDE**

Variáveis	Coefficientes estimados
Rondônia	-0.2488
Acre	-0.7472
Amazonas	-0.2567
Roraima	-1.2390
Pará	-0.0508*
Amapá	-0.3250
Tocantins	-0.2528
Maranhão	-0.4743
Piauí	-0.2609
Ceará	-0.2986
Rio Grande do Norte	-0.4314
Paraíba	-0.1019
Pernambuco	-0.1088
Alagoas	-0.2942
Sergipe	-0.1405
Bahia	-0.1585
Minas Gerais	-0.0357
Espírito Santo	-0.0279*
Rio de Janeiro	-0.2441
Paraná	-0.3192
Santa Catarina	-0.2961
Rio Grande do Sul	-0.1618
Mato Grosso do Sul	-0.0280*
Mato Grosso	-0.5040
Goiás	-0.0468
Distrito Federal	-0.4804

(continua)

(continuação)

Variáveis	Coeficientes estimados
Sexo	-0.1042
Idade	-0.0038
Idade <sup>2</sup>	0.0001
Raça	0.1109
Metropolitana	0.4702
Urbana	0.3056
Log renda familiar <i>per capita</i>	0.4625
Até 3 anos de estudo	-1.1730
De 4 a 7	-0.8728
De 8 a 10	-0.5897
De 11 a 14	-0.3029
Filhos de 0 e 1 ano	0.0844
Filhos de 2 a 7	0.0322
Filhos de 8 a 15	0.0618
Filhos de 16 a 25	0.0268
Saúde auto-avaliada	0.0035*
Número de componentes na família	-0.0204
Empregado sem carteira	-0.7196
Trabalhador doméstico	-0.7514
Conta-própria	-0.6952
Empregador	-0.5348
Não-remunerado	-0.2746
Não-PEA	-0.2174
Desocupados	-0.1784
Constante	-2.1646

\* Não-significativo.

TABELA I.5

**MODELO DE UTILIZAÇÃO DE CONSULTAS MÉDICAS, DIAS DE INTERNAÇÃO E NÚMERO DE INTERNAÇÕES PARA PLANOS NÃO-CORPORATIVOS (ADESÃO INDIVIDUAL)**

Variáveis	Coeficientes estimados		
	Consultas médicas	Dias de internação	Número de internações
Log renda	0.0030*	-0.0207*	0.0050*
De 0 a 2 anos	0.5508	0.5641	0.1567*
De 3 a 5	0.3144	0.5728	0.2824
De 11 a 25	-0.1964	0.2959	0.4502
De 26 a 40	-0.0870	0.6421	0.5761
De 41 a 50	-0.0988	0.7712	0.6057
De 51 a 60	-0.0753	0.8228	0.6955
De 60 a 70	-0.0737	0.9555	0.8470
71 ou mais	-0.0821	1.4508	1.1357
Sexo	-0.3285	-0.0999*	-0.2611
Metropolitana	0.2168	0.0255*	-0.3594
Urbana	0.0319*	-0.0073*	-0.1795
Rondônia	-0.0482*	0.2149*	0.5396
Acre	0.1713*	0.7758*	0.4431
Amazonas	-0.0501*	-0.6870	-0.4342
Roraima	0.1018*	-0.2946*	0.2908*
Pará	-0.0554*	0.1051*	0.2345
Amapá	0.1058*	1.3518*	-0.5768*
Tocantins	-0.2019	-0.8079	0.0934*
Maranhão	0.0032*	-0.4742*	-0.2145*
Piauí	-0.1635*	-0.8970	-0.2924*
Ceará	0.0674*	-0.3375	0.0021*
Rio Grande do Norte	0.1399	-0.0782*	0.0564*
Paraíba	0.1427	0.3100*	0.2856

(continua)

(continuação)

Variáveis	Coeficientes estimados		
	Consultas médicas	Dias de internação	Número de internações
Pernambuco	0.1298	-0.2375*	-0.2005
Alagoas	0.2679	-0.2713*	-0.1070*
Sergipe	0.2445	-0.4806	0.1940*
Bahia	0.0831	0.0269*	0.1552*
Minas Gerais	-0.0393*	-0.1553*	0.0000*
Espírito Santo	0.2109	0.3253*	0.0627*
Rio de Janeiro	0.0256*	-0.2001*	-0.3142
Paraná	0.0241*	-0.0091*	0.1567
Santa Catarina	-0.0974	0.2078*	0.0184*
Rio Grande do Sul	0.0864	-0.0569*	-0.0646*
Mato Grosso do Sul	0.0522*	0.1316*	0.2952
Mato Grosso	-0.1462	0.4793	0.5408
Goiás	-0.0411*	-0.2923	0.0971*
Distrito Federal	0.0837*	-0.3811	-0.0474*
Até 3 anos de estudo	-0.1061	-0.0874*	-0.0100*
De 4 a 7	-0.0769	-0.0217*	0.0080*
De 8 a 10	-0.0519*	-0.0857*	0.0061*
De 11 a 14	-0.0265*	-0.1897	0.0359*
Filhos de 0 e 1 ano	0.2498	1.1732	1.1083
Filhos de 2 a 7	-0.0458	-0.0572*	0.0253*
Filhos de 8 a 15	-0.0785	-0.0675*	-0.0643
Filhos de 16 a 25	-0.0481	-0.0728*	-0.0641
Saúde auto-avaliada	-0.4471	-0.9963	-0.8174
Coluna	0.1647	0.0333*	-0.0625*
Artrite	0.0227*	-0.2133	-0.0385*

(continua)

(continuação)

Variáveis	Coeficientes estimados		
	Consultas médicas	Dias de internação	Número de internações
Câncer	0.6337	1.4949	1.3032
Diabete	0.2533	0.5680	0.2915
Asma	0.3530	0.5523	0.5878
Hipertensão	0.2089	0.0362*	0.0200*
Coração	0.2261	0.8149	0.5806
Renal	0.2331	1.0781	0.4938
Depressão	0.3315	0.4185	0.2469
Tuberculose	0.4334*	-2.0012	-0.7395*
Tendinite	0.1887	-0.0020*	0.0553*
Cirrose	0.2921	0.8735	1.0503
Acesso	0.1627	0.3573*	0.3235
Qualidade	0.3512	0.5334	0.3941
Probabilidade estimada de possuir plano de adesão individual	0.1234*	-0.2680*	-0.0832*
Constante	1.2216	-1.0871	-2.3815

\* Não-significativo.

**TABELA I.6  
MODELO DE UTILIZAÇÃO DE CONSULTAS MÉDICAS, DIAS DE INTERNAÇÃO E NÚMERO DE  
INTERNAÇÕES PARA PLANOS CORPORATIVOS (ADESÃO VIA TRABALHO)**

Variáveis	Coeficientes estimados		
	Consultas médicas	Dias de internação	Número de internações
Log renda	0.0426	-0.2389	-0.1224
De 0 a 2 anos	0.4716	0.0655*	-0.1664*
De 3 a 5	0.2586	-0.4621	-0.1237*
De 11 a 25	-0.1945	-0.4214	0.0046*
De 26 a 40	-0.0963	-0.1135*	0.2354
De 41 a 50	-0.0715	-0.0370*	0.2441
De 51 a 60	-0.1501	-0.4337	0.1321*
De 60 a 70	-0.1645	0.3457*	0.3165
71 ou mais	-0.0653*	0.5290	0.7223
Sexo	-0.3168	-0.0365*	-0.2858
Metropolitana	0.3271	-0.0026*	-0.4563
Urbana	0.0502*	-0.0963*	-0.1328*
Rondônia	-0.2343	0.2693*	0.2687*
Acre	-0.2463	-0.5037*	-0.0619*
Amazonas	-0.1630	-0.6290	-0.2445*
Roraima	0.0539*	-1.1840*	-0.4509*
Pará	-0.2175	-0.4304*	0.1513*
Amapá	-0.1758*	-0.4955*	-0.3842*
Tocantins	-0.2619	-0.4851*	0.2388*
Maranhão	-0.1255*	-1.0732	0.1109*
Piauí	0.0997*	0.0122*	0.1422*
Ceará	-0.0820*	-0.8382	0.0611*
Rio Grande do Norte	0.3814	-0.8146	0.0087*
Paraíba	0.0722*	-0.0715*	0.3121

(continua)

(continuação)

Variáveis	Coeficientes estimados		
	Consultas médicas	Dias de internação	Número de internações
Pernambuco	-0.0150*	-0.4321*	0.2900
Alagoas	-0.1027*	-0.7237*	-0.8984
Sergipe	0.1668	-0.3392*	0.0398*
Bahia	0.0449*	-0.8578	-0.1157*
Minas Gerais	-0.0402*	-0.4580	0.0665*
Espírito Santo	0.3774	-0.4577*	0.2815*
Rio de Janeiro	-0.0829	-0.4283*	-0.1001*
Paraná	-0.0281*	-0.4166*	0.3769
Santa Catarina	-0.0776*	0.0213*	0.0996*
Rio Grande do Sul	0.0540*	0.0007*	0.2681
Mato Grosso do Sul	0.0064*	0.3281*	0.2254*
Mato Grosso	-0.0605*	-0.0160*	0.0890*
Goiás	-0.1139	0.0167*	0.3646
Distrito Federal	0.0065*	-0.1996*	0.2370
Até 3 anos de estudo	-0.1954	-0.4845	-0.1107*
De 4 a 7	-0.1336	-0.2517*	-0.0650*
De 8 a 10	-0.0658	-0.0697*	0.0490*
De 11 a 14	-0.0034*	-0.2214*	0.0610*
Filhos de 0 e 1 ano	0.2647	0.8372	0.9080
Filhos de 2 a 7	-0.0193*	-0.1253	-0.0416*
Filhos de 8 a 15	-0.0916	-0.0906	-0.0945
Filhos de 16 a 25	-0.0615	-0.0835*	-0.0584
Saúde auto-avaliada	-0.4764	-1.0635	-0.8427
Coluna	0.1897	0.0833*	-0.0335*
Artrite	0.0423*	-0.2851	-0.1215*

(continua)

(continuação)

Variáveis	Coeficientes estimados		
	Consultas médicas	Dias de internação	Número de internações
Câncer	0.5169	1.6978	1.1916
Diabete	0.3516	-0.1005*	0.1564*
Asma	0.3214	0.7813	0.6249
Hipertensão	0.2322	0.5640	0.3020
Coração	0.1747	0.7255	0.4235
Renal	0.1975	0.3567	0.5529
Depressão	0.1900	0.5510	0.3042
Tuberculose	0.0937*	1.4941	1.0885
Tendinite	0.2642	0.0689*	-0.0816*
Cirrose	0.2715	1.3668	1.5007
Acesso	0.3799	0.3398	0.3394
Qualidade	0.2195	0.4193	0.2297
Constante	0.9570	1.4603	-1.2949

\* Não-significativo.

## ANEXO II

QUADRO II.1  
DEFINIÇÃO DAS VARIÁVEIS EXPLICATIVAS

<b>Características socioeconômicas</b>	
Renda	Logaritmo da renda familiar <i>per capita</i> .
Idade e idade <sup>2</sup>	
Gênero	<i>Dummy</i> , sendo 0 para mulheres e 1 para homens
Área de residência	Conjunto de <i>dummy</i> , área de residência metropolitana, urbana e rural, padronizado pela região rural.
Região	Conjunto de seis variáveis <i>dummies</i> , região Norte, Nordeste, Sudeste (exceto São Paulo), Estado de São Paulo, regiões Sul e Centro-Oeste. A referência adotada é o Estado de São Paulo.
Raça	<i>Dummy</i> , sendo 0 para branco e 1 para não-branco
Anos de estudo	Conjunto de <i>dummies</i> para os seguintes níveis de escolaridade: 0 a 3 anos de estudo; 4 a 7; 8 a 10; 11 a 14; e 15 anos ou mais. A categoria de referência usada é de 15 anos ou mais de estudo.
Número de filhos	Conjunto de quatro variáveis discretas referente ao número de filhos na família com as seguintes faixas de idade: 0 e 1 ano; 2 a 7; 8 a 15; e 16 a 25.
Posição na ocupação	Conjunto de oito variáveis <i>dummies</i> , empregado com carteira, empregados sem carteira, trabalhador doméstico, conta-própria, empregador, trabalhador não-remunerado ou na produção para o próprio uso/consumo, não-PEA e desocupados. Padronizado pelos empregados com carteira.
Número de componentes na família	Variável discreta.
Tipo de família	Quatro variáveis <i>dummies</i> : casal, casal com filhos, mãe com filhos e outros tipos de família.
Acesso	Variável <i>dummy</i> igual a 1 se o indivíduo teve problema de acesso e 0, do contrário. A construção dessa variável consiste em uma combinação de três questões referentes à utilização dos serviços de saúde. Variável V1350: nas duas últimas semanas, procurou algum lugar, serviço ou profissional de saúde para atendimento relacionado à própria saúde? Caso a resposta seja afirmativa: variável V1354: nessa primeira vez que procurou atendimento de saúde nas duas últimas semanas foi atendido? Sim (não houve problema de acesso). Não (houve problema de acesso). Caso a resposta a V1350 seja negativa, ir para variável V1368: nas duas últimas semanas, por que motivo não procurou atendimento de saúde? Não houve necessidade (não houve problema de acesso). Categorias de resposta de 02 a 10, por ex: não tinha dinheiro, horário incompatível (houve problema de acesso).
Probabilidade de ter plano	Valor predito da variável dependente do modelo <i>probit</i> .
<b>Características do estado de saúde</b>	
Saúde auto-avaliada	<i>Dummy</i> igual a 1 se o indivíduo avalia o seu estado de saúde como bom ou muito bom e 0 se avalia como regular, ruim e muito ruim.
Presença de doença crônica	A PNAD questiona sobre a presença de 12 tipos diferentes de doenças crônicas: coluna, artrite ou reumatismo, câncer, diabetes, bronquite ou asma, hipertensão, doença do coração, doença renal crônica, depressão, tuberculose, tendinite e cirrose. Foi incluída uma variável <i>dummy</i> para cada doença. Caso o indivíduo responda que possui a doença, a categoria de resposta é igual a 1, do contrário, 0.

## QUADRO II.2

**VARIÁVEIS DE PLANO DE SAÚDE E VARIÁVEIS DE UTILIZAÇÃO****Variáveis de plano de saúde**

Plano de saúde	1 se o indivíduo possui plano de saúde, 0 se não.
Plano privado via empregador	1 se o indivíduo possui plano de saúde privado via empregador, 0 se possui outro tipo de plano de saúde.
Plano por intermédio do vínculo trabalhista com o setor público	1 se o indivíduo possui plano de saúde por intermédio do vínculo trabalhista e esse plano é gerido por uma instituição pública, 0 se possui outro tipo de plano de saúde.
Plano voluntário	1 se o indivíduo possui plano voluntário, 0 se possui outro tipo de plano de saúde.

**Variáveis de utilização do cuidado de saúde**

Número de consultas	Número de consultas realizadas pelo indivíduo nos últimos 12 meses.
Número de dias internado	Número de dias internado na última internação dos últimos 12 meses.

TABELA II.1  
MODELO DE SELEÇÃO PARA DADOS DE CONTAGEM: PLANO DE SAÚDE

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
<b>Equação de utilização</b>		
Constante	1,8026***	0,0391
Norte	-0,2163***	-0,0226
Nordeste	-0,0267***	0,0590***
Sudeste (exceto São Paulo)	-0,0174*	0,1572***
Sul	-0,0041	0,1355***
Centro-Oeste	-0,0542***	0,1470***
Sexo	-0,4698***	-0,1978***
Idade	-0,0118***	-0,0026
idade <sup>2</sup>	0,849199D-04***	0,331742D-04 <sup>+</sup>
Raça	0,0113*	0,1095***
Metropolitana	0,2790***	-0,0185
Urbana	0,0201	-0,1176***
Renda familiar <i>per capita</i>	0,0136***	-0,0818***
Coluna	0,1565***	0,0996***
Artrite	0,0546***	-0,0983***
Câncer	0,7147***	1,0000***
Diabetes	0,2954***	0,4551***
Asma	0,2317***	0,3680***
Hipertensão	0,1790***	0,0954***
Coração	0,2251***	0,6070***
Renal	0,1368***	0,4858***
Depressão	0,2762***	0,2418***
Tendinite	0,1769***	0,1865***
Cirrose	0,2867***	1,1757***

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Tuberculose	0,4069***	0,2962*
Até 3 anos de estudo	-0,2280***	-0,1443***
De 4 a 7	-0,1502***	-0,0148
De 8 a 10	-0,1070***	-0,0618***
De 11 a 14	-0,0651***	-0,1527***
Filhos de 0 e 1 anos	0,0653***	0,2605***
Filhos de 2 a 7	-0,0783***	-0,2744***
Filhos de 8 a 15	-0,0181***	0,0599***
Filhos de 16 a 25	0,0257***	0,0187**
Saúde auto-avaliada	-0,5649***	-0,9037***
Casal com filhos	-0,0168	0,0064
Mãe com filhos	-0,1025***	-0,2255***
Outro tipo de família	-0,1449***	-0,3054***
Acesso	-0,1330***	0,1482***
Equação de seleção		
Constante	-1,2351***	-1,2351***
Norte	-0,2752***	-0,2752***
Nordeste	-0,2962***	-0,2962***
Sudeste (exceto São Paulo)	-0,2086***	-0,2086***
Sul	-0,2490***	-0,2490***
Centro-Oeste	-0,2259***	-0,2259***
Sexo	-0,1297***	-0,1297***
Idade	0,0205***	0,0205***
Idade <sup>2</sup>	-0,0001***	-0,0001***
Raça	0,1378***	0,1378***
Metropolitana	0,5608***	0,5608***

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Urbana	0,3877***	0,3877***
Log renda familiar <i>per capita</i>	0,1997***	0,1997***
Até 3 anos de estudo	-1,6304***	-1,6304***
De 4 a 7	-1,1882***	-1,1882***
De 8 a 10	-0,8378***	-0,8378***
De 11 a 14	-0,4702***	-0,4702***
Filhos de 0 e 1 ano	0,0451**	0,0451**
Filhos de 2 a 7	-0,0679***	-0,0679***
Filhos de 8 a 15	-0,0516***	-0,0516***
Filhos de 16 a 25	-0,0382***	-0,0382***
Saúde auto-avaliada	0,0346***	0,0346***
Número de componentes na família	0,0448***	0,0448***
Empregado sem carteira	-0,8426***	-0,8426***
Trabalhador doméstico	-0,9618***	-0,9618***
Conta-própria	-0,8264***	-0,8264***
Empregador	-0,4465***	-0,4465***
Não-remunerado	-0,4363***	-0,4363***
Não-PEA	-0,3521***	-0,3521***
Desocupados	-0,5166***	-0,5166***
Sigma	0,4192***	0,6927***
Rho	-0,344776D-08	-0,183432D-07

\* Significativo a 10%.

\*\* Significativo a 5%.

\*\*\* Significativo a 1%.

TABELA II.2  
MODELO DE SELEÇÃO PARA DADOS DE CONTAGEM: PLANO PÚBLICO

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
<b>Equação de utilização</b>		
Constante	1,4469***	1,9268***
Norte	-0,0309	-0,7342***
Nordeste	0,0911	-0,7653***
Sudeste (exceto São Paulo)	0,1064*	-0,3976***
Sul	0,1190**	-0,0813
Centro-Oeste	0,1064	-0,3367***
Sexo	-0,4611***	-0,2473***
Idade	-0,0233***	-0,0281***
Idade <sup>2</sup>	0,0002***	-0,128421D-04
Raça	0,0213	0,2228***
Metropolitana	0,1513*	-0,3825***
Urbana	-0,1090	-0,5654***
Renda familiar <i>per capita</i>	0,0266**	-0,0719***
Coluna	0,1882***	0,2018***
Artrite	0,0798**	-0,2852***
Câncer	0,8612***	0,6848***
Diabetes	0,4785***	0,3896***
Asma	0,2168***	0,5522***
Hipertensão	0,2820***	0,4130***
Coração	0,3335***	0,5308***
Renal	0,1962***	0,2478***
Depressão	0,2655***	0,4286***
Tendinite	0,3328***	0,5613***
Cirrose	0,2698	0,7682***

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Tuberculose	-0,7797	-0,1099
Até 3 anos de estudo	-0,3905***	-0,0492
De 4 a 7	-0,2861***	-0,0628
De 8 a 10	-0,2824***	0,0343
De 11 a 14	-0,1244***	-0,0968
Filhos de 0 e 1 ano	0,0231	0,1083
Filhos de 2 a 7	-0,0227	-0,0022
Filhos de 8 a 15	-0,0404**	0,0495
Filhos de 16 a 25	-0,0027	0,1030***
Saúde auto-avaliada	-0,5248***	-1,1050***
Casal com filhos	-0,0253	-0,5039***
Mãe com filhos	-0,0681	-0,7252***
Outro tipo de família	-0,1431**	-0,8995***
Acesso	-0,1494*	0,1753*
Equação de seleção		
Constante	-1,0533***	-1,0532***
Norte	1,1054***	1,1054***
Nordeste	0,7544***	0,7543***
Sudeste (exceto São Paulo)	0,4511***	0,4506***
Sul	0,6061***	0,6059***
Centro-Oeste	1,0433***	1,0430***
Sexo	-0,1031***	-0,1029***
Idade	0,0142***	0,0142***
Idade <sup>2</sup>	-0,217778D-04	-0,219838D-04
Raça	-0,1611***	-0,1611***
Metropolitana	-0,0747	-0,0746

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Urbana	0,0648	0,0648
Renda familiar <i>per capita</i>	-0,0382***	-0,0381***
Até 3 anos de estudo	-0,2648***	-0,2649***
De 4 a 7	-0,2496***	-0,2498***
De 8 a 10	-0,1067***	-0,1067***
De 11 a 14	0,0321	0,0321
Filhos de 0 e 1 ano	0,0195	0,0197
Filhos de 2 a 7	0,0802***	0,0805***
Filhos de 8 a 15	0,0640***	0,0643***
Filhos de 16 a 25	0,0727***	0,0730***
Saúde auto-avaliada	-0,0538**	-0,0540**
Número de componentes na família	-0,0365***	-0,0366***
Empregado sem carteira	-0,0287	-0,0300
Trabalhador doméstico	-0,4198***	-0,4220***
Conta-própria	-0,3357***	-0,3364***
Empregador	-0,7708***	-0,7719***
Não-remunerado	-0,0945	-0,0947
Não-PEA	0,0724**	0,0736**
Desocupados	-0,1023	-0,1001
Sigma	0,9274***	1,1838***
Rho	0,0366++	-0,191312D-06

\* Significativo a 10%.

\*\* Significativo a 5%.

\*\*\* Significativo a 1%.

TABELA II.3  
MODELO DE SELEÇÃO PARA DADOS DE CONTAGEM: PLANO PRIVADO VIA EMPREGADOR

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
<b>Equação de utilização</b>		
Constante	1,7762***	-0,6083**
Norte	-0,2010***	0,9024***
Nordeste	0,0239	0,4824***
Sudeste (exceto São Paulo)	-0,0058	0,6713***
Sul	0,0568***	0,0773*
Centro-Oeste	0,0568***	0,6860***
Sexo	-0,5260***	-0,6347***
Idade	-0,0126***	-0,0071
Idade <sup>2</sup>	0,792919D-04***	-0,0001
Raça	0,0130	0,4496***
Metropolitana	0,2898***	0,5093***
Urbana	0,0747***	0,5632***
Renda familiar <i>per capita</i>	0,0250***	-0,1007***
Coluna	0,2312***	0,0729*
Artrite	0,0756***	-0,1171**
Câncer	0,6770***	1,4117***
Diabetes	0,1360***	0,6881***
Asma	0,1126***	0,2692***
Hipertensão	0,1570***	0,1184***
Coração	0,2132***	1,1234***
Renal	0,1265***	0,2694***
Depressão	0,3335***	0,1542***
Tendinite	0,1831***	-0,0970
Cirrose	0,1618	2,3246***

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Tuberculose	1,2810***	Omitida
Até 3 anos de estudo	-0,1337***	-0,2626***
De 4 a 7 do	-0,1788***	0,0496
De 8 a 10	-0,1485***	-0,3282***
De 11 a 14	-0,0760***	-0,2890***
Filhos de 0 e 1 ano	0,0632***	0,2265***
Filhos de 2 a 7	-0,0886***	-0,5179***
Filhos de 8 a 15	0,0196**	0,2133***
Filhos de 16 a 25	0,0505***	0,1156***
Saúde auto-avaliada	-0,6165***	-0,9956***
Casal com filhos	-0,0311	0,2522***
Mãe com filhos	-0,1023***	-0,3245***
Outro tipo de família	-0,1190***	0,4033***
Acesso	-0,0900**	0,1786***
Equação de seleção		
Constante	0,2463*	0,2463*
Norte	-0,8714***	-0,8714***
Nordeste	-0,4662***	-0,4662***
Sudeste (exceto São Paulo)	-0,2412***	-0,2412***
Sul	-0,1395***	-0,1395***
Centro-Oeste	-0,5441***	-0,5441***
Sexo	0,3145***	0,3145***
Idade	-0,0214***	-0,0214***
Idade <sup>2</sup>	0,361317D-04	0,361317D-04
Raça	-0,0077	-0,0077
Metropolitana	0,1237**	0,1237**

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Urbana	0,0262	0,0262
Log renda familiar <i>per capita</i>	0,0304***	0,0304***
Até 3 anos de estudo	0,3721***	0,3721***
De 4 a 7	0,3537***	0,3537***
De 8 a 10	0,2009***	0,2009***
De 11 a 14	0,1157***	0,1157***
Filhos de 0 e 1 ano	-0,0189	-0,0189
Filhos de 2 a 7	-0,0534***	-0,0534***
Filhos de 8 a 15	-0,0168	-0,0168
Filhos de 16 a 25	-0,0169	-0,0169
Saúde auto-avaliada	0,0858***	0,0858***
Número de componentes na família	0,0636***	0,0636***
Empregado sem carteira	-0,6592***	-0,6592***
Trabalhador doméstico	-0,4835***	-0,4835***
Conta-própria	-1,0234***	-1,0234***
empregador	-1,0137***	-1,0137***
Não-remunerado	-0,8106***	-0,8106***
Não-PEA	-0,7306***	-0,7306***
Desocupados	-0,7519***	-0,7519***
Sigma	0,3909***	0,8512***
Rho	-0,537193D-08	-0,680729D-07

\* Significativo a 10%.

\*\* Significativo a 5%.

\*\*\* Significativo a 1%.

TABELA II.4  
MODELO DE SELEÇÃO PARA DADOS DE CONTAGEM: PLANO VOLUNTÁRIO

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
<b>Equação de utilização</b>		
Constante	1,6615***	-2,0924***
Norte	-0,1547***	-0,0516
Nordeste	0,0877***	0,2867***
Sudeste (exceto São Paulo)	0,0615***	0,1282***
Sul	0,0075	0,1540***
Centro-Oeste	0,0443**	-0,0735**
Sexo	-0,4752***	0,1599***
Idade	-0,0077***	0,0147***
Idade <sup>2</sup>	0,595880D-04***	0,477500D-04
Raça	0,0058	0,0241
Metropolitana	0,2711***	0,6284***
Urbana	-0,0018	0,4720***
Renda familiar <i>per capita</i>	0,0042	-0,0888***
Coluna	0,1300***	-0,0422
Artrite	-0,0228	0,0867***
Câncer	0,8279***	1,0223***
Diabetes	0,3576***	0,3269***
Asma	0,2894***	0,3265***
Hipertensão	0,1687***	-0,0881***
Coração	0,2547***	0,2760***
Renal	0,2045***	1,0539***
Depressão	0,2844***	0,1973***
Tendinite	0,0928***	0,2317***
Cirrose	0,4371***	0,7213***

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Tuberculose	0,0291	Omitida
Até 3 anos de estudo	-0,2004***	0,2216***
De 4 a 7	-0,0281	0,3407***
De 8 a 10	0,0098	0,0614*
De 11 a 14	-0,0320*	0,0293
Filhos de 0 e 1 ano	0,1137***	0,6027***
Filhos de 2 a 7	-0,1424***	-0,4572***
Filhos de 8 a 15	-0,0396***	-0,1316***
Filhos de 16 a 25	0,0251***	-0,0925***
Saúde auto-avaliada	-0,5300***	-0,4209***
Casal com filhos	0,0025	0,3006***
Mãe com filhos	-0,1042***	0,2016***
Outro tipo de família	-0,1133***	0,0562
Acesso	-0,2764***	-0,7081***
Equação de seleção		
Constante	-1,0399***	-1,0399***
Norte	-0,2492***	-0,2492***
Nordeste	-0,1978***	-0,1978***
Sudeste (exceto São Paulo)	-0,1033***	-0,1033***
Sul	-0,3913***	-0,3913***
Centro-Oeste	-0,4641***	-0,4641***
Sexo	-0,2067***	-0,2067***
Idade	0,0210***	0,0210***
Idade <sup>2</sup>	-0,0002***	-0,0002***
Raça	0,1882***	0,1882***
Metropolitana	-0,0597	-0,0597

(continua)

(continuação)

Variáveis	Coeficientes estimados	
	Consultas médicas	Dias de internação
Urbana	-0,0918*	-0,0918*
Log renda familiar <i>per capita</i>	0,0089	0,0089
Até 3 anos de estudo	-0,1498***	-0,1498***
De 4 a 7	-0,1565***	-0,1565***
De 8 a 10	-0,1400***	-0,1400***
De 11 a 14	-0,1776***	-0,1776***
Filhos de 0 e 1 ano	-0,0291	-0,0291
Filhos de 2 a 7	-0,0597***	-0,0597***
Filhos de 8 a 15	-0,0802***	-0,0802***
Filhos de 16 a 25	-0,0655***	-0,0655***
Saúde auto-avaliada	-0,0239	-0,0239
Número de componentes na família	-0,0157	-0,0157
Empregado sem carteira	0,8079***	0,8079***
Trabalhador doméstico	1,0005***	1,0005***
Conta-própria	1,3170***	1,3170***
Empregador	1,5858***	1,5858***
Não-remunerado	0,9640***	0,9640***
Não-PEA	0,7231***	0,7231***
Desocupados	0,9375***	0,9375***
Sigma	0,4251***	0,5465***
Rho	-0,343565D-08	-0,735829D-08

\* Significativo a 10%.

\*\* Significativo a 5%.

\*\*\* Significativo a 1%.

TABELA II.5  
MÉDIA E VARIÂNCIA DAS VARIÁVEIS SEGUNDO A AMOSTRA DE INDIVÍDUOS

Variáveis	Toda a amostra	Amostra com plano	Amostra sem plano	Amostra com plano público	Amostra com plano privado	Amostra com plano voluntário
Norte	0,69 (0,25)	0,54 (0,22)	0,07 (0,26)	0,95 (0,29)	0,28 (0,16)	0,04 (0,21)
Nordeste	0,29 (0,45)	0,20 (0,40)	0,33 (0,47)	0,25 (0,43)	0,17 (0,37)	0,20 (0,40)
Sudeste (exceto São Paulo)	0,21 (0,41)	0,23 (0,42)	0,20 (0,40)	0,20 (0,40)	0,23 (0,42)	0,26 (0,44)
Estado de São Paulo	0,12 (0,32)	0,18 (0,39)	0,10 (0,30)	0,07 (0,27)	0,23 (0,42)	0,22 (0,41)
Sul	0,17 (0,38)	0,20 (0,40)	0,17 (0,37)	0,19 (0,39)	0,23 (0,42)	0,18 (0,38)
Centro-Oeste	0,11 (0,31)	0,10 (0,30)	0,11 (0,310)	0,16 (0,37)	0,08 (0,27)	0,07 (0,26)
Sexo	0,72 (0,44)	0,71 (0,45)	0,73 (0,44)	0,71 (0,45)	0,82 (0,37)	0,67 (0,46)
Idade	44,41 (15,70)	45,46 (14,79)	44,03 (16,00)	45,46 (14,79)	39,90 (11,76)	48,95 (14,36)
Raça	0,52 (0,49)	0,67 (0,46)	0,47 (0,49)	0,67 (0,46)	0,67 (0,46)	0,74 (0,43)
Metropolitana	0,40 (0,49)	0,56 (0,49)	0,35 (0,47)	0,56 (0,49)	0,58 (0,49)	0,56 (0,49)
Urbana	0,44 (0,49)	0,40 (0,49)	0,45 (0,49)	0,40 (0,49)	0,38 (0,48)	0,39 (0,48)
Rural	0,15 (0,35)	0,31 (0,17)	0,19 (0,39)	0,03 (0,17)	0,03 (0,17)	0,03 (0,17)
Renda familiar <i>per capita</i> (ln)	4,70 (2,14)	5,86 (1,47)	4,30 (2,19)	5,86 (1,47)	5,79 (1,24)	6,08 (1,60)
Coluna	0,31 (,046)	0,26 (0,44)	0,32 (0,46)	0,26 (0,44)	0,21 (0,40)	0,27 (0,44)
Artrite	0,14 (0,35)	0,10 (0,31)	0,16 (0,36)	0,10 (0,31)	0,58 (0,23)	0,12 (0,33)
Câncer	0,003 (0,62)	0,004 (0,67)	0,003 (0,60)	0,004 (0,67)	0,002 (0,51)	0,005 (0,07)

(continua)

(continuação)

	Toda amostra	Amostra com plano	Amostra sem plano	Amostra com plano público	Amostra com plano privado	Amostra com plano voluntário
Diabetes	0,36 (0,18)	0,46 (0,21)	0,33 (0,17)	0,46 (0,21)	0,02 (0,16)	0,06 (0,23)
Asma	0,42 (0,20)	0,40 (0,19)	0,43 (0,20)	0,40 (0,19)	0,03 (0,18)	0,04 (0,19)
Hipertensão	0,19 (0,39)	0,20 (0,40)	0,19 (0,39)	0,20 (0,40)	0,14 (0,35)	0,24 (0,43)
Coração	0,07 (0,26)	0,07 (0,26)	0,07 (0,25)	0,76 (0,26)	0,04 (0,20)	0,09 (0,28)
Renal	0,04 (0,21)	0,32 (0,17)	0,52 (0,22)	0,32 (0,17)	0,02 (0,14)	0,03 (0,18)
Depressão	0,78 (0,26)	0,76 (0,26)	0,78 (0,26)	0,07 (0,26)	0,04 (0,21)	0,08 (0,27)
Tuberculose	0,001 (0,04)	0,0006 (0,02)	0,001 (0,44)	0,006 (0,02)	0,0003 (0,01)	0,0006 (0,02)
Tendinite	0,31 (0,17)	0,42 (0,20)	0,27 (0,16)	0,42 (0,20)	0,03 (0,19)	0,04 (0,20)
Cirrose	0,002 (0,54)	0,002 (0,48)	0,003 (0,56)	0,002 (0,04)	0,001 (0,04)	0,002 (0,04)
Até 3 anos de estudo	0,35 (0,47)	0,11 (0,32)	0,43 (0,49)	0,11 (0,32)	0,09 (0,28)	0,11 (0,32)
De 4 a 7	0,30 (0,46)	0,23 (0,42)	0,33 (0,47)	0,23 (0,42)	0,24 (0,43)	0,22 (0,41)
De 8 a 10	0,12 (0,33)	0,15 (0,36)	0,11 (0,31)	0,15 (0,36)	0,17 (0,38)	0,14 (0,35)
De 11 a 14	0,14 (0,35)	0,30 (0,45)	0,09 (0,29)	0,30 (0,45)	0,31 (0,46)	0,26 (0,44)
Mais de 15	0,06 (0,24)	0,19 (0,39)	0,16 (0,12)	0,19 (0,39)	0,16 (0,37)	0,23 (0,42)
Filhos de 0 e 1 ano	0,12 (0,35)	0,09 (0,30)	0,13 (0,37)	0,83 (0,28)	0,12 (0,35)	0,06 (0,26)
Filhos de 2 a 7	0,39 (0,70)	0,28 (0,57)	0,43 (0,74)	0,29 (0,60)	0,38 (0,63)	0,21 (0,49)

(continua)

(continuação)

	Toda amostra	Amostra com plano	Amostra sem plano	Amostra com plano público	Amostra com plano privado	Amostra com plano voluntário
Filhos de 8 a 15	0,55 (0,91)	0,45 (0,76)	0,58 (0,95)	0,47 (0,79)	0,55 (0,82)	0,36 (0,68)
Filhos de 16 a 25	0,40 (0,79)	0,40 (0,77)	0,40 (0,80)	0,44 (0,80)	0,38 (0,76)	0,41 (0,76)
Saúde auto-avaliada	0,68 (0,46)	0,76 (0,42)	0,65 (0,47)	0,76 (0,42)	0,83 (0,37)	0,73 (0,43)
Casal	0,12 (0,33)	0,14 (0,35)	0,12 (0,32)	0,12 (0,33)	0,13 (0,33)	0,16 (0,37)
Casal com filhos	0,55 (0,49)	0,55 (0,49)	0,54 (0,49)	0,53 (0,49)	0,67 (0,46)	0,50 (0,50)
Mãe com filhos	0,17 (0,37)	0,15 (0,36)	0,18 (0,38)	0,18 (0,38)	0,09 (0,28)	0,16 (0,37)
Outros tipos de família	0,15 (0,35)	0,15 (0,35)	0,14 (0,35)	0,15 (0,36)	0,10 (0,30)	0,16 (0,37)
Acesso	0,05 (0,23)	0,24 (0,15)	0,68 (0,25)	0,02 (0,16)	0,01 (0,12)	0,01 (0,13)
Número de componentes na família	3,47 (1,64)	3,25 (1,40)	3,55 (1,71)	3,30 (1,45)	3,47 (1,35)	3,12 (1,35)
Empregado com carteira	0,29 (0,45)	0,50 (0,49)	0,21 (0,41)	0,54 (0,49)	0,77 (0,41)	0,23 (0,42)
Empregado sem carteira	0,12 (0,32)	0,05 (0,22)	0,14 (0,35)	0,05 (0,22)	0,03 (0,18)	0,06 (0,24)
Trabalhador doméstico	0,03 (0,18)	0,009 (0,09)	0,04 (0,20)	0,005 (0,07)	0,006 (0,07)	0,01 (0,11)
Conta-própria	0,22 (0,42)	0,11 (0,31)	0,27 (0,44)	0,08 (0,27)	0,04 (0,21)	0,20 (0,40)
Empregador	0,04 (0,20)	0,07 (0,26)	0,03 (0,18)	0,02 (0,16)	0,03 (0,18)	0,16 (0,37)
Não-remunerado	0,02 (0,14)	0,01 (0,10)	0,02 (0,15)	0,01 (0,10)	0,004 (0,06)	0,01 (0,12)
Não-PEA	0,21 (0,40)	0,21 (0,41)	0,21 (0,40)	0,25 (0,43)	0,08 (0,27)	0,26 (0,44)
Desocupados	0,04 (0,20)	0,01 (0,13)	0,04 (0,21)	0,01 (0,13)	0,01 (0,10)	0,02 (0,15)

### ANEXO III

#### DESCRÍÇÃO DAS VARIÁVEIS

		<b>tbl_PlanoMedico</b>	<b>tbl_ContasMedicas</b>
CodSabesperv	Código unificado da Sabesperv		Código unificado da Sabesperv
SeqSabesperv	Desconsiderar		Número da guia
PlanoMedico	Código completo da carteirinha		Número de seqüencial de itens dentro da guia
PlanoAtual	Plano atual		Código do credenciado
DtInclPlano	Data de inclusão no plano		Código do procedimento AMB
CodMotAlteracao	Código do motivo de alteração		Tipo da tabela AMB utilizada
DescMotAlteracao	Descrição do motivo de alteração		Data do atendimento
DtAlteracao	Data de alteração		Data do pagamento pago pelo atendimento
DescModPagamento	Modo de pagamento		Ano/mês de referência de pagamento
		<b>tbl_Beneficiarios</b>	<b>tbl_CredenciadoMedico</b>
CodSabesperv	Código unificado da Sabesperv		Tipo de internação
Nome	Nome completo		Descrição do tipo de internação
Sexo	Sexo		Descrição da função cirúrgica
DtNascimento	Data de nascimento		Descrição do tipo da cirurgia
DescEstCivil	Estado civil		Quantidade cobrada do procedimento AMB
Endereco	Endereço + número + complemento		Quantidade de CH utilizado
Bairro	Bairro		Valor do serviço
Cidade	Cidade		Valor do reembolso
Estado	Estado		Data da internação
CEP	CEP		Data de alta
DDD	DDD		Código internacional de doença (CID)
Telefone	Telefone		Valor da glosa
TelefoneRecado	Telefone para recado		PlanoMedico
Banco	Número do banco		
Agencia	Número da agência		
ContaCorrente	Número da conta corrente		
DescGrParent	Grau de parentesco em relação ao titular		
CIC	Número do CPF		
RG	Número do RG		
		<b>tbl_CadastroEmpregados</b>	<b>tbl_AMB</b>
CodSabesperv	Código unificado da Sabesperv		Tipo da tabela AMB utilizada
SeqSabesperv	Desconsiderar		Código do procedimento AMB
Nome	Nome completo		Descrição do procedimento AMB
DescTpEmpregado	Descrição do tipo de empregado		Desconsiderar
Unidade	Unidade de trabalho do empregado		Desconsiderar
Salario	Salário		Desconsiderar
PercentualATS	Percentual do adicional por tempo de serviço		Porte_Anes
ComissFuncao	Percentual da comissão		
ComissCargoConf	Gratificação de cargo		
CentroCusto	Centro de custo		
NomeCargo	Nome do cargo		
DescTpCargo	Tipo do cargo		
DescSitAtual	Situação atual		
DtSituacaoAtual	Data da situação atual		
DescGrInstrucao	Grau de instrução		
DtAdmissao	Data de admissão		

## ANEXO IV

### GASTOS NÃO COBERTOS PELOS PLANOS

Não se incluem na assistência médico-hospitalar prestada pelo Sabesperv-Saúde, os seguintes serviços e materiais:

I - Procedimentos não-éticos, experimentais ou não reconhecidos pela prática médica.

II - Procedimentos clínicos ou cirúrgicos com finalidade estética, excetuando-se as cirurgias plásticas reparadoras decorrentes de acidentes pessoais, sequelas de cirurgias mutilantes, de queimaduras e defeitos congênitos, mediante a realização de perícias prévias.

III - Internações em casa de repouso, clínicas de idosos ou de emagrecimento estético.

IV - Escleroterapia estética de veias superficiais de membros inferiores.

V - Transplantes, à exceção dos mencionados no Artigo 16, inciso XII desse regulamento.

VI - Atendimentos domiciliares de rotina ou eletivos, excetuando-se os previstos em subprograma específico.

VII - Atendimento em especialidades não reconhecidas pelo Conselho Federal de Medicina (CFM).

VIII - Aluguel de equipamentos para uso domiciliar, excetuando-se os previstos em subprograma específico.

IX - Reembolso de serviço de auxiliares/instrumentadores cirúrgicos não-médicos.

X - Fertilização *in vitro*, bem como procedimentos de apoio a essa terapêutica.

XI - Próteses e órteses não implantadas cirurgicamente (botas ortopédicas, palmilhas, óculos, lentes de contato, meias elásticas, cintas abdominais, pernas e braços mecânicos etc.) e as com finalidade estética.

XII - Próteses de implante cirúrgico importadas na existência de similar nacional e de menor custo.

XIII - Enfermagem em caráter particular, excetuando-se as previstas em subprograma específico.

XIV - Terapia ocupacional.

XV - Vacinas preventivas constantes do calendário nacional de imunizações.

XVI - Condicionamento físico, à exceção do paciente cardíaco.

XVII - Tratamento de dependência química aos participantes não-titulares em situações que não de emergência, com crise de risco de vida ou de danos físicos, ou ainda que extrapolem o prazo aqui estabelecido de cobertura de 15 (quinze) dias de internação por ano civil para quadros de intoxicação ou abstinência.

XVIII - Medicamentos não-registrados na Divisão de Controle de Medicamentos do Ministério da Saúde (Dimed), ou importados não nacionalizados.

XIX - Medicamentos para tratamento domiciliar e utilizados em regime ambulatorial à exceção de quimioterápicos, antineoplásicos para tumores malignos, e implantes radioterápicos.

XX - Procedimentos odontológicos, salvo os envolvendo cirurgia bucomaxilofacial.

XXI - Remoções aéreas.

XXII - Tratamento de enfermidades ou acidentes decorrentes de perturbações da ordem pública, atentados, operações militares, convulsões sociais e catástrofes públicas, quando declarados por autoridade competente.

XXIII - Procedimentos e tratamentos realizados no exterior.

## **ANEXO V**

### **CÓDIGOS DA AMB UTILIZADOS NA CONSTRUÇÃO DO BANCO DE DADOS**

Consultas (estes códigos não caracterizam pacientes internados)

00010006 - Consulta oftalmológica

00010014 - Consulta em consultório

00010030 - Consulta oftalmológica

00010073 - Consulta médica em pronto socorro

Materiais e Medicamentos (estes códigos não caracterizam pacientes internados)

89030028 - Materiais em consultório, ambulatório ou P.S.

89040023 - Medicamentos e gases usados ambulatorialmente

- 98013017 - Materiais e/ou filmes radiológicos ambulatoriais  
98014013 - Medicamentos/gases e/ou contrastes ambulatoriais  
98043013 - Materiais usados em serviços especiais (Sadt)  
98044010 - Medicamentos e/ou gases em serviços especiais  
98063014 - Materiais e/ou filmes radiológicos em medicina ocupacional  
98064010 - Medicamentos/gases e/ou contrastes em medicina ocupacional  
98513010 - Materiais e/ou filmes radiológicos ambulatoriais  
98514016 - Medicamentos/gases e/ou contrastes ambulatoriais  
98563017 - Materiais e/ou filmes radiológicos em medicina ocupacional

Diárias e Taxas (estes códigos não caracterizam pacientes internados)

- 27040330 - Taxa de irradiação de unidade hemoterápica  
27040348 - Taxa de utilização de bolsa plástica p/ unidade  
27040356 - Taxa de utilização de descartáveis p/ aplicação  
27040461 - Taxa de leucotização p/componente  
29010012 - Aerossolterapia sem pressão positiva  
29010020 - Aerossolterapia com pressão positiva  
29010039 - Prova de complacência pulmonar  
29010047 - Curva fluxo-volume estudo dos fluxos pulmonares  
29010055 - Determinação da mecânica respiratória  
29010063 - Espirografia simples  
29010080 - Estudo da função pulmonar (pulgões em separado)  
29010128 - Medida da difusão alvéolo-capilar  
29010136 - Prova ventilatória completa (05+06+07+14+15)  
29010144 - Prova fármaco-dinâmica  
29010179 - Reeducação funcional respiratória  
33019991 - Uso de Doppler/colorido  
88010015 - Tratamento ambulatorial intensivo fase I

- 88010023 - Programa para recaída
- 89020014 - Taxa de sala cirúrgica pequena
- 89020057 - Taxa de sala cirúrgica ambulatorial
- 89029992 - Taxas diversas
- 89050010 - Remoção simples — sem médico acompanhante
- 89050029 - Remoção com acompanhamento médico
- 98012010 - Taxas diversas ou taxa de observação
- 98037013 - Taxas, materiais e/ou medicamentos em Sadt
- 98042017 - Taxas diversas em serviços especiais (Sadt)
- 98045016 - Remoção de pacientes ambulatoriais
- 98061011 - Diária (qualquer) de paciente de medicina ocupacional
- 98062018 - Taxas diversas/de observação em medicina ocupacional
- 98065017 - Despesas de remoção em medicina ocupacional
- 98512013 - Taxas diversas ou taxa de observação
- 98537016 - Taxas, materiais e/ou medicamentos em Sadt

#### Diagnose

- 29020158 - Nebulização simples
- 50070053 - Vitrectomia anterior
- 50070061 - Vitrectomia (vias pars plana)
- 50100033 - Cirurgias antiglaucomatosas (qualquer técnica)
- 50140035 - Retinoplexia com introflexão escleral (qualquer técnica)
- 56010222 - Ereção fármaco-induzida

#### Terapia Psiquiatria / Psicologia

- 70010013 - Entrevista inicial em psiquiatria
- 70020043 - Psiquiatria: orientação familiar
- 70020051 - Psiquiatria: psicoterapia grupal

- 70020060 - Psiquiatria: psicoterapia individual
- 70030014 - Entrevista inicial em psicologia
- 70030022 - Segunda entrevista em psicologia
- 70040010 - Psicologia: avaliação completa (testes)
- 70040028 - Psicologia: avaliação de personalidade
- 70040052 - Psicologia: avaliação psicomotora
- 70040060 - Psicologia: teste de Rorsharch
- 70050023 - Psicologia: ludoterapia individual
- 70050031 - Psicologia: orientação familiar
- 70050040 - Psicologia: psicomotricidade grupal
- 70050058 - Psicologia: psicomotricidade individual
- 70050066 - Psicoterapia grupal
- 70050074 - Psicoterapia individual
- 72010029 - Audiometria tonal limiar
- 72010045 - Audiometria vocal limiar de discriminação
- 72010053 - Audiometria vocal limiar de inteleibilidade
- 72010126 - Impedanciometria (inclui timpanometria)
- 72010185 - Pesquisa de pares cranianos (VIII par)
- 72010207 - Rinomanometria
- 72010266 - Testes vestibulares sem eletronistagmografia
- 72010282 - Testes vestibulares com vecto-eletronistagmo
- 72010290 - Teste decay do reflexo estapédico
- 72010312 - Pesquisa do nistagmo optocinético
- 72010320 - Pesquisa de nistagmo pendular
- 72010371 - Audiometria ocupacional ou de seleção
- 72020016 - Fonoaudiologia: entrevista c/paciente ou pais
- 72020024 - Avaliação fonoaudiológica
- 72030011 - Fonoaudioterapia individual
- 72030038 - Orientação de pais

## Terapias Pré e Pós-Operatórias

- 25010050 - Exame muscular
- 25010077 - Pesquisa eletroneuromiográfica
- 25020013 - Fisioterapia em ataxias
- 25020021 - Fisioterapia em distrofia simpático reflexo
- 25020030 - Fisioterapia em hemiplegia e hemiparesia
- 25020048 - Fisioterapia em lesão nervosa periférica (+um nervo)
- 25020056 - Fisioterapia em lesão nervosa periférica (um nervo)
- 25020064 - Fisioterapia em miopatias
- 25020072 - Fisioterapia em monoplegia
- 25020080 - Fisioterapia em paraplegia e paraparesia
- 25020110 - Fisioterapia em moléstia de Parkinson
- 25020129 - Fisioterapia em quadriplegia e quadriparesia
- 25020137 - Fisioterapia em retardo do desenvolvimento psicomotor
- 25020153 - Assistência fisiátrica em neurologia
- 25030019 - Fisioterapia em distúrbios circulatórios
- 25040014 - Assistência fisiátrica respiratória — doente clínico internado
- 25040022 - Assistência fisiátrica respiratória — pac. intern. vent. mec.
- 25040030 - Fisioterapia respiratória em pré/pós-operatório
- 25040049 - Doenças pulmonares atendidas em ambulatório
- 25040057 - Fisioterapia em paciente com DPOC
- 25050010 - Fisioterapia em reumatologia (1 membro)
- 25050028 - Fisioterapia em reumatologia (+ de um membro)
- 25050036 - Fisioterapia em um segmento da coluna vertebral
- 25050044 - Fisioterapia em mais de 1 segmento da coluna vertebral
- 25050052 - Fisioterapia p/ coluna vertebral e membros
- 25060015 - Fisioterapia em amputação unilateral
- 25060031 - Fisioterapia para contusões

- 25060040 - Fisioterapia desvios posturais coluna vertebral  
25060058 - Fisioterapia para entorses  
25060066 - Fisioterapia pós-operatória/imobiliz. 1 membro  
25060074 - Fisioterapia pós-operat./imobiliz.+ de 1 membro  
25060082 - Fisioterapia pós-operat./imobiliz. coluna vertebral  
25060090 - Fisioterapia articulação têmporo-mandibular  
25060104 - Fisioterapia p/ seqüela de traumatismos dos tendões  
25060112 - Fisioterapia traumatismos torácicos-abdominais  
25060139 - Fisioterapia ortopédica de um membro  
25060147 - Fisioterapia ortopédica de mais de um membro  
25060171 - Fisioterapia pós-operat./imobiliz. de 1 membro  
25060180 - Fisioterapia pós-operat./imobiliz. mais de 1 membro  
25070010 - Fisioterapia p/ doença isquêmica do coração  
25070037 - Fisioterapia pós-operat. cirurgia cardíaca  
25070061 - Fisioterapia cardíaca ambulatorial  
25090011 - Fisioterapia pré/pós-operatória em neurologia  
25100017 - Doenças urológicas  
25100050 - Fisioterapia domiciliar  
30010012 - Quimioterapia sistêmica inicial (por 7 dias)  
30010020 - Quimioterapia sistêmica inicial (subseqüente 7 dias)  
30010128 - Quimioterapia sistêmica inicial  
30010136 - Quimioterapia sistêmica (por dia até 7)  
30020018 - Quimioterapia sistêmica por períodos de 7 dias  
30020026 - Quimioterapia sistêmica subseqüente por períodos de 7 dias  
30020034 - Quimioterapia ambulatorial (primeiro dia)  
30020042 - Quimioterapia ambulatorial (por dia até 7)  
35010010 - Radioterapia — megavoltagem (acel. linear etc.)  
35010029 - Radioterapia — betaterapia p/ aplicação

- 35010045 - Radioterapia — cheque — filme (filme à parte)
- 35010088 - Radioterapia — máscara ou blocos de colimação
- 35010100 - Radioterapia — simulação de tratamento
- 35010118 - Planejam. técnico (teleterapia/branquiterapia)
- 51010011 - Audiometria (tipo Von Bekesy)
- 88250016 - Fisioterapia domiciliar (neuro/resp/motora)
- 88370011 - Oxigênioterapia hiperbárica p/ meia hora

## **ANEXO VI**

TABELA VI.1  
**DISTRIBUIÇÃO DOS TITULARES DA SABESPREV, SEGUNDO IDADE E SEXO**

Grupo etário	Homem	(%)	Mulher	(%)	Total	(%)
15-19	0	0,05	3	0,03	3	0,00
20-24	36	0,73	18	1,36	54	0,08
25-29	492	1,50	924	5,08	1.416	2,09
30-34	1.014	2,40	3.438	9,03	4.452	6,57
35-39	1.626	2,79	6.117	10,68	7.743	11,43
40-44	1.890	3,07	7.233	11,23	9.123	13,47
45-49	2.082	3,25	7.605	14,25	9.687	14,30
50-54	2.202	2,21	9.654	12,69	11.856	17,51
55-59	1.500	1,35	8.592	7,23	10.092	14,90
60-64	915	0,84	4.899	4,04	5.814	8,58
65-69	567	0,60	2.733	2,46	3.300	4,87
70-74	405	0,53	1.665	1,04	2.070	3,06
75-79	357	0,28	702	0,58	1.059	1,56
80 e +	411	0,32	645	0,37	1.056	1,55

Fonte: Elaboração própria a partir dos dados da Sabesperv

**TABELA VI.2**  
**DISTRIBUIÇÃO DOS BENEFICIÁRIOS DA SABESPREV, SEGUNDO IDADE E SEXO**

Grupo etário	Homem	(%)	Mulher	(%)	Total	(%)
0-5	252	2,03	201	1,95	453	0,19
5-10	4.707	1,92	4.569	3,25	9.276	3,79
10-15	8.733	3,57	7.965	4,00	16.698	6,82
15-20	9.846	4,02	9.777	4,80	19.623	8,02
20-25	12.297	5,03	11.745	5,77	24.042	9,82
25-30	13.797	5,64	14.127	3,43	27.924	11,41
30-35	7.005	2,86	8.382	3,51	15.387	6,29
35-40	8.046	3,29	8.592	3,75	16.638	6,80
40-45	8.679	3,55	9.189	4,31	17.868	7,30
45-50	9.012	3,68	10.557	4,60	19.569	8,00
50-55	11.046	4,51	11.253	3,68	22.299	9,11
55-60	9.651	3,94	8.997	2,45	18.648	7,62
60-65	5.856	2,39	5.988	1,68	11.844	4,84
65-70	3.477	1,42	4.116	1,29	7.593	3,10
70-75	2.196	0,90	3.147	1,10	5.343	2,18
75-80	1.353	0,55	2.700	0,97	4.053	1,66
80 e +	2.289	0,94	5.169	1,96	7.458	3,05

Fonte: Elaboração própria a partir dos dados da Sabesperv.

TABELA VI.3  
**DESCRIÇÃO DA SITUAÇÃO ATUAL DOS TITULARES DO PLANO DE SAÚDE SABESPREV — 2003**

Descrição da situação atual	Número	(%)
Acidente do trabalho	52	0,23
Aposentadoria — inval. outros	3	0,01
Ativo	93	0,41
Auxílio doença	451	2,00
Auxílio doença — crem. Individual	26	0,12
Auxílio maternidade	25	0,11
Dem. s/ justa causa	2	0,01
Dem. término contrato/prazo determinado	3	0,01
Demissão voluntária / incentivada	109	0,48
Desligado — aposentado	1	0,00
Desligado — aposentado por inval. acidente de trabalho	7	0,03
Desligado — aposentado por inval. doença profissão	1	0,00
Desligado — aposentado por inval. outros	173	0,77
Desligado — no período de experiência	18	0,08
Desligado — pedido de demissão	264	1,17
Desligado justa causa	33	0,15
Desligado sem justa causa	767	3,40
Desligado término contrato/prazo determinado	57	0,25
Falecimento	288	1,28
Falecimento por acidente de trabalho	12	0,05
Férias	219	0,97
Licença-maternidade	1	0,00
Licença-prêmio	11	0,05
Licença sem direito a vencimentos	16	0,07
Pedido de demissão	218	0,97
Retorno	16.232	87,38

Fonte: Elaboração própria a partir dos dados da Sabesperv.

**TABELA VI.4  
MOTIVO DA SAÍDA DOS TITULARES DO PLANO DE SAÚDE SABESPREV — 2003**

Descrição da situação atual	Número	(%)
Desligado — aposentado	1	0,06
Desligado — aposentado por invalidez acidente de trabalho	7	0,43
Desligado — aposentado por invalidez doença da profissão	1	0,06
Desligado — aposentado por invalidez outros	173	10,57
Desligado — no período de experiência	18	1,10
Desligado — pedido de demissão	264	16,13
Desligado por justa causa	33	2,02
Desligado sem justa causa	767	46,85
Desligado término contrato/prazo determinado	57	3,48
Falecimento	288	17,59
Falecimento por acidente de trabalho	12	0,73
Licença sem direito a vencimentos	16	0,98

Fonte: Elaboração própria a partir dos dados da Sabesperv.

## Ipea – Instituto de Pesquisa Econômica Aplicada

### **Editorial**

#### **Coordenação**

Ronald do Amaral Menezes

#### **Supervisão**

Marcos Hecksher

#### **Revisão**

Lucia Duarte Moreira

Alejandro Sainz de Vicuña

Eliezer Moreira

Elisabete de Carvalho Soares

Marcio Alves de Albuquerque

Míriam Nunes da Fonseca

Roberta da Costa de Sousa

#### **Capa**

Joanna Silvestre Friques de Sousa

#### **Editoração**

Roberto das Chagas Campos

Carlos Henrique Santos Vianna

Joanna Silvestre Friques de Sousa

Emilia Teles da Silva

### **Brasília**

SBS – Quadra 1 – Bloco J – Ed. BNDES, 9º andar

70076-900 – Brasília – DF

Fone: (61) 3315-5090

Fax: (61) 3315-5314

Correio eletrônico: editbsb@ipea.gov.br

### **Rio de Janeiro**

Av. Nilo Peçanha, 50 – 6º andar (Grupo 609)

20044-900 – Rio de Janeiro – RJ

Fone: (21) 2215-1043/R. 233

Fax: (21) 2215-1043/R. 234

Correio eletrônico: editrj@ipea.gov.br

### **Comitê Editorial**

#### **Secretário-Executivo**

Marco Aurélio Dias Pires

SBS – Quadra 1 – Bloco J – Ed. BNDES

9º andar, sala 908

70076-900 – Brasília – DF

Fone: (61) 3315-5406

Correio eletrônico: madp@ipea.gov.br



A aplicação de métodos quantitativos na defesa da concorrência, ao contrário de desonerar os juristas, requer-lhes que repensem uma vasta gama de regras e princípios de direito: que disciplina outorgar a esses estudos econométricos do ponto de vista do direito da prova? Em que momento do processo administrativo, e sob quais regras, tais estudos devem ser introduzidos, quer pelos órgãos instrutórios, quer pelas partes ou terceiros interessados no processo?

Neste compêndio, os economistas encontrarão um primeiro passo na tentativa de sistematizar o ferramental que se tornará, inexoravelmente, parte do cotidiano da análise antitruste. De outro lado, os juristas hão de vislumbrar, aqui, uma nova ordem de desafios, instigantes e complexos.

Daniel K. Goldberg

*Secretário de Direito Econômico do Ministério da Justiça*

**Governo Federal**

**Ministério da Justiça**

**Secretaria de  
Direito Econômico**

**Ministério do Planejamento,  
Orçamento e Gestão**

**Ipea – Instituto de Pesquisa  
Econômica Aplicada**

ISBN 85-86170-79-8

