NOTAS DE AULA

Economia do Setor Público

Professor Victor Oliveira

DEPARTAMENTO DE ECONOMIA

SETOR DE CIÊNCIAS SOCIAIS APLICADAS

UNIVERSIDADE FEDERAL DO PARANÁ

Acordei hoje com tal nostalgia de ser feliz. Eu nunca fui feliz na minha vida inteira. Por dentro eu sempre me persegui. Eu me tornei intolerável para mim mesma. Vivo numa dualidade dilacerante. Eu tenho uma aparente liberdade mas estou presa dentro de mim. Eu queria uma liberdade olímpica. Mas essa liberdade só é conhecida aos seres imateriais. Enquanto eu tiver corpo ele me submeterá às suas exigências. Vejo a liberdade como uma forma de beleza e essa beleza me falta.

Clarice Lispector, Um Sopro de Vida

Sumário

1 Introdução

Young economists are often extremely diffident about presenting their work. Remember, though, good research is rarely done in solitary confinement; and appearing inexperienced in public leaves you no worse off than obscurity and anonymity. A presentation can improve your work, acquaint senior scholars with it, and raise your visibility with editors and potential employers.

Hamermesh, 1992

1.1 Utilidade

A ideia de utilidade como expressão de grau de felicidade é de Jeremy Bentham. Essa ideia já foi ultrapassada há muito tempo.

Ken Binmore (*Rational Decisions*, Princeton University Press, 2009, pp. 19–22) denomina essa concepção de falácia da utilidade causal, o termo "causal" aqui designando a ideia de que é a utilidade que explica a ação, não o contrário. O erro é propagado pelos próprios economistas neoclássicos nos livros-textos, quando dão a entender que "o indivíduo escolhe A em vez de B porque ele prefere A a B", mas não explicam o que isso realmente significa. Se você pensar bem, verá que essa concepção contraria a teoria da preferência revelada, segundo a qual é a observação empírica da escolha de A, em vez de B, que revela que o indivíduo prefere A a B, não o contrário. Para resumir o que disse até aqui, temos duas concepções distintas:

- teoria benthamista (concepção antiga) o indivíduo escolhe A em detrimento de B porque ele prefere A a B;
- teoria da preferência revelada (concepção moderna): é porque o indivíduo escolhe A em detrimento de B, que podemos dizer que ele prefere A a B.

É claro que a definição moderna de utilidade é dada em termos de equivalência: "a alternativa A é preferida a B se, e somente se, a utilidade de A é maior que a utilidade de B". Note, porém, que essa definição nada diz quanto à ação tomada: ela apenas conecta uma estrutura de preferências com uma representação numérica. Tudo permanece, portanto, no mundo das preferências, não da ação.

A concepção que a Teoria Econômica adota hoje é a da preferência revelada. Como diz Binmore, "o fato de que a utilidade costumava significar uma coisa e hoje significa outra, compreensivelmente é causa de muita confusão" (op. cit., p. 19). A concepção moderna liberta a teoria da preferência revelada de quaisquer pretensões psicológicas.

Para entendermos a moderna Teoria Econômica, devemos voltar a Lionel Robbins e a um termo que ele utilizou: consistência. Consistência nada mais é que transitividade. Ele chama de consistência porque é essa propriedade o que nos garante observar consistência nas escolhas dos indivíduos.

A melhor explicação para isso é evolucionária e quem a fornece é Armen Alchian. A natureza tende a eliminar os indivíduos que expressam intransitividades, não importa se consciente ou inconscientemente. Quando falamos de racionalidade em termos evolucionários, é preciso distinguir entre causas próximas e causas últimas. Isso é mais ou menos o que eu expliquei em outro texto sobre preferências instrumentais e intrínsecas. Por que os pássaros cantam na primavera? A causa próxima é que as cordas vocais emitem sons, motivadas por reações químicas e neuronais. A causa última é que assim eles sinalizam seu território e evitam conflitos desnecessários, além de sinalizar a disposição para reproduzir, garantindo a preservação da espécie. Os pássaros não sabem e nem se preocupam se esse comportamento é racional ou não, mas esse comportamento é racional, no sentido de ser o melhor comportamento para a sobrevivência. Se os pássaros fossem conscientemente racionais, tomariam essa decisão. Eles não são, mas os que sobreviveram aos tempos são justamente aqueles cujos comportamentos consistentemente se alinharam à causa última. De modo similar, Alchian argumenta que as forças econômicas tenderão a eliminar do mercado os investidores que consistentemente não procuram maximizar os lucros.

Observe que a teoria econômica não nega que eventualmente alguns indivíduos expressem intransitividades (ou inconsistência ou irracionalidade), como os críticos em geral gostam de alardear. O que os críticos não levam em conta é o aspecto evolucionário. A realização de intransitividades causa perdas no ambiente que prejudicam o indivíduo e o eliminam. É justamente esse aspecto que os críticos ignoram. Seguindo Binmore, suponha que um indivíduo X tem preferências cíclicas, $A > B > C > A$, e tem recursos de \$10. Um negociante Y lhe oferece A em troca de B e mais \$1, o que X aceita. Estamos aqui supondo que a unidade monetária é apropriada para que a troca acompanhada da cessão de $1 seja compatível com as preferências de X. O negociante Y então oferece C em troca de A e mais $1, o que X mais uma vez aceita. Depois Y lhe oferece B em troca de C e mais $1. Findo esse ciclo, X volta à situação inicial, só que $3 mais pobre. Y então repete o processo até que X fique sem recursos e seja eliminado.

A teoria da preferência revelada é uma filha da teoria neoclássica e, como tal, é a doutrina oficial da economia neoclássica, patente em todos os livros-textos (Binmore, p. cit., p. 20). Neste ponto, passo a palavra ao próprio Binmore:

> *O tipo de crítico que pensa que os economistas são sacanas, pessoas desajustadas que só pensam em dinheiro, geralmente ignoram o pensamento oficial em favor de um espantalho que é fácil de nocautear. Dizem que a economia neoclássica é baseada no princípio de que as pessoas são egoístas. (...) Não é verdade que o egoísmo é um axioma da teoria econômica. Eu suspeito que esse erro tão generalizado exista porque as pessoas pensam que os agentes racionais devem agir por auto-interesse porque eles maximizam sua própria função de utilidade, em vez de alguma outra função-objetivo social. Mas dizer tais coisas é mostrar que não entendeu nada da teoria da preferência revelada.*

O argumento evolucionário de Alchian significa que as pessoas agem como se maximizassem a utilidade, não que elas efetivamente maximizam. Elas nem pensam nisso, mas seu comportamento, se consistente, não será desconforme com isso.

Demos acima a definição de utilidade, aquela que todos conhecemos dos livros-textos, expressa em termos de uma equivalência entre ordenação no âmbito das preferências e uma correspondente ordenação no âmbito das utilidades numéricas. Nesse sentido, a utilidade de fato é uma construção baseada nas preferências, mas devemos entender que as preferências não são o elemento primitivo da teoria da preferência revelada. Os elementos primitivos são as escolhas observadas!

O objetivo do modelo de livro-texto é fornecer um *modus ratiocinandi*, um modo de conectar conceitos para explicar fenômenos, e não simplesmente achar que a teoria é igual ao modelo.

1.2 O Mercado

Considere o mercado de um bem qualquer, digamos, frango. É importante especificar o período de tempo em que a transação é relevante, por exemplo, por semana. Assim, quando se diz que um consumidor consome ou demanda um frango, deve-se subentender que ele demanda um frango por semana. O mesmo vale para a oferta e a mesma unidade de tempo deve subjazer a todas as quantidades mencionadas, a todas as utilidades e funções de custo consideradas. Sem essa especificação, os conceitos de oferta e demanda perdem totalmente o sentido.

Para simplificar, em vez de considerar uma curva de oferta agregada e uma de demanda agregada genéricas, impessoais, suponha que cada demandante demande um e apenas um frango e que, similarmente, cada ofertante oferte um e apenas um frango. Cada um deles terá um nome (os ofertantes e os demandantes). Os demandantes são ordenados de forma descendente a partir da valoração mais alta[1], ou seja, desde aquele que está disposto a pagar mais até aquele disposto a pagar menos, como na tabela abaixo.

Assim, o número 3 relativo a Maria significa que Maria possui a terceira maior disposição a pagar pelo frango (R$ 7), não que ela demande três unidades.

Os ofertantes também são pessoas, só que eles serão ordenados de forma ascendente, desde o que está disposto a ofertar por menos até aquele disposto a ofertar por mais, como na tabela abaixo.

Juntando tudo numa tabela só, temos o seguinte quadro de preços de demanda e de oferta para as diversas quantidades de frango transacionadas:

[1] O preço de demanda é o preço que as pessoas estão dispostas a pagar por bens e serviços quando uma determinada quantidade ou quantidade está disponível.

Tabela 1.1 – *Preços de Demanda*

Demandante	Preço de demanda
1. João	9
2. Pedro	8
3. Maria	7
4. Carla	$6,50$
5. Antônio	$5,50$
6. Dora	3
7. Gustavo	1
8. José	0
9. Patrícia	0
10. Amélia	0

Tabela 1.2 – *Preços de Oferta*

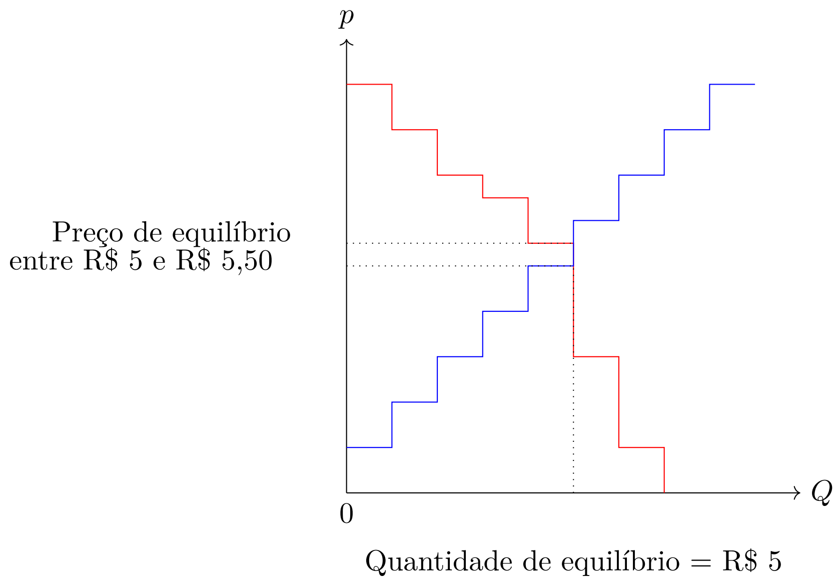
Ofertante	Preço de oferta
1. Catarina	1
2. Augusto	2
3. Roberto	3
4. Aline	4
5. Paula	5
6. Marcelo	6
7. Tiago	7
8. Luíza	8
9. Agnes	9
10. César	10

Tabela 1.3 – *Preços de Demanda e de Oferta por Unidade*

Quantidade	Preço de demanda	Preço de oferta
1	9	1
2	8	2
3	7	3
4	$6,50$	4
5	$5,50$	5
6	3	6
7	1	7
8	0	8
9	0	9
10	0	10

A lista de preços de oferta e de demanda pode ser melhor visualizada mediante sua representação gráfica. No eixo horizontal coloco as quantidades e no eixo vertical o valor monetário correspondente ao preço, seja de oferta ou de demanda, como na Figura 1.1.

Figura 1.1 – *Curvas de Oferta e de Demanda do Mercado de Frangos*



Pela tabela acima – e com a ajuda do gráfico –, podemos ver que ao preço de R$ 5 por frango, a quantidade ofertada por semana é igual à quantidade demandada. Um equilíbrio de mercado é o comércio de 5 frangos no período ao preço de R$ 5 a unidade.

Para mostrarmos que a quantidade de equilíbrio é de 5 unidades, vejamos o que ocorre quando a quantidade comercializada é de 4 unidades ou de 6 unidades. Se a quantidade comercializada fosse de 4 unidades, tanto o quinto ofertante (Paula) quanto o quinto demandante (Antônio) veriam a possibilidade de ganhos no comércio da quinta unidade. Com efeito, o demandante Antônio pagaria R$ 5,50 pelo frango e a ofertante Paula aceitaria R$ 5 por ele, de modo que existe espaço para a troca. Se o sexto ofertante, Marcelo, decidisse ofertar a sexta unidade, ele teria um custo marginal de R$ 6, mas a sexta demandante, Dora, só estaria disposta a pagar R$ 3 pela sexta unidade. Assim, Marcelo não conseguiria cobrir o seu custo marginal.

Disse acima que um equilíbrio de mercado é o comércio de 5 frangos no período ao preço de R$ 5 a unidade. Ora, outro equilíbrio seria o comércio de 5 frangos ao preço de R$ 5,50 a unidade. Procedendo do mesmo modo como fiz acima, é fácil ver que esse também é um equilíbrio. Mas qual é a diferença entre os dois? É o preço de equilíbrio! Paula exige no mínimo R$ 5 para ofertar a quinta unidade e Antônio está disposto a pagar no máximo R$ 5,50 por ela. Portanto, qualquer preço entre R$ 5 e R$ 5,50 pode equilibrar a oferta com a demanda. Qual preço prevalecerá? É nesse ponto que a Microeconomia abre espaço para a barganha e para isso há uma literatura abundante e cursos específicos. Independentemente do resultado da barganha, o importante é que o preço final acordado não afetará os ganhos totais de troca, afetará, porém, toda a distribuição dos ganhos de troca entre ofertantes e demandantes.

Com esta alocação final de equilíbrio, nenhuma outra revisão seria mutuamente aceitável. Esta é uma situação de compensação de mercado (*market-clearing*).

A interação entre a demanda e a oferta é importante não simplesmente porque estabelece um preço, mas porque, no processo, revela valores subjetivos relativos; estabelece um preço que permite que as pessoas troquem para que cada uma delas consiga uma combinação preferencial de bens.

Ninguém precisa conhecer a estrutura de demanda dos demais potenciais demandantes. Ninguém precisa conhecer a própria estrutura de demanda. Tudo o que é requerido é que diante de uma oportunidade de comprar ou vender, o indivíduo pode tomar uma decisão.

Qual o propósito dos conceitos de demanda e de oferta?

1. explicar como os mercados reduzem os custos dos agentes para ajustarem seu

consumo diante de mudanças nos gostos.

2. mostrar como a competição interpessoal pelos bens existentes é resolvida no mercado.

3. explicar como a negociação ou ajuste de preços facilita a realocação de bens.

4. ver como o mercado economiza os custos do agente para coletar informações.

5. comparar o sistema de negociação em uma situação de liberdade com uma situação em que haja restrições.

1.3 Teoremas do Bem-Estar

Como tal, o estudo dos mercados em economia está intrinsecamente ligado a problemas de divisão justa. Tais problemas perguntam: como um conjunto de mercadorias pode ser razoavelmente dividido entre várias partes? Obviamente, os problemas de divisão justa são de grande significado prático e os mercados são um mecanismo do mundo real pelo qual bens e serviços podem ser alocados entre várias partes. Existem muitos tipos de problemas de divisão justa, que dependem de diversos fatores, incluindo a natureza das preferências dos participantes, os tipos de bens divididos e os critérios de justiça desejados. A pesquisa sobre problemas de divisão justa abrangeu uma série de abordagens, incluindo estudos sobre a existência ou não de divisões justas, as propriedades de divisões justas e algoritmos para produzir divisões justas. Em tais problemas, os atores atribuem valores diferentes a bens diferentes, de acordo com suas próprias funções de utilidade.

Existem muitos critérios pelos quais a conveniência de uma divisão pode ser avaliada. Consequentemente, o sucesso do mercado pode ser medido pela conveniência da alocação produzida.

Na economia, no entanto, a noção de eficiência tem sido o principal critério pelo qual o sucesso do mercado na alocação de bens é avaliado. Em particular, a eficiência do mercado é considerada propriedade da otimização de Pareto: sob uma divisão ideal de Pareto, é impossível melhorar uma parte individual sem piorar a outra. Entendido em outras palavras, em divisões que não são ótimas para Pareto, é possível realocar de forma que todas as partes estejam em melhor situação.

Grosso modo, o primeiro teorema fundamental da economia do bem-estar afirma que os mercados competitivos tenderão ao equilíbrio de alocações eficientes. Serve de justificativa teórica para a eficácia dos mercados. Qual o motivo para conhecermos os teoremas do bem-estar? Se valessem as condições do primeiro e segundo teoremas de bem-estar na prática todo o problema do setor público teria solução teórica trivial. Senão vejamos. Para que apresentemos os teoremas de bem-estar precisamos de algumas definições.

Definição 1.1: Alocação Factível

Uma alocação $\left(\{\mathbf{x}^n\}_{n=1}^N, \{\mathbf{y}^h\}_{h=1}^H\right)$ é dita factível se $\sum\limits_n \mathbf{x}^n \leq \sum\limits_n \overline{\mathbf{x}}^n + \sum\limits_h \mathbf{y}^h$. Ou seja, alocações factíveis são aquelas tais que os indivíduos não consomem mais do que aquilo que existe após as decisões de produção das firmas.

Definição 1.2: Alocação Pareto-Eficiente

Uma alocação factível é dita Pareto-eficiente se não existe nenhuma outra alocação factível tal que $\mathbf{x}^n \succsim_n \tilde{\mathbf{x}}^n$ para todo n e $\mathbf{x}^n \succ_n \tilde{\mathbf{x}}^n$ para pelo menos um n. Uma alocação $\tilde{\mathbf{x}}$ é dita eficiente no sentido de Pareto se não existir uma forma de melhorar uma pessoa sem piorar outra.

Os dois teoremas de bem-estar vão relacionar alocações eficientes com as resultantes de um equilíbrio competitivo.

O primeiro teorema diz, essencialmente, que se todo bem relevante é negociado em um mercado com preços conhecidos publicamente (ou seja, se mercados são completos) e as firmas e os domicílios são tomadores de preços então o resultado de mercado é Pareto ótimo. Em poucas palavras, com mercados completos todo equilíbrio competitivo é necessariamente Pareto eficiente.

Formalmente, temos o teorema a seguir.

Definição 1.3: Primeiro Teorema do Bem-Estar

Seja $\left(\{\hat{\mathbf{x}}^n\}_{n=1}^N, \{\hat{\mathbf{y}}^h\}_{h=1}^H, \hat{\mathbf{p}}\right)$ um equilíbrio competitivo com nenhum consumidor localmente saciado, então $\left(\{\hat{\mathbf{x}}^n\}_{n=1}^N, \{\hat{\mathbf{y}}^h\}_{h=1}^H\right)$ é um ótimo de Pareto.

No caso do segundo teorema do bem-estar social, sua importância reside no fato de que, se válido, qualquer alocação eficiente pode ser atingida com uma simples redistribuição das dotações iniciais seguida do mecanismo de mercado.

Definição 1.4: Segundo Teorema do Bem-Estar

Suponha que $\left(\{\hat{\mathbf{x}}^n\}_{n=1}^N, \{\hat{\mathbf{y}}^h\}_{h=1}^H\right)$ é um ótimo de Pareto tal que pelo menos um domicílio não esteja saciado. Então, com:

1. preferências convexas
2. conjuntos de produção convexos
3. alocação $\hat{\mathbf{x}} \in \mathbb{X}^h, \ \forall h$
4. continuidade das preferências

existe $\hat{\mathbf{p}}$, tal que $\left(\{\hat{\mathbf{x}}^n\}_{n=1}^N, \{\hat{\mathbf{y}}^h\}_{h=1}^H, \hat{\mathbf{p}}\right)$ é um equilíbrio competitivo. Em palavras, se as preferências individuais e os conjuntos de possibilidade de produção das firmas são convexos, existe um conjunto completo de mercados com preços publicamente conhecidos e todos os agentes são tomadores de preços, então toda alocação Pareto eficiente pode ser alcançada como o equilíbrio competitivo para uma distribuição adequada das dotações iniciais.

Os pressupostos implícitos do PTBE são: i) não há externalidades no consumo; ii) economia competitiva e iii) existe um equilíbrio. As implicações do teorema são que os preços são estatística suficiente para todas as informações de que os agentes precisam para seu processo decisório.

A noção de eficiência do equilíbrio competitivo remonta a Adam Smith e sua metáfora da mão invisível. A hipótese fundamental subjacente aos teoremas está na existência de um conjunto completo de mercados. Como sabemos, a definição de bens nos dá grande flexibilidade para incorporar escolhas intertemporais e escolha sob incerteza. Porém, essa mesma flexibilidade torna a ideia de completeza muito mais delicada, em particular na presença de assimetrias de informação e de custos de transação. Uma outra crítica importante diz respeito à hipótese de concorrência propriamente dita, que elimina a possibilidade de poder de mercado em geral, o que não ocorre com muita facilidade. Há outras falhas de mercado diretamente concernentes ao estudo da economia do setor público, como bens públicos (quando o consumo de um bem por um agente não impede o consumo por outro agente), e externalidades (tanto no consumo quanto na produção).

As implicações do STBE são que os problemas de distribuição e alocação podem ser separados. Podemos redistribuir as dotações de bens para avaliar a riqueza dos agentes e usar os preços para indicar a escassez relativa. Valendo as condições do segundo teorema é que qualquer alocação eficiente pode ser atingida com uma simples redistribuição das dotações iniciais seguida do mecanismo de mercado. Se as preferências individuais e os conjuntos de possibilidade de produção das firmas são convexos, existe um conjunto completo de mercados com preços publicamente conhecidos e todos os agentes são tomadores de preços, então toda alocação Pareto eficiente pode ser alcançada como o equilíbrio competitivo para uma distribuição adequada das dotações iniciais. Redistribuição da alocação (transferências *lump-sum*) tem um papel fundamental na teoria de finanças públicas, porque permite uma solução eficiente do problema distributivo.

Para o segundo teorema, a hipótese de convexidade é especialmente problemática no que concerne ao conjunto de produção. Eliminam-se assim os retornos crescentes de escala que parecem caracterizar o processo produtivo de muitos bens. Para que uma transferência seja *lump-sum* é necessário que os domicílios não possam afetar o tamanho da transferência com mudanças em seu comportamento. Assim, uma contribuição uniforme é sempre (ou quase sempre) possível. Porém, não é uma transferência. Mais interessante é uma transferência efetiva das dotações iniciais, onde dotações de alguns agentes sejam transferidas para outros agentes. Nesse caso, é importante que o governo possa observar essas dotações para efetuar as transferências. Algumas das mais importantes dotações iniciais, porém, não são observáveis, como o talento, a inteligência, etc.

A alternativa para o governo seria perguntar para as pessoas quais as suas dotações iniciais e promover a transferência com base nessas informações. Naturalmente, as pessoas só falariam a verdade se isso fosse de seu interesse, o que tende a reduzir a importância prática do segundo teorema, mas define a essência do *trade-off* entre distribuição e eficiência.

No que se segue, vamos considerar os efeitos da não-observação das condições dos teoremas de bem-estar. Com relação ao primeiro teorema, vamos analisar o que acontece quando há bens-públicos e externalidades. Com relação ao segundo, vamos admitir a impossibilidade de tributação *lump-sum*. Neste caso, vamos admitir que o governo somente pode redirecionar recursos por meio de tributos. Assim, vamos estudar como ocorre a incidência de impostos na economia, vamos avaliar os custos sociais da utilização de tributos que não são *lump-sum* e como os bens são afetados pela tributação. Por fim, vamos tentar entender o federalismo e a discussão sobre dívida pública.

2 Bens Públicos

I think that it is a relatively good approximation to truth – which is much too complicated to allow anything but approximations – that mathematical ideas originate in empirics. [...] As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality", it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely l'art pour l'art. This need not be bad, if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste. But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much "abstract" inbreeding, a mathematical subject is in danger of degeneration. [...] [W]henever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the reinjection of more or less directly empirical ideas. I am convinced that this was a necessary condition to conserve the freshness and the vitality of the subject and that this will remain equally true in the future.

von Neumann, 1947

2.1 Introdução

Um bem é chamado de bem público puro se o consumo de cada indivíduo de tal bem não levar a nenhuma subtração do consumo de qualquer outro indivíduo[2] (Samuelson, 1954, p. 387)[3]. Esta propriedade é chamada de não-rivalidade no uso. Existem duas áreas importantes da economia em que os bens públicos desempenham um papel importante. A primeira é aquela em que se avalia os gastos em coisas como defesa nacional: o custo de prover um sistema de defesa de míssil é independente do número de pessoas que habitam a área protegida, e é impossível defender alguns mas não todos os habitantes. Este é o exemplo prototípico usado para motivar o papel do governo no fornecimento de tais bens públicos. Historicamente, esse foi o conjunto de problemas que motivaram o interesse em bens públicos. Embora interessante, mais recentemente, uma segunda classe de problemas em que os bens públicos desempenham um papel importante é a economia familiar. Um casal consome conjuntamente muitos bens e compartilha o custo desses bens. A maneira como um casal decide a quantidade de tempo gasto com a criação dos filhos e o quanto cada um deles contribui para essa quantia é fundamental para entender o desenvolvimento infantil, a participação na força de trabalho, o grau de correspondência entre a proporcionalidade e as probabilidades de divórcio etc.

Duas características dos bens públicos:

1. Não rivalidade: os bens privados beneficiam apenas um único usuário, enquanto os bens públicos fornecem benefícios para um número maior de usuários simultaneamente. Se o bem público puder acomodar qualquer número de usuários ele é puro. Neste caso, dada a existência do bem público nessa escala, então o custo marginal de adicionar outro usuário é zero. Se ocorrer congestionamento, é impuro, isto é, o custo marginal de adicionar outro usuário é maior do que

[2] Samuelson se baseia nos trabalhos de Sax, Wicksell, Lindhal, Musgrave e Bowen.
[3] Para detalhes, ver "The Pure Theory of Public Expenditure (1954)".

zero. Mais genericamente, um bem público puro é caracterizado pela não rivalidade. O consumo do bem público não reduz a quantidade disponível para consumo para os outros. Ex: uma rádio estatal.

2. Não exclusivo: é possível excluir indivíduos de consumir o bem público? Exemplo: uma rádio estatal é impossível de excluir, enquanto o ensino sim. É impossível, ou extremamente caro excluir o consumo do bem público. A maioria das análises econômicas se concentra em bens públicos puros.

Nenhum bem público é realmente puro, mas é uma referência útil. Bens que satisfazem as duas condições acima (não rivais no consumo e não excludentes) até certo ponto, mas não totalmente, são chamados de bens públicos impuros. A Tabela 2.1 resume esses pontos.

Tabela 2.1 – *Definindo Bens*

		O bem é rival?	
		Sim	Não
O bem é exclusivo?	Sim	Bem privado (sorvete)	*Club good* (calçada lotada)
	Não	Bem público impuro (*Netflix*)	Bem público puro (defesa)

Um sinal de farol é um exemplo clássico de bem público puro, em que a oferta é não rival e não excludente. As apresentações teatrais e os eventos esportivos não televisados são exemplos interessantes de um bem público local (*club good*), em que a oferta não é rival, mas exclui-se. O mercado não é o único mecanismo pelo qual bens e serviços são fornecidos em uma economia moderna (Coase, 1974); os bens públicos e os bens de clube são caracterizados por serem fornecidos inteiramente por meio de um processo político, uma vez que, por sua própria natureza, não são comercializáveis.

A principal razão pela qual a falha de mercado persiste reflete-se na incapacidade dos cidadãos de agirem cooperativamente e é essa falta de cooperação que exige um papel alocativo para o governo na economia. Um bem público que se torna excludente é um bem de clube (McNutt, 1996). A análise econômica dos clubes iniciada por Buchanan (1965) pode ser aplicada ao fornecimento de bens públicos locais, que vão desde o fornecimento de bens públicos regionais descentralizados (conselhos locais de saúde) a projetos comunitários e esquemas de bairro, como clubes esportivos comunitários e associações residenciais.

Na teoria dos clubes, entretanto, existe um consumo coletivo, mas com um princípio de exclusão, por exemplo, uma taxa de adesão. Pode-se pensar nos bens do clube como bens públicos sem impossibilidade de exclusão. Há economias de escala porque os sócios adicionais reduzem o custo médio do bem do clube. Mas membros adicionais também geram aglomeração, o que, a longo prazo, pode ser considerado a introdução do consumo rival. De fato, os bens do clube têm extremos polares, conforme observado por Mueller (1989, p. 131): “para um bem público puro, a adição de mais um membro ao clube nunca diminui os benefícios da afiliação ao clube”.

Buchanan (1965), que foi um dos primeiros estudiosos a considerar as propriedades de eficiência dos clubes voluntários, derivou as condições econômicas sob as quais uma provisão ótima de um bem público local poderia ser alcançada. Este trabalho inicial esboçou uma justificativa para a análise do clube na explicação de por que os clubes deveriam se organizar. Buchanan e Olson (1965) reconheceram independentemente que os clubes permitem aos membros explorar economias de escala na provisão do bem público e compartilhar o custo de sua provisão. Cada um deles abordou a questão das restrições de associação, com Olson distinguindo entre clubes exclusivos e clubes inclusivos sem restrições de associação.

Da mesma forma, Tiebout (1956) havia abordado muito antes uma questão relacionada a clubes em seu trabalho sobre a mobilidade da população e o tamanho do governo local. Outros estudiosos, nomeadamente Schelling (1969) e McGuire (1974), justificaram a formação de clubes com base no “gosto pela associação”. Desde então, isso foi traduzido na literatura dos clubes como a suposição de homogeneidade (gostos idênticos), uma suposição que levantou a questão política de se os clubes mistos são ou não ideais. Por exemplo, se os clubes mistos não são ideais, então a política de segregação de grupo é ótima. A questão da otimização, no entanto, não está completamente resolvida na literatura do clube.

Exemplo 2.1. *Vamos fazer um exemplo de provisão ótima de dois bens privados: sorvete e brownies. Vamos assumir que o preço de sorvete é* P_S *e o o preço dos brownies é* P_B*. Também iremos assumir que* $P_B = 1$*, isto é, que os brownies serão nosso numerário (aquele bem que tomamos como referência para os preços). Sejam* A *e* G *dois consumidores que demandem esses produtos em diferentes quantidades mas no mesmo mercado. Sabemos de Teoria Microeconômica que* $TMS_{S,B} = \dfrac{UMgS}{UMgB}$ *é o número de brownies que o consumidor deseja abrir mão para consumir uma unidade de sorvete adicional. A condição de otimização para o consumo de bens privados é escrita como* $TMS^A_{S,B} = TMS^G_{S,B} = \dfrac{P_S}{P_B} = P_S$*. O equilíbrio do lado da oferta requer que o custo marginal da produção de sorvete seja igual ao preço, isto é,* $CMgS = P_S$*. Em equilíbrio, portanto,* $TMS^A_{S,B} = TMS^G_{S,B} = CMgS$*.*

Exemplo 2.2. *Suponha que existam apenas duas pessoas que vivem nas margens da Lagoa dos Patos. Ana gosta de esquiar na água e Bruno gosta de tomar sol. Ambas as atividades são seriamente afetadas pelo nível da água no lago. Quando há muita água no lago, é bom praticar esqui aquático, mas a linha de água é tão alta que não há praia para se bronzear. Quando há muito menos água, o banho de sol é bom, mas o lago é raso demais para esquiar na água. Portanto, Ana prefere que o lago tenha muita água, e Bruno prefere que ele tenha muito menos água. Felizmente, é possível aumentar ou diminuir o nível da água sem custos, abrindo uma barragem em uma extremidade do lago ou na outra extremidade. Infelizmente, não está claro em que nível a água deve ser ajustada.*

Para medir a quantidade de água no lago, vamos usar a profundidade da água em um local especificado no lago: deixe x *denotar a profundidade da água (em metros) nesse local. As preferências de Ana e Bruno são descritas pelas seguintes funções de* *utilidade:*

$$u^A(x, y_A) = y_A - (15 - x)^2 \tag{2.1}$$

$$u^B(x, y_B) = y_B - \frac{1}{2}(6 - x)^2 \tag{2.2}$$

em que x indica o nível da água e y_A e y_B são o consumo diário de Ana e Bruno de outros bens, medidos em reais.

Suponha que Ana e Bruno tenham uma renda de R\$ 100 por dia. Observe que as taxas marginais de substituição de Ana e Bruno são

$$TMS^A_{x,y_A} = \frac{\frac{\partial u^A(x,y_A)}{\partial x}}{\frac{\partial u^A(x,y_A)}{\partial y_A}} = \frac{-2(15-x)(-1)}{1} = 30 - 2x \tag{2.3}$$

$$TMS^B_{x,y_B} = \frac{\frac{\partial u^B(x,y_B)}{\partial x}}{\frac{\partial u^B(x,y_B)}{\partial y_B}} = \frac{-\frac{2}{2}(6-x)(-1)}{1} = 6 - x \tag{2.4}$$

O nível de água mais preferido de Ana é $\hat{x}_A = 15$ e o nível mais preferido de Bruno é $\hat{x}_B = 6$. Observe que, se o nível da água estiver acima de $\hat{x}_B$, Bruno estaria disposto a pagar para ter x reduzido, e que Ana estaria igualmente disposta a pagar para reduzir x se estiver acima do nível ideal, $\hat{x}_A$.

O que torna essa situação diferente de tudo o que vimos antes é que a variável x não pode estar em níveis diferentes para pessoas diferentes. Não é como pizza ou cerveja, onde Ana pode consumir uma quantidade e Bruno uma quantidade diferente. Nesse caso, o nível da água pode variar, mas será o mesmo para os dois. É por isso que não usamos os subscritos A e B na variável x: é apenas uma variável, não duas. O nível da água neste exemplo é um bem público.

Vamos tentar determinar quais resultados são eficientes de Pareto. Vamos começar perguntando se um nível de água de $x = 8$ metros é eficiente. Em $x = 8$, *as TMS's de Ana e Bruno são* $TMS^A_{x,y_A} = 14$ *e* $TMS^B_{x,y_B} = -2$*, respectivamente. Ana estaria disposta a pagar R$ 14 para aumentar o nível da água em um metro e Bruno estaria disposta a pagar R$ 2 para diminuí-lo em um metro – ou Bruno estaria disposta a aceitar um pagamento de R$ 2 como compensação pelo aumento do nível da água em um metro. Portanto, se aumentássemos o nível da água em um metro e se Ana compensasse Bruno pagando a ele, digamos, R$ 6, os dois estariam melhor. De fato, você pode calcular que a utilidade de Ana aumentará de* 51 *para* 58 *e que a utilidade de Bruno aumentará de* 98 *para* $101\frac{1}{2}$.

E o novo nível de água de $x = 9$ *metros é eficiente? Temos que* $TMS^A_{x,y_A} = 12$ *e* $TMS^B_{x,y_B} = -3$*, então poderíamos aumentar o nível da água em mais um metro, com Ana pagando outros R$ 6 para compensar Bruno. O pagamento de R$ 6 é menor do que os R$ 12 que Ana estaria disposto a pagar, e mais do que os R$ 3 que Bruno estariam dispostos a aceitar como compensação, então eles estão novamente melhor com o aumento de um metro com a compensação de R$ 6. Você pode calcular que suas utilidades terão aumentado novamente, para* $u^A = 63$ *e* $u^B = 104$.

Agora está ficando claro que, enquanto Ana estiver disposta a pagar mais por um aumento do que Bruno estaria disposto a aceitar como compensação, essa barganha – aumentar o nível da água, com Ana compensando Bruno – melhorará os dois. Em outras palavras, o nível da água não é Pareto eficiente desde que $TMS^A_{x,y_A} > -TMS^B_{x,y_B}$ *– ou seja, desde que* $TMS^A_{x,y_A} + TMS^B_{x,y_B} > 0$*. Quando* $TMS^A_{x,y_A} + TMS^B_{x,y_B} > 0$*, o valor social marginal de um aumento em* x *é positivo, então* x *deve ser aumentado. Da mesma forma, poderíamos mostrar que, se* $TMS^A_{x,y_A} + TMS^B_{x,y_B} < 0$*,* x *deve ser diminuído (porque o valor social marginal de um aumento é negativo, portanto, o valor social marginal de uma diminuição de* x *é positivo).*

Os resultados eficientes de Pareto são, portanto, aqueles que satisfazem a condição de teste $TMS^A_{x,y_A} + TMS^B_{x,y_B} = 0$*. No nosso exemplo, é fácil resolver o nível* *de água eficiente na lagoa:*

$$\begin{aligned} TMS^A_{x,y_A} + TMS^B_{x,y_B} &= 0 \\ (30-2x)+(6-x) &= 0 \\ 36-3x &= 0 \\ x &= 12 \end{aligned} \tag{2.5}$$

O nível eficiente da água é de 12 metros, em que $TMS^A_{x,y_A} = 6$ *e* $TMS^B_{x,y_B} = -6$.

Agora sabemos o nível de água que é Pareto eficiente. Mas qual nível será realmente escolhido? Como sempre, isso depende dos arranjos institucionais usados para escolher o nível da água. Por exemplo, as partes afetadas podem votar no nível que desejam. O resultado dessa instituição pode ser analisado usando a teoria dos jogos, o que não faremos aqui. Vamos supor que Bruno seja dono da lagoa ou, pelo menos, que ele tenha o direito de escolher o nível da água. Qual será o nível da água? Parece que Bruno escolherá o nível que ele mais gosta, ou seja, $x = 6$ *metros.*

Mas já vimos que, em um nível tão baixo de água, Bruno estaria melhor ao permitir um nível mais alto se Ana a compensasse adequadamente. De fato, esperamos que eles cheguem a uma negociação mutuamente agradável, na qual não haja mais ganhos a serem obtidos com o comércio ou a negociação - ou seja, um resultado Pareto eficiente. No nosso exemplo, isso significa que o nível da água será de 12 metros, com Ana pagando a Bruno uma quantia como compensação pelo aumento de 6 para 12 metros.

No caso de bens públicos estaremos considerando na maioria das vezes o conceito de equilíbrio de Nash. Neste caso, cada agente escolhe sua contribuição ótima (do ponto de vista privado, naturalmente) tomando como dada a contribuição dos demais agentes. Em ambientes mais complexos, com dinâmica e/ou incerteza e/ou assimetria de informação, tem-se em geral uma multiplicidade de equilíbrios. Usam-se, então, refinamentos como: equilíbrio perfeito em sub-jogos; equilíbrio bayesiano, equilíbrio sequencial, etc.

2.2 Provisão Ótima de um Bem Público Puro: Caso Discreto

Por simplicidade consideramos uma economia com n consumidores e dois bens: um bem público e um bem privado.

Seja g_i a contribuição feita pelo consumidor i, de tal modo que

$$x_i + g_i = w_i \tag{2.6}$$

$$\sum_{i=1}^{n} g_i = \nu \tag{2.7}$$

Assuma que $u_i(x_i, y)$ é estritamente crescente (monotonicamente) e contínua.

Seja c o custo de produzir o bem público tal que a tecnologia de produção é dada por

$$y = \begin{cases} 1 & \text{se} \quad \sum_{i=1}^{n} g_i \geq c \\ 0 & \text{se} \quad \text{caso contrário} \end{cases} \tag{2.8}$$

Queremos saber sob quais condições oferecer o bem público será Pareto em relação a não ofertar, isto é, existe $(g_1, \ldots, g_n)$ tal que $\sum_{i=1}^{n} g_i \geq c$ e

$$u_i(w_i - g_i, 1) > u_i(w_i, 0) \quad \forall i \tag{2.9}$$

Seja r_i a disposição máxima a pagar (preço de reserva) do indivíduo i, isto é, r_i deve satisfazer

$$u_i(w_i - r_i, 1) = u_i(w_i, 0) \tag{2.10}$$

Se produzir o bem público domina não produzir o bem público, temos que

$$u_i(w_i - g_i, 1) > u_i(w_i, 0) = u_i(w_i - r_i, 1) \quad \text{para todo } i \tag{2.11}$$

Supomos a monotonicidade[4] da função de utilidade. Logo,

$$w_i - g_i > w_i - r_i \tag{2.12}$$

Assim,

$$r_i > g_i \tag{2.13}$$

e, portanto,

[4] Há duas suposições, a saber, monotonicidade e convexidade, que são frequentemente usadas para garantir um bom comportamento das funções de demanda do consumidor. Primeiro, definimos vários tipos de monotonicidade usados na teoria do consumidor. Monotonicidade fraca: se $x \geqq y$, então $x \succeq y$. Ou seja, se as cestas contém as mesmas quantidades de bens, então x é tão boa quanto y. Monotonicidade: se $x > y$, então $x \succ y$. Ou seja, é preferível mais a menos quando um bem é desejável. Monotonicidade forte: se $x \geqq y$ e $x \neq y$, então $x \succ y$. Se comparada a y, x possui pelo menos as mesmas quantidades de todos os bens e uma quantidade maior de, pelo menos, um bem, então $x \succ y$. Outra suposição que é mais fraca que qualquer tipo de monotonicidade ou monotonicidade forte é a de Não-Saciedade Local: dado qualquer x em X e qualquer $\varepsilon > 0$, então há uma cesta $y \in X$ com $|x - y| < \varepsilon$ tal que $y \succ x$. Não-Saciedade: dado qualquer x em X, então há uma cesta $y \in X$ tal que $y \succ x$. A monotonicidade das preferências pode ser interpretada como o desejo dos indivíduos por bens: quanto mais, melhor. A não-saciedade local diz que sempre é possível estar um pouco melhor, mesmo se estiver restrito a apenas pequenas mudanças na cesta. Assim, a não-saciedade local significa que os desejos dos indivíduos são ilimitados. Você deve verificar que a monotonicidade (forte) implica não-saciedade local e a não-saciedade local implica em não-saciedade, mas não vice-versa.

$$\sum_{i=1}^{n} r_i > \sum_{i=1}^{n} g_i \geq c \tag{2.14}$$

Ou seja, a soma da disposição a pagar pelo bem público deve exceder o custo de fornecê-lo. Esta condição é necessária. Na verdade, esta condição também é suficiente.

2.3 Provisão Ótima de um Bem Público Puro: Caso Contínuo

Com base nesse último exemplo vamos formalizar nossos resultados e ver que eles não decorrem de nossas escolhas particulares de bens e indivíduos. Para tanto, sejam as seguintes características do modelo:

- existem n consumidores, indexados por $i = 1, 2, \ldots, n$
- x_i é o consumo do bem privado
- G é o consumo (comum) do bem público
- a preferência do agente i é descrita pela função de utilidade

$$u_i(x_i, G) \tag{2.15}$$

  que é diferenciável (o que nos permite computar a utilidade marginal), crescente em ambos os argumentos (quanto maiores forem os valores de x_i e de G maior será a satisfação do consumidor, em decorrência do princípio de preferências monotônicas), quase-côncava (o que garante que a sua solução ótima é de fato um máximo) e satisfaz as condições de Inada[5] (o que garante a existência de uma solução interna)

[5] As condições de Inada são condições impostas sobre a função de produção (ou função de utilidade) em modelos econômicos, especialmente em modelos de crescimento econômico. Essas condições garantem certos comportamentos desejáveis da função de produção à medida que a quantidade de insumos, como capital ou trabalho, tende a zero ou ao infinito. As condições de Inada têm várias

- w_i é a dotação do bem privado i e $W = \sum_{i=1}^{n} w_i$ é o total de dotações de bens privados; o indivíduo não tem nenhuma dotação inicial de bens públicos

- o bem público pode ser produzido a partir do bem privado de acordo com uma função de produção f, em que $f' > 0$ e $f'' < 0$: isto é, se z é o total de unidades de bens privados que são usados como insumos para produzir o bem público, o nível de bem público produzido será

$$G = f(z) \tag{2.16}$$

Primeiro, fazemos a pergunta normativa sobre qual é o nível ótimo de oferta de um bem público puro. Assumimos que o governo de uma economia totalmente controlada escolha o nível de G e a alocação de bens privados $\vec{x} = (x_1, \dots, x_n)$ para os agentes de acordo com o critério de Pareto. Para resolvermos o problema precisamos de duas definições. Sejam as seguintes definições.

Uma alocação (x, G) é viável se existe um $z \geq 0$ tal que

$$\sum_{i=1}^{n} x_i + z \leq W \tag{2.17}$$

---

implicações importantes em modelos econômicos:

1. Garantia de Convergência em Modelos de Crescimento: As condições de Inada são frequentemente usadas em modelos de crescimento econômico, como o modelo de Solow, para garantir que a economia convergirá para um estado estacionário, onde o capital e o trabalho estão em equilíbrio.
2. Produtividade Marginal Decrescente: As condições refletem a ideia de produtividade marginal decrescente, onde o aumento contínuo de um insumo, mantendo os outros constantes, resulta em incrementos cada vez menores na produção.
3. Evitando Extremos Irrealistas: As condições de Inada evitam que os modelos prevejam resultados extremos, como produtividade marginal constante ou crescente indefinidamente, o que seria irrealista em contextos econômicos.
4. Sustentabilidade do Crescimento: Elas também garantem que, no longo prazo, a economia não possa crescer indefinidamente apenas aumentando o capital ou o trabalho, pois ambos enfrentam rendimentos marginais decrescentes.

$$G \leq f(z) \tag{2.18}$$

Uma alocação viável (x, G) é um ótimo de Pareto se não existe outra alocação viável (x', G') tal que

$$u_i(x_i', G') \geq u_i(x_i, G), \quad \forall i = 1, \ldots, n \tag{2.19}$$

e para algum i

$$u_i(x_i', G') > u_i(x_i, G) \tag{2.20}$$

Ou seja, uma alocação viável (factível) (x, G) é ótima de Pareto se não houver maneira de tornar um agente estritamente melhor sem prejudicar alguém. Com isso em mente, podemos caracterizar o conjunto de alocações ótimas de Pareto.

O problema de otimização é

$$\max_{x,G,z} u_1(x_1, G) \tag{2.21}$$

$$\text{sujeito a} \quad u_i(x_i, G) - \underline{u}_i \geq 0 \quad \text{para } i = 2, 3, \ldots, n \tag{2.22}$$

$$W - \sum_{i=1}^{n} x_i - z \geq 0 \tag{2.23}$$

$$f(z) - G \geq 0 \tag{2.24}$$

$$G \geq 0, z \geq 0, x_i \geq 0 \qquad \forall i = 1, 2, \ldots, n \tag{2.25}$$

Para resolver este problema montamos o lagrangeano, como segue:

$$\begin{aligned} L(x_i, G, z, \lambda, \mu, \alpha, \beta, \delta) = u_1(x_1, G) - \gamma_i \left[-u_i(x_i, G) + \underline{u}_i\right] - \lambda \left[-W + \sum_{i=1}^{n} x_i + z\right] \\ - \mu \left[-f(z) + G\right] - \alpha(-G) - \beta(-z) - \delta(-x_i) \end{aligned} \tag{2.26}$$

As condições de primeira ordem (condições necessárias e que nos dão as soluções) são:

$$\frac{\partial L}{\partial x_i} = 0 \iff \gamma_i \frac{\partial u_i(x_i, G)}{\partial x_i} - \lambda + \delta = 0 \tag{2.27}$$

$$\frac{\partial L}{\partial G} = 0 \iff \sum_{i=1}^{n} \gamma_i \frac{\partial u_i(x_i, G)}{\partial G} - \mu + \alpha = 0 \tag{2.28}$$

$$\frac{\partial L}{\partial z} = 0 \iff -\lambda + \mu f'(z) + \beta = 0 \tag{2.29}$$

$$\gamma_i \left[-u_i(x_i, G) + \underline{u}_i\right] = 0 \tag{2.30}$$

$$\lambda \left[-W + \sum_{i=1}^{n} x_i + z\right] = 0 \tag{2.31}$$

$$\mu \left[-f(z) + G\right] = 0 \tag{2.32}$$

$$\alpha(-G) = 0 \tag{2.33}$$

$$\beta(-z) = 0 \tag{2.34}$$

$$\delta(-x_i) = 0 \tag{2.35}$$

$$\gamma_i \geq 0 \tag{2.36}$$

$$\lambda \geq 0 \tag{2.37}$$

$$\mu \geq 0 \tag{2.38}$$

$$\alpha \geq 0 \tag{2.39}$$

$$\beta \geq 0 \tag{2.40}$$

$$\delta \geq 0 \tag{2.41}$$

$$u_i(x_i, G) - \underline{u}_i \geq 0 \tag{2.42}$$

$$W - \sum_{i=1}^{n} x_i - z \geq 0 \tag{2.43}$$

$$f(z) - G \geq 0 \tag{2.44}$$

$$G \geq 0 \tag{2.45}$$

$$z \geq 0 \tag{2.46}$$

$$x_i \geq 0 \tag{2.47}$$

As condições de Inada que assumimos serem válidas na função de utilidade implicam que as restrições de não-negatividade podem ser ignoradas (equações (2.45), (2.46) e (2.47), e, consequentemente, as equações (2.33), (2.34), (2.35), (2.39), (2.40) e (2.41)). Também sabemos que as condições de Inada garantem que a solução será interior e, portanto, as equações (2.42), (2.43) e (2.44) valem com a igualdade em vez de desigualdade. Disso decorre que γ_i, λ e μ, nas equações (2.36), (2.37) e (2.38), são todos estritamente positivos. Com isso, as condições necessárias e suficientes (suficientes devido à quase-concavidade em u e f) podem ser resumidas em 3 equações (e não em 21):

$$\frac{\partial L}{\partial x_i} = 0 \iff \gamma_i \frac{\partial u_i(x_i, G)}{\partial x_i} - \lambda = 0 \tag{2.48}$$

$$\frac{\partial L}{\partial G} = 0 \iff \sum_{i=1}^{n} \gamma_i \frac{\partial u_i(x_i, G)}{\partial G} - \mu = 0 \tag{2.49}$$

$$\frac{\partial L}{\partial z} = 0 \iff -\lambda + \mu f'(z) = 0 \tag{2.50}$$

em que $\gamma_1 = 1$ por convenção.

Temos da equação (2.48) que:

$$\gamma_i \frac{\partial u_i(x_i, G)}{\partial x_i} = \lambda \tag{2.51}$$

e da equação (2.50) que

$$\lambda = \mu f'(z) \tag{2.52}$$

Igualando (2.51) e (2.52) para eliminar λ e rearranjando, temos:

$$\gamma_i \frac{\partial u_i(x_i, G)}{\partial x_i} = \mu f'(z)$$

$$\frac{\mu}{\gamma_i \frac{\partial u_i(x_i, G)}{\partial x_i}} = \frac{1}{f'(z)}$$

$$\frac{\sum_{i=1}^{n} \gamma_i \dfrac{\partial u_i(x_i, G)}{\partial G}}{\gamma_i \dfrac{\partial u_i(x_i, G)}{\partial x_i}} = \frac{1}{f'(z)} \qquad \left[\text{usando } \mu = \sum_{i=1}^{n} \gamma_i \frac{\partial u_i(x_i, G)}{\partial G} \text{ [ver (2.49)]}\right]$$

$$\sum_{i=1}^{n} \frac{\dfrac{\partial u_i(x_i^*, G^*)}{\partial G^*}}{\dfrac{\partial u_i(x_i^*, G^*)}{\partial x_i^*}} = \frac{1}{f'(z^*)} \tag{2.53}$$

A equação (2.53) é referida como condição de Samuelson, condição de Lindahl-Samuelson, ou às vezes até mesmo condição de Bowen-Lindahl-Samuelson.

O lado esquerdo da equação acima é a soma das taxas marginais de substituição dos n agentes. Para ver isto, note que a partir da curva de indiferença do agente i, o termo $\dfrac{\dfrac{\partial u_i(x_i, G)}{\partial G}}{\dfrac{\partial u_i(x_i, G)}{\partial x_i}}$ denota a quantidade do bem privado que o agente está disposto a abrir mão para ter um pequeno aumento no nível de consumo do bem público. O lado direito da equação acima é a quantidade do bem privado necessária para produzir uma unidade adicional de bem público (também conhecida como taxa marginal de transformação). Com base na regra acima, a condição de Samuelson diz o seguinte: qualquer alocação ótima é tal que a soma da quantidade de bens privados que os n consumidores estariam dispostos a desistir para ter uma unidade adicional de bem público deve ser igual à quantidade do bem privado que é realmente necessária para produzir a unidade adicional do bem público.

Se houver mais de uma mercadoria privada, digamos, k bens privados, e o bem público é produzido de acordo com

$$f(z_1, \dots, z_k) \tag{2.54}$$

então a condição correspondente de Samuelson para o nível ótimo de bens públicos é dada por

$$\sum_{i=1}^{n} \frac{\dfrac{\partial u_i(x_{ij}^*, G^*)}{\partial G^*}}{\dfrac{\partial u_i(x_{ij}^*, G^*)}{\partial x_{ij}^*}} = \frac{1}{\dfrac{\partial f(z_1^*, \ldots, z_k^*)}{\partial z_j^*}}, \quad \forall j = 1, \ldots, k \tag{2.55}$$

Uma ilustração gráfica da condição de Samuelson para o caso onde há dois indivíduos e dois bens é dada na Figura 2.1. Na Figura, a parte superior mostra as curvas de indiferença para o agente 1 e a restrição de produção AB. Suponha que nós fixamos o agente 1 na curva de indiferença $\underline{u}_1$; então as possibilidades para o agente 2 são mostradas na parte inferior da Figura por CD (que é a diferença entre AB e $\underline{u}_1$). Claramente, a eficiência de Pareto requer que a taxa marginal de substituição do agente 2 seja igual à inclinação da curva CD (isto é, no ponto E). Mas esta é justamente a diferença entre a taxa marginal de transformação (a inclinação da fronteira de possibilidades de produção) e a taxa marginal de substituição do agente 1 (a inclinação de sua curva de indiferença).

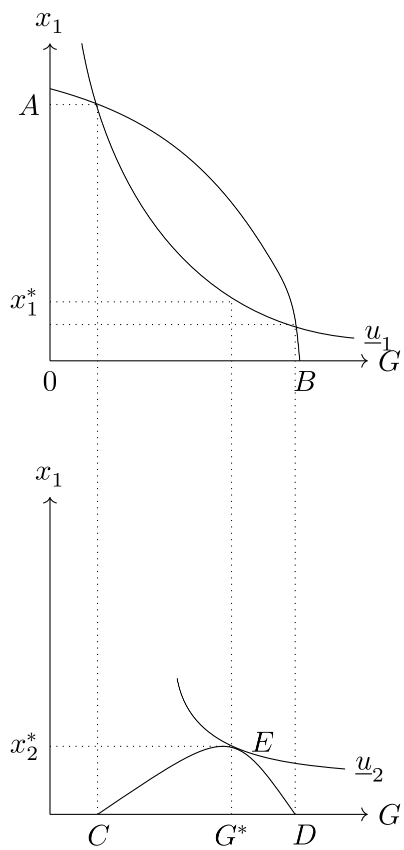
Assim nós temos

$$TMgS_2 = TMgT - TMgS_1 \tag{2.56}$$

Como fazer a implementação da alocação ótima? Se o governo é capaz de cobrar impostos de montante fixo (*lump-sum*[6]) tanto para financiar as despesas quanto para redistribuir a renda, então a alocação ótima acima pode ser alcançada. Se os impostos de montante fixo não forem viáveis, o governo precisa usar distorções como impostos. Por exemplo, o imposto sobre a renda do trabalho pode ser utilizado para financiar

[6] O imposto *lump sum*, ou de soma fixa, corresponde a um valor fixo que não depende da quantidade produzida ou vendida. Este imposto pode ser considerado como regressivo, pois é maior para aqueles que produzem e vendem menos. O contraste com um imposto por unidade, cobrado a cada unidade de produção produzida ou consumida, é que o imposto aumenta de tamanho à medida que a produção ou o consumo aumenta. Um imposto de montante fixo aumenta o custo fixo médio das empresas e, portanto, o custo total médio, mas não afeta o custo marginal ou o custo variável médio.

Figura 2.1 – *Provisão Ótima de Bens Públicos*



os bens públicos.

Essa solução funciona em uma situação "totalmente controlada", na qual o governo tem informações perfeitas sobre as preferências e, em seguida, pode definir G de maneira ideal. A fórmula acima não leva em conta quaisquer impostos distorcivos necessários para obter fundos para financiar os bens públicos. A exclusibilidade não desempenha nenhum papel na análise. A possibilidade de exclusão é relevante apenas para determinar os mecanismos de provisão possíveis. Nesse sentido, a análise de Samuelson pode ser considerada como uma análise de *first-best*[7]. A pergunta prática é: como a provisão ótima de um bem público pode ser descentralizada, dadas as ferramentas políticas disponíveis e respeitando as restrições de informação? Vamos tentar responder essa pergunta na próxima seção.

[7] Em economia, as análises *first-best* (de primeira melhor escolha) e *second-best* (de segunda melhor escolha) referem-se à avaliação das políticas e dos resultados econômicos em cenários ideais e não ideais, respectivamente. Esses conceitos são fundamentais na teoria do bem-estar e na formulação de políticas econômicas. A análise de *first-best* ocorre em um cenário ideal onde todas as condições necessárias para a eficiência econômica são atendidas. Nesse contexto, as alocações de recursos são consideradas eficientes de acordo com o critério de Pareto, ou seja, não é possível melhorar a situação de alguém sem piorar a de outra pessoa. Características do cenário *first-best*:

- Mercados Competitivos Perfeitos: Todos os mercados funcionam sem falhas, como externalidades, poder de monopólio ou informação assimétrica.
- Alocação Eficiente: Os recursos são alocados de maneira que maximiza o bem-estar social.
- Intervenção Mínima: O governo não precisa intervir na economia, exceto para garantir que os mercados permaneçam competitivos e sem falhas.

A análise de *second-best* se aplica quando uma ou mais das condições necessárias para o cenário de *first-best* não são satisfeitas. Quando é impossível alcançar a primeira melhor solução devido à presença de falhas de mercado, restrições institucionais ou políticas, a melhor alternativa possível é denominada solução de *second-best*. Características do cenário *second-best*:

- Falhas de Mercado: Existem imperfeições no mercado, como externalidades, monopólios, assimetrias de informação ou mercados incompletos.
- Intervenção Necessária: O governo ou outra autoridade reguladora pode precisar intervir para melhorar a alocação de recursos, mas essa intervenção pode não restaurar completamente a eficiência de Pareto.
- Compromissos Inevitáveis: A solução de *second-best* pode envolver compromissos onde corrigir uma falha em uma parte do mercado pode exacerbar problemas em outra, resultando em uma situação que não é ideal, mas que é melhor do que não intervir.

As dificuldades da provisão pública de ordem conceituais são: i) *crowding out*[8]

[8] O termo *crowding out* refere-se a um fenômeno econômico em que o aumento do gasto público resulta na redução do investimento privado. Esse conceito é frequentemente discutido em macroeconomia, particularmente no contexto da política fiscal e de seus efeitos sobre a economia. O *crowding out* ocorre quando o governo aumenta seus gastos, especialmente em um cenário onde esses gastos são financiados por meio de empréstimos. Esse aumento na demanda por empréstimos por parte do governo pode levar a um aumento nas taxas de juros. As taxas de juros mais altas tornam o custo do financiamento mais caro para as empresas e os indivíduos, o que pode levar à redução do investimento privado. Mecanismo do *crowding out*:

- Gastos Públicos Aumentados: O governo decide aumentar os gastos para estimular a economia, financiar grandes projetos de infraestrutura ou responder a uma crise.
- Financiamento via Empréstimos: Para cobrir esses gastos, o governo pode precisar tomar empréstimos, o que aumenta a demanda por crédito no mercado financeiro.
- Aumento das Taxas de Juros: Com a maior demanda por crédito, as taxas de juros tendem a subir. Isso ocorre porque os credores exigem uma compensação maior pelo aumento da demanda de fundos.
- Redução do Investimento Privado: As taxas de juros mais altas tornam mais caro para as empresas financiarem novos projetos de investimento e para os consumidores financiarem compras de bens duráveis, como casas e carros. Como resultado, o investimento privado diminui, o que pode contrabalançar o efeito estimulante dos gastos públicos.

Tipos de *crowding out*:

1. *crowding out* Completo: Isso ocorre quando o aumento dos gastos do governo é totalmente compensado por uma queda equivalente no investimento privado. Nesse caso, o efeito líquido sobre a demanda agregada seria nulo.
2. *crowding out* Parcial: Aqui, o aumento dos gastos públicos é parcialmente compensado por uma redução no investimento privado. O efeito líquido sobre a demanda agregada ainda seria positivo, mas menor do que o pretendido pelo governo.
3. *crowding out* Indireto: Além do efeito direto via aumento das taxas de juros, o *crowding out* pode ocorrer de maneira indireta, por meio de expectativas de inflação futura ou aumento da dívida pública, o que pode reduzir a confiança dos investidores e consumidores.

O *crowding out* é mais provável de ocorrer em certas condições econômicas:

1. Economia em Pleno Emprego: Se a economia já está operando em pleno emprego, o aumento do gasto público pode competir diretamente com o setor privado pelos recursos disponíveis, levando a um *crowding out* mais significativo.
2. Mercado de Crédito Limitado: Se o mercado de crédito é restrito ou os fundos disponíveis para empréstimos são limitados, o aumento da demanda por parte do governo pode rapidamente aumentar as taxas de juros.

Economistas keynesianos e neoclássicos frequentemente debatem a magnitude e a importância do *crowding out*. Keynesianos argumentam que, especialmente em uma economia com desemprego elevado e baixo uso de capacidade, o *crowding out* é menos relevante, pois o aumento do gasto público pode levar ao crescimento econômico que, por sua vez, pode estimular o investimento privado. Já os neoclássicos tendem a enfatizar os riscos do *crowding out*, sugerindo que o aumento dos gastos do governo pode prejudicar o crescimento de longo prazo ao reduzir o investimento privado.

da contribuição privada; ii) revelação das preferências; e iii) financiamento. A dificuldade prática é a análise de custos e benefícios.

Exemplo 2.3. *Considere três consumidores* $(i = 1, 2, 3)$ *que derivam utilidade de seus consumos de um bem privado e de um bem público. Suas funções utilidade são da forma* $u_i = x_i G$*, em que* x_i *é o consumo do bem privado pelo consumidor* i *e* G *é a quantidade total do bem público consumida pelos três indivíduos. O custo unitário do bem privado é R\$* $1,00$ *e o custo unitário do bem público é R\$* $10,00$*. Os níveis de renda individuais em R\$ são* $\omega_1 = 30$*,* $\omega_2 = 50$ *e* $\omega_3 = 20$*. Qual a quantidade eficiente de bem público a ser consumida?*

Calculando a taxa marginal de substituição para o consumidor i*:*

$$TMS^i_{G,x_i} = \frac{\partial u_i(x_i, G)/\partial G}{\partial u_i(x_i, G)/\partial x_i} = \frac{x_i}{G}, \quad i = 1, 2, 3 \tag{2.57}$$

Assumindo produção competitiva, temos que a taxa marginal de transformação é igual a razão de preços:

$$TMT_{G,x_i} = \frac{p_G}{p_{x_i}} = 10 \tag{2.58}$$

Logo, da condição BLS:

$$\begin{aligned} \sum_{i=1}^{3} TMS^i_{G,x_i} &= TMT_{G,x_i} \\ \frac{x_1}{G} + \frac{x_2}{G} + \frac{x_3}{G} &= 10 \\ x_1 + x_2 + x_3 &= 10G \end{aligned} \tag{2.59}$$

Pela Lei de Walras[9]*, o valor do excesso de demanda de todos os bens é igual a zero:*

$$
\begin{aligned}
p_x(x_1 + x_2 + x_3) + p_G G - (\omega_1 + \omega_2 + \omega_3) &= 0 \\
x_1 + x_2 + x_3 + 10G &= 100 \\
10G + 10G &= 100 \\
G &= 5 \qquad (2.60)
\end{aligned}
$$

Exemplo 2.4. *A e B estão pensando em comprar um sofá. A função utilidade de A é dada por*

[9] A Lei de Walras é um princípio fundamental na teoria econômica que surgiu do trabalho do economista francês Léon Walras, um dos fundadores da teoria do equilíbrio geral. Esta lei pode ser expressa de forma simples: em um mercado competitivo com múltiplos bens, se todos os mercados, exceto um, estão em equilíbrio (ou seja, a oferta iguala a demanda), então o mercado restante também estará em equilíbrio. Aqui estão alguns pontos-chave sobre a Lei de Walras:

- Equilíbrio Geral: A Lei de Walras faz parte da teoria do equilíbrio geral, que examina como todos os mercados de uma economia interagem e chegam a um ponto de equilíbrio simultâneo.
- Condicionalidade: A lei é condicional no sentido de que se todos os mercados, exceto um, estão em equilíbrio, o mercado restante será automaticamente equilibrado. Em outras palavras, não é necessário analisar todos os mercados individualmente para garantir o equilíbrio geral; é suficiente verificar a condição de equilíbrio para todos, exceto um mercado.
- Ajuste de Preços: A Lei de Walras sugere que, em um sistema de mercados livres e competitivos, os preços se ajustarão de maneira que a oferta e a demanda sejam iguais em todos os mercados. Se houver excesso de oferta ou demanda em um mercado, o ajuste dos preços permitirá que esses desequilíbrios sejam eliminados.
- Implicações: A Lei de Walras é importante porque fornece uma base teórica para a ideia de que a concorrência e o ajuste dos preços podem levar a um equilíbrio econômico onde todos os mercados são equilibrados simultaneamente.
- Críticas e Limitações: Embora a Lei de Walras seja uma peça central da teoria econômica, ela também tem suas limitações. Por exemplo, assume que os mercados são perfeitamente competitivos e que todos os agentes econômicos têm informação perfeita, o que não é sempre o caso no mundo real.

Em resumo, a Lei de Walras é um conceito central na teoria econômica que ajuda a entender como os mercados interagem para alcançar o equilíbrio geral, sob a suposição de que todos os mercados, exceto um, estão em equilíbrio.

$$u_A(s, m_A) = (1 + s)m_A \tag{2.61}$$

e a função de utilidade de B é

$$u_B(s, m_B) = (2 + s)m_B \tag{2.62}$$

em que $s = 0$ *se eles não comprarem o sofá e* $s = 1$ *se comprarem. Além disso,* m_A *e* m_B *correspondem à quantidade de dinheiro que eles dispõem, respectivamente, para gastar no bem privado. Cada um deles tem R\$* $100,00$ *para gastar. Qual a maior quantia que eles poderiam pagar pelo sofá que deixasse ambos em situação melhor do que sem ter o sofá?*

Se eles não compram o sofá $(s = 0, m_A = m_B = 100)$, *suas utilidades serão:*

$$u_A(0, 100) = 100 \tag{2.63}$$

$$u_B(0, 100) = 200 \tag{2.64}$$

Se eles compram o sofá pagando p_A *e* p_B *respectivamente* $(s = 1, m_A = 100 - p_A, m_B = 100 - p_B)$, *suas utilidades serão*

$$u_A(1, 100 - p_A) = 2(100 - p_A) \tag{2.65}$$

$$u_B(1, 100 - p_B) = 3(100 - p_B) \tag{2.66}$$

Logo, A estaria melhor com o sofá se pagasse no máximo p_A^* *tal que*

$$u_A(0, 100) = u_A(1, 100 - p_A^*) \quad \Longrightarrow \quad p_A^* = 50 \tag{2.67}$$

E B estaria melhor com o sofá se pagasse no máximo p_B^* tal que

$$u_B(0, 100) = u_A(1, 100 - p_B^*) \quad \Longrightarrow \quad p_B^* = \frac{100}{3} \tag{2.68}$$

Dessa forma, o máximo que A e B podem pagar pelo sofá conjuntamente para que estejam melhor com ele é

$$p_A^* + p_B^* = \frac{250}{3} \tag{2.69}$$

2.4 Pode a Alocação Ótima ser Descentralizada?

As alocações ótimas caracterizadas pela condição de Samuelson podem ser descentralizadas (sem a participação do governo)? Imagine que existam mercados competitivos para os bens privados e públicos. Vamos assumir que o bem privado é o numerário (ou seja, o preço do bem privado é normalizado para 1). Seja p o preço do bem público (em termos do bem privado). Seja g_i a quantidade de bem público comprada pelo agente i. Sem perda de generalidade, supomos que exista uma única firma maximizadora de lucro tomadora de preços que opera no mercado. Faremos a seguinte suposição: supomos que todos os agentes são tomadores de preço (ou seja, sua escolha não afeta o nível de preços), mas levam em conta que a compra deles pode afetar o nível agregado de bens públicos.

Dado o consumo de bens públicos por outros agentes, que vamos denotar por $\overline{g}_{-i} = (g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$, a melhor resposta do agente i, dado o preço p, é definido como o conjunto de alocações ótimas $\beta_i(\overline{g}_{-i}, p)$. Dessa forma,

$$\beta_i(\overline{g}_{-i}, p) = \arg\max_{\{g_i\}} u_i \left(w_i - pg_i, g_i + \sum_{j \neq i} g_j \right) \tag{2.70}$$

$$\text{sujeito a} \quad g_i \geq 0 \tag{2.71}$$

$$w_i - pg_i \geq 0 \tag{2.72}$$

Para descobrirmos qual é o nível ótimo de bens públicos que deve ser fornecido escrevemos o lagrangeano. Assim obtemos:

$$L(g_i, \lambda, \mu) = u_i \left(w_i - pg_i, g_i + \sum_{j \neq i} g_j \right) - \lambda(-g_i) - \mu(-w_i + pg_i) \tag{2.73}$$

As condições de primeira ordem (condições necessárias e que nos dão as soluções) são dadas por:

$$\frac{\partial L}{\partial g_i} = 0 \quad \Longleftrightarrow \quad -\frac{\partial u_i(x_i, G)}{\partial x_i} p + \frac{\partial u_i(x_i, G)}{\partial G} + \lambda - \mu p = 0 \tag{2.74}$$

$$\lambda g_i = 0 \tag{2.75}$$

$$\mu(w_i - pg_i) = 0 \tag{2.76}$$

$$\lambda \geq 0 \tag{2.77}$$

$$\mu \geq 0 \tag{2.78}$$

$$g_i \geq 0 \tag{2.79}$$

$$w_i - pg_i \geq 0 \tag{2.80}$$

Assumindo que u_i é estritamente quase-côncava, existe uma solução única para o problema de maximização do agente, dado $\overline{g}_{-i}$ e p, que é caracterizada por

$$-\frac{\partial u_i(x_i, G)}{\partial x_i} p + \frac{\partial u_i(x_i, G)}{\partial G} + \lambda - \mu p = 0 \tag{2.81}$$

$$\lambda g_i = 0 \tag{2.82}$$

$$\mu(w_i - pg_i) = 0 \tag{2.83}$$

Como a função de utilidade satisfaz as condições de Inada, isto é, existe solução interior para o problema, temos que $\mu = 0$. Portanto:

$$p \geq \frac{\frac{\partial u_i(x_i, G)}{\partial G}}{\frac{\partial u_i(x_i, G)}{\partial x_i}} \tag{2.84}$$

O produtor do bem público maximiza os lucros, dado o preço p, isto é,

$$\max_{z \geq 0} \pi = pf(z) - z \tag{2.85}$$

cuja condição ótima implica:

$$\frac{\partial \pi}{\partial z} = 0 \quad \Longleftrightarrow \quad p^* = \frac{1}{f'(z^*)} \tag{2.86}$$

Um equilíbrio competitivo consiste de p^* e $G^* = (g_1^*, \dots, g_n^*)$ de tal modo que:

1. para cada i, dado p^* e $\overline{g}^*_{-i} = (g_1^*, \dots, g_{i-1}^*, g_{i+1}^*, \dots, g_n^*)$,

$$g_i^* \in \beta_i\left(\overline{g}^*_{-i}, p^*\right) \tag{2.87}$$

2. a firma é otimizadora, isto é,

$$p^* = \frac{1}{f'\left[f^{-1}\left(\sum_{i=1}^{n} g_i^*\right)\right]} \tag{2.88}$$

Por (2.16), $G \equiv \sum_{i=1}^{n} g_i = f(z)$. Logo, $z = f^{-1}\left(\sum_{i=1}^{n} g_i\right)$.

Ou seja, um equilíbrio competitivo constitui-se de uma alocação e um vetor de preços que satisfazem algumas condições. A primeira condição diz que dado o vetor de preço a cesta de escolha referente ao agente i é uma solução para o seu problema do consumidor. A outra condição simplesmente diz que o vetor de preços resolve o problema das firmas. E, por fim, as alocações são factíveis.

Utilizando novamente da validade das condições de Inada, temos que para algum $i \in \{1, \dots, n\}$,

$$p = \frac{\frac{\partial u_i(x_i, G)}{\partial G}}{\frac{\partial u_i(x_i, G)}{\partial x_i}} \tag{2.89}$$

que, conjuntamente com as condições de otimização das firmas, resulta, para algum i, em

$$\frac{1}{f'(z)} = \frac{\frac{\partial u_i(x_i, G)}{\partial G}}{\frac{\partial u_i(x_i, G)}{\partial x_i}} \tag{2.90}$$

No equilíbrio competitivo, chegamos a:

$$\sum_{i=1}^{n} \frac{\frac{\partial u_i(x_i^*, G^*)}{\partial G^*}}{\frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*}} > \frac{1}{f'(z^*)} \tag{2.91}$$

Portanto, há sub-provisão do bem público em relação ao nível prescrito pela condição de Samuelson. A intuição é a seguinte: cada agente ao decidir quanto do bem público irá comprar, não considera o benefício para outros agentes da produção que ele comprou. Isso é verdade para cada agente e, consequentemente, como um grupo, os agentes compram menos do que a quantidade desejável para a otimização de Pareto.

Exemplo 2.5. *Suponha que* $u_i(x_, G) = \delta \ln G + \ln x_i$, $w_i = \dfrac{W}{n}$ *e* $f(z) = z$. *Encontre a alocação Pareto ótima e a alocação de equilíbrio competitivo.*

O lagrangeano desse problema é dado por

$$\begin{aligned} L(x_i, G, z, \lambda, \mu, \alpha, \beta, \delta) = \delta \ln G + \ln x_1 - \gamma_i \left[-\delta \ln G - \ln x_i + \underline{u}_i \right] \\ - \lambda \left[-W + \sum_{i=1}^{n} x_i + z \right] - \mu \left[-f(z) + G \right] \\ - \alpha(-G) - \beta(-z) - \delta(-x_i) \end{aligned} \tag{2.92}$$

Sabemos que:

$$\gamma_i \frac{1}{x_i} = \lambda \tag{2.93}$$

$$\sum_{i=1}^{n} \gamma_i \frac{\delta}{G} - \mu = 0 \tag{2.94}$$

$$-\lambda + \mu f'(z) = 0 \tag{2.95}$$

Encontramos que $\gamma_i = \lambda x_i$ *e que* $\mu = \lambda$. *Assim,*

$$\begin{aligned} \sum_{i=1}^{n} \lambda x_i \frac{\delta}{G} - \lambda = 0 \\ \lambda \frac{\delta}{G} \sum_{i=1}^{n} x_i - \lambda = 0 \\ \lambda \left(\frac{\delta}{G} \sum_{i=1}^{n} x_i - 1 \right) = 0 \\ \frac{\delta}{G} \sum_{i=1}^{n} x_i = 1 \\ \sum_{i=1}^{n} x_i = \frac{G}{\delta} \end{aligned} \tag{2.96}$$

Usando o fato de que $W - \sum_{i=1}^{n} x_i - G = 0$*, encontramos:*

$$\begin{aligned}
W - \sum_{i=1}^{n} x_i - G &= 0 \\
W - \frac{G}{\delta} - G &= 0 \\
G &= \frac{\delta}{1+\delta} W
\end{aligned} \tag{2.97}$$

Lembre-se que $\sum_{i=1}^{n} x_i = \frac{G}{\delta}$*. Portanto,*

$$\begin{aligned}
\sum_{i=1}^{n} x_i &= \frac{G}{\delta} \\
\sum_{i=1}^{n} x_i &= \frac{1}{\delta} \frac{\delta}{1+\delta} W \\
n\overline{x}_i &= \frac{1}{\delta} \frac{\delta}{1+\delta} W \\
\overline{x}_i &= \frac{W}{n(1+\delta)}
\end{aligned} \tag{2.98}$$

Observe que podemos obter o mesmo resultado acima usando a condição BLS:

$$\begin{aligned}
\sum_{i}^{n} TMS_{G,x_i} &= \frac{1}{f'(z)} \\
\sum_{i}^{n} \frac{\frac{\delta}{G}}{\frac{1}{x_i}} &= \frac{1}{1} \\
\frac{\delta}{G} \sum_{i}^{n} x_i &= 1 \\
\sum_{i}^{n} x_i &= \frac{G}{\delta}
\end{aligned} \tag{2.99}$$

Usando a Lei de Walras, temos que:

$$\begin{aligned}
p_x(x_1 + x_2 + ... + x_n) + p_G G - W = 0 \\
\sum_i^n x_i + G - W = 0 \\
\frac{G}{\delta} + G - W = 0 \\
G = \frac{\delta}{1+\delta} W
\end{aligned} \tag{2.100}$$

No caso da alocação de equilíbrio competitivo, sabemos que:

$$-\frac{1}{x_i} p + \frac{\delta}{G} + \lambda - \mu p = 0 \tag{2.101}$$

$$\lambda G = 0 \tag{2.102}$$

$$\mu(w_i - pG) = 0 \tag{2.103}$$

Como as condições de Inada são válidas, $\mu = 0$ e $\lambda = 0$. Assim,

$$\begin{aligned}
-\frac{1}{x_i} p + \frac{\delta}{G} = 0 \\
x_i = \frac{G}{\delta} \quad [p = f'(z) \longrightarrow p = 1]
\end{aligned} \tag{2.104}$$

Logo,

$$\begin{aligned}
W - \sum_{i=1}^n x_i - G = 0 \\
W - \sum_{i=1}^n \frac{G}{\delta} - G = 0 \\
G = \frac{\delta}{n+\delta} W
\end{aligned} \tag{2.105}$$

e

$$\begin{aligned} x_i &= \frac{G}{\delta} \\ &= \frac{W}{n+\delta} \end{aligned} \tag{2.106}$$

O aumento de n leva a uma redução de G e, portanto, a sub-provisão do bem público. Perceba as diferenças entre o ótimo de Pareto e a alocação de equilíbrio competitivo.

Nas próximas seções vamos tentar responder a seguinte pergunta: é possível implementar alocações eficientes na presença de bens públicos? Para tentar responder a essa pergunta vamos estudar quais os mecanismos que garantem que isso seja possível. Eles incluem: o mecanismo de alocação e de preços de Lindhal, os mecanismos de votação e o mecanismo de revelação de preferências de Vickrey-Clark-Groves.

2.5 Equilíbrio de Lindhal

Enquanto o equilíbrio competitivo com um preço fixo do bem público produzirá uma alocação ineficiente, existe uma instituição de mercado muito estudada que, em princípio, alcançaria eficiência. A ideia é pensar na quantia comprada por cada agente como uma mercadoria distinta e fazer com que cada agente enfrente um preço personalizado p_i, de forma que todos os agentes concordem com o nível do bem público. Seja $s_i \in [0,1]$, com $\sum_{i=1}^{n} s_i = 1$, o *share* do lucro da firma que o indivíduo obtém.

Um equilíbrio de Lindhal é um vetor de preços $p^* = (p_1^*, \dots, p_n^*)$ e uma alocação $(x_1^*, \dots, x_n^*, G^*)$ tal que:

- A firma maximiza lucros, isto é,

$$G^* = \arg\max_{G \geq 0} \left\{ \left(\sum_{i=1}^{n} p_i^* \right) G^* - f^{-1}(G^*) \right\} \tag{2.107}$$

- Cada consumidor maximiza utilidade, isto é,

$$(x_i^*, G^*) = \arg\max_{x_i, G} u_i(x_i^*, G^*) \tag{2.108}$$

$$\text{sujeito a } w_i + s_i \left(\sum_i p_i^* G^* - f^{-1}(G^*) \right) - x_i^* - p_i^* G^* \geq 0 \tag{2.109}$$

- A condição de equilíbrio do mercado requer que

$$\sum_{i=1}^{n} x_i^* + f^{-1}(G^*) \leq \sum_{i=1}^{n} w_i \tag{2.110}$$

O equilíbrio de Lindahl é um equilíbrio competitivo numa economia fictícia em que o espaço de bens foi expandido para $(n+1)$ bens, bens privados e n bens públicos personalizados. Os n bens são produzidos "em conjunto", de modo que devemos encontrar um vetor de preços para o qual todos os agentes exijam quantidades iguais do bem público. Nós mostramos agora que um equilíbrio de Lindahl é de fato Pareto ótimo.

Para ver isso, montamos o lagrangeano para o problema da firma:

$$L(G) = \left(\sum_{i=1}^{n} p_i \right) G - f^{-1}(G) \tag{2.111}$$

cuja condição de primeira ordem para a maximização do lucro da empresa[10] resulta em

$$\frac{\partial L}{\partial G} = 0 \quad \Longleftrightarrow \quad \sum_{i=1}^{n} p_i^* = \frac{1}{f'\left(f^{-1}\left(G^*\right)\right)} \tag{2.112}$$

O problema do consumidor é

$$L(x_i, G, \lambda) = u_i(x_i, G) - \lambda \left[w_i + s_i \left(\sum_i p_i G - f^{-1}(G) \right) - x_i - p_i G \right] \tag{2.113}$$

[10] Aqui usamos o teorema da função inversa.

cujas condições de primeira ordem para a maximização da utilidade do indivíduo i resultam em

$$\frac{\partial L}{\partial x_i} = 0 \iff \frac{\partial u_i(x_i, G)}{\partial x_i} + \lambda = 0 \tag{2.114}$$

$$\frac{\partial L}{\partial G} = 0 \iff \frac{\partial u_i(x_i, G)}{\partial G} - \lambda \left[s_i \sum_i p_i - s_i \left(\frac{1}{f'\left(f^{-1}(G)\right)} \right) - p_i \right] = 0 \tag{2.115}$$

Substituindo a primeira condição de primeira ordem, usando o fato de que $\lambda = -\dfrac{\partial u_i(x_i, G)}{\partial x_i}$, na segunda condição de primeira ordem, temos:

$$\begin{aligned}
\frac{\partial u_i(x_i, G)}{\partial G} + \frac{\partial u_i(x_i, G)}{\partial x_i} \left[s_i \left(\frac{1}{f'(f^{-1}(G^*))} \right) - s_i \left(\frac{1}{f'\left(f^{-1}(G)\right)} \right) - p_i \right] &= 0 \\
\frac{\partial u_i(x_i^*, G^*)}{\partial G^*} - \frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*} p_i &= 0 \\
\frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*} p_i^* &= \frac{\partial u_i(x_i^*, G^*)}{\partial G^*} \\
\sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*} p_i^* &= \sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial G^*} \\
\sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*} \sum_{i=1}^{n} p_i^* &= \sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial G^*} \\
\sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*} \frac{1}{f'(f^{-1}(G^*))} &= \sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial G^*}, \left[\text{por } (2.112) \ \sum_{i=1}^{n} p_i^* = \frac{1}{f'(f^{-1}(G^*))} \right] \\
\frac{1}{f'(f^{-1}(G^*))} \sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*} &= \sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial G^*} \\
\frac{\sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial G^*}}{\sum_{i=1}^{n} \frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*}} &= \frac{1}{f'(f^{-1}(G^*))} \\
\sum_{i=1}^{n} \frac{\frac{\partial u_i(x_i^*, G^*)}{\partial G^*}}{\frac{\partial u_i(x_i^*, G^*)}{\partial x_i^*}} &= \frac{1}{f'\left(f^{-1}\left(G^*\right)\right)}
\end{aligned} \tag{2.116}$$

que satisfaz a condição de Samuelson. Além disso, todos os conjuntos de orçamentos dos agentes devem manter a igualdade, o que significa que o mercado se ajusta. Portanto, o equilíbrio de Lindahl é eficiente.

O equilíbrio de Lindahl é mais uma prescrição normativa para a alocação de bens públicos do que uma descrição positiva do mecanismo de mercado. A razão é simples: pela definição do preço personalizado no equilíbrio de Lindahl, um agente rapidamente se inclinará para que ele não se comporte de maneira competitiva (uma suposição que sempre foi justificada pela existência de um grande número de participantes do mercado). Ele terá incentivo para declarar erroneamente seu desejo pelo bem público. Ao contrário do caso de bens privados, onde o incentivo para revelar falsas funções de demanda diminui com o número de agentes, um aumento no número de agentes no caso de bem público apenas agrega o problema.

Duas restrições práticas que limitam o uso do preço de Lindahl:

- Necessidade de excluir um consumidor do uso do bem público (não é possível trabalhar com um bem público não excludente);

- Cada agente tem que enfrentar um preço personalizado p_i. O problema disso é que é necessário conhecer as preferências individuais para obter esses preços. Os agentes não têm incentivos para revelar suas preferências. De fato, cada agente tem interesse em fingir que tem pouco gosto pelo bem público.

Exemplo 2.6. *Vamos supor que tenhamos um bem privado, x_i, e um bem público, y. Imagine que o objetivo seja implementar o esquema de Lindahl. Para tanto, vamos assumir que temos a seguinte função de utilidade:*

$$u(x_i, y) = x_i^{\alpha_i} y^{(1-\alpha_i)}, \quad 0 < \alpha_i < 1 \tag{2.117}$$

e que

$$y = \frac{f^{-1}(G)}{p} = \frac{\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} p_i} \tag{2.118}$$

A restrição orçamentária é dada por

$$x_i + p_i y = w_i \tag{2.119}$$

A função demanda pelo bem x_i *é dada pela condição:*

$$\begin{aligned}
TMS_{x_i,y} &= \frac{p_{x_i}}{p_i} \\
\frac{\alpha_i x_i^{\alpha_i - 1} y^{1-\alpha_i}}{(1-\alpha_i) x_i^{\alpha_i} y^{-\alpha_i}} &= \frac{1}{p_i} \\
\frac{\alpha_i}{(1-\alpha_i)} \frac{y}{x_i} &= \frac{1}{p_i} \\
\frac{\alpha_i}{(1-\alpha_i)} \frac{\left(\dfrac{w_i - x_i}{p_i}\right)}{x_i} &= \frac{1}{p_i} \\
\frac{\alpha_i}{(1-\alpha_i)} \left(\frac{w_i - x_i}{x_i p_i}\right) &= \frac{1}{p_i} \\
\alpha_i w_i - \alpha_i x_i &= x_i - \alpha_i x_i \\
x_i &= \alpha_i w_i
\end{aligned} \tag{2.120}$$

A função demanda pelo bem y_i *é dada pela condição:*

$$\begin{aligned}
TMS_{x_i,y} &= \frac{p_{x_i}}{p_i} \\
\frac{\alpha_i x_i^{\alpha_i - 1} y^{1-\alpha_i}}{(1-\alpha_i) x_i^{\alpha_i} y^{-\alpha_i}} &= \frac{1}{p_i} \\
\frac{\alpha_i}{(1-\alpha_i)} \frac{y}{x_i} &= \frac{1}{p_i}
\end{aligned}$$

$$
\begin{aligned}
\frac{\alpha_i}{(1-\alpha_i)}\frac{y}{(w_i-p_iy)} &= \frac{1}{p_i} \\
\frac{\alpha_i}{(1-\alpha_i)} &= \frac{w_i-p_iy}{p_iy} \\
\frac{\alpha_i}{(1-\alpha_i)} &= \frac{w_i}{p_iy}-1 \\
\frac{1}{1-\alpha_i} &= \frac{w_i}{p_iy} \\
y &= (1-\alpha_i)\frac{w_i}{p_i} \qquad (2.121)
\end{aligned}
$$

Em equilíbrio, todos os consumidores demandam a mesma quantia do bem público, isto é, $y_1 = y_2 = ... = y_n = y^*$*. Portanto, temos:*

$$
\begin{aligned}
p_iy^* &= (1-\alpha_i)w_i \\
\sum_{i=1}^{n} p_iy^* &= \sum_{i=1}^{n}(1-\alpha_i)w_i \\
y^*\sum_{i=1}^{n} p_i &= \sum_{i=1}^{n}(1-\alpha_i)w_i \\
y^*p &= \sum_{i=1}^{n}(1-\alpha_i)w_i \\
y^* &= \frac{\sum_{i=1}^{n}(1-\alpha_i)w_i}{p} \qquad (2.122)
\end{aligned}
$$

Disso decorre que

$$
\begin{aligned}
p_i &= \frac{(1-\alpha_i)w_i}{y^*} \\
p_i &= \frac{(1-\alpha_i)w_i}{\frac{\sum_{i=1}^{n}(1-\alpha_i)w_i}{p}} \\
p_i &= \frac{p(1-\alpha_i)w_i}{\sum_{i=1}^{n}(1-\alpha_i)w_i} \qquad (2.123)
\end{aligned}
$$

A diferença chave entre o equilíbrio de Lindahl e o equilíbrio padrão é que nenhum mecanismo descentralizado irá gerar o vetor de preços correto. Esse ponto já foi apontado por Samuelson e ocorre devido ao fato de que as pessoas têm interesses próprios e emitem uma sinalização incorreta, isto é, dão informações erradas. Podemos contornar esse problema ao analisarmos um mecanismo de revelação de preferências. Se queremos encontrar um equilíbrio de Lindhal, precisamos conhecer as preferências ou a taxa marginal de substituição de cada consumidor. Porém, devido ao problema do *free-rider*, é muito difícil para os consumidores revelarem suas preferências. Vamos analisar esse problema antes.

2.6 O Problema do *Free-Rider*

Quando a taxa marginal de substituição é conhecida, a alocação eficiente de Pareto pode ser determinada a partir da condição de Lindahl-Samuelson ou da solução de Lindahl. Depois disso, a contribuição de cada consumidor é dada por $g_i = w_i - x_i$. No entanto, é difícil a sociedade conhecer as informações sobre a taxa marginal de substituição. Obviamente, o método ingênuo é que poderíamos pedir a cada indivíduo que revelasse suas preferências e, assim, determinasse a sua disposição a pagar. Contudo, como cada consumidor tem interesse próprio, cada pessoa quer ser uma pessoa livre. Portanto, não está disposta a contar a sua verdadeira taxa marginal de substituição. Se os consumidores perceberem que as parcelas da contribuição para a produção de bens públicos (ou preços personalizados) dependem de suas respostas, eles terão "incentivos para trapacear". Ou seja, quando for solicitado que os consumidores relatem suas funções de utilidade ou taxas marginais de substituição, eles têm incentivos para reportar uma taxa marginal de substituição menor, para que possam pagar menos e consumir o bem público (*free riders*). Isso causa os maiores problemas em teoria da escolha pública.

Observe que o objetivo social é atingir alocações eficientes de Pareto para a economia de bens públicos, mas, por interesse pessoal, cada pessoa procura otimizar a sua função de utilidade:

$$\max u_i(x_i, y) \tag{2.124}$$

$$\text{sujeito a } g_i \in [0, w_i] \tag{2.125}$$

$$x_i + g_i = w_i \tag{2.126}$$

$$y = f\left(g_i + \sum_{j \neq i}^{n} g_j\right) \tag{2.127}$$

Ou seja, cada consumidor adota as estratégias dos outros como determinadas e maximiza seus ganhos. A pergunta que resulta disso é: quão eficaz é um mercado privado no fornecimento de bens públicos? A resposta, como mostrada abaixo, é que não podemos esperar que uma decisão puramente independente resulte em uma quantidade eficiente do bem público que está sendo produzido.

Observe que podemos reescrever a função *payoff* do agente i como

$$\phi_i(g_i, g_{-i}) = u_i\left[(w_i - g_i), f\left(g_i + \sum_{j \neq i}^{n} g_j\right)\right] \tag{2.128}$$

A condição de primeira ordem do problema acima, obtida pela regra da cadeia, é

$$\begin{aligned}\frac{\partial \phi_i(g_i, g_{-i})}{\partial g_i} = 0 \quad &\Longleftrightarrow \quad \frac{\partial u_i(x_i, y)}{\partial x_i}\frac{\partial x_i}{\partial g_i} + \frac{\partial u_i(x_i, y)}{\partial y}\frac{\partial y}{\partial g_i} = 0 \\ &\Longleftrightarrow \quad \frac{\partial u_i(x_i, y)}{\partial x_i}(-1) + \frac{\partial u_i(x_i, y)}{\partial y}\left[f'\left(g_i^* + \sum_{j \neq i} g_j\right)\right] = 0\end{aligned} \tag{2.129}$$

Então, utilizando a equação (2.129) para escrever a taxa marginal de substituição entre o bem público e os bens privados, temos como solução:

$$\frac{\frac{\partial u_i(x_i^*, y^*)}{\partial y^*}}{\frac{\partial u_i(x_i^*, y^*)}{\partial x_i^*}} = \frac{1}{f'\left(g_i^* + \sum_{j \neq i} g_j\right)} \tag{2.130}$$

que não satisfaz a condição de Samuelson. Portanto, o equilíbrio obtido não necessariamente resulta em uma alocação Pareto eficiente.

Exemplo 2.7. *Para ver isso, suponha que*

- *A disposição a pagar por um bem seja* $r_i = 100$, $i = 1, 2$.
- *O custo* c *do bem seja R\$* 150.
- $g_i = 150$ *se um agente contribuir ou* $g_i = 75$ *se ambos contribuírem*

Cada pessoa decide independentemente se quer ou não comprar o bem público. Como resultado, cada um tem um incentivo para ser um free-rider do outro, como mostrado na matriz de payoff a seguir.

		Agente 2	
		Comprar	Não comprar
Agente 1	Comprar	$(25, 25)$	$(-50, 100)$
	Não comprar	$(100, -50)$	$(100, 100)$

Note que os payoffs líquidos são definidos por $r_i - g_i$. *Assim, é dado por* $100 - \frac{150}{2} = 25$ *quando ambos os consumidores estão dispostos a produzir o projeto público, e* $100 - 150 = -50$ *quando apenas uma pessoa quer comprar, mas a outra pessoa não. Este é o dilema do prisioneiro. O equilíbrio de estratégia dominante neste jogo é (não comprar, não comprar). Assim, nenhum agente quer dividir o custo de produção do projeto público, mas quer aproveitar com o outro consumidor.*

Como resultado, o bem público não é fornecido, mesmo que seja considerado eficiente. Assim, a contribuição voluntária em geral não resulta no nível eficiente do bem público.

Como podemos resolver o problema do *free-rider*? Para isso precisamos da teoria de desenho de mecanismo, que não é o foco aqui, mas vamos ver duas alternativas: o mecanismo de votação e o mecanismo de revelação de preferências.

2.7 Mecanismo de Votação

Desenvolvido inicialmente por Bowen (1943), Black (1948), Donws (1957), entre outros, o Modelo do Eleitor Mediano diz que, sob a hipótese de que as preferências dos eleitores apresentem "pico único", em um sistema eleitoral majoritário, os eleitores escolherão o candidato cuja cesta ofertada de bens e serviços públicos mais se aproxime da cesta demandada pelo eleitor mediano.

Na prática, o nível de provisão de bens públicos é frequentemente determinado pelo processo político, com os partidos concorrendo nos sistemas eleitorais prometendo diferentes níveis de provisão do bem público. A eleição de uma das partes por votação determina o nível de provisão do bem público. O que queremos fazer aqui é fornecer um contraste entre o resultado da votação e o nível eficiente de provisão de bem público, quando as pessoas diferem em gostos e (possivelmente) níveis de renda. Considere uma população de consumidores que determina a quantidade de bem público a ser fornecida pelo voto da maioria. O custo do bem público é compartilhado igualmente entre os consumidores e, portanto, se G unidades do bem público são fornecidas, o custo para cada consumidor é $\frac{G}{N}$. Com a renda M_i, um consumidor pode comprar bens privados no valor de $M_i - \frac{G}{N}$ depois de pagar pelo bem público. Isso fornece um preço efetivo de $\frac{1}{N}$ para cada unidade do bem público e um nível de utilidade $U_i\left(M_i - \frac{G}{N}, G\right)$. A restrição orçamentária, as curvas de indiferença mais altas atingíveis e a quantidade preferida de bem público são mostradas na parte superior da Figura 2.2 (assumindo, por conveniência, os mesmos níveis de renda para todos os consumidores).

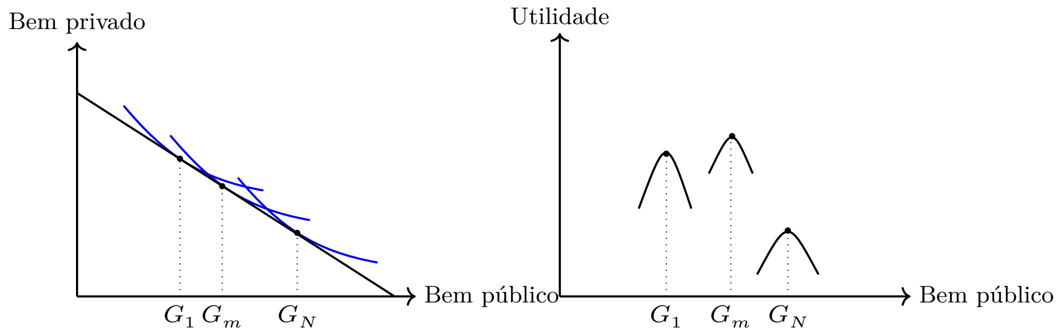
Sendo as preferências são de pico único, a votação majoritária fornecerá uma agregação consistente de preferências dos eleitores individuais. O Teorema do Eleitor Mediano afirma que a votação majoritária produzirá o resultado preferido pelo eleitor mediano se as preferências tiverem um único pico. O eleitor mediano é o eleitor cujos gostos estão no meio do conjunto de eleitores; portanto, um número igual de eleitores prefere mais e um número igual de eleitores prefere menos do bem público.

Para que o Teorema do Eleitor Mediano[11] possa ser aplicado, pressuponha que exista um número ímpar, N, de consumidores, em que $N > 2$, e que cada um dos consumidores tenha preferências únicas para o bem público. Essa segunda suposição implica que, quando o nível de utilidade é representado graficamente em relação à quantidade de bem público, haverá um valor único de G_i que maximiza a utilidade para o consumidor i. Tais preferências são ilustradas no painel à esquerda da Figura 2.2. Os consumidores são numerados de modo que seus níveis preferidos de bem público satisfaçam $G_1 < G_2 < ... < G_N$.

Por essas premissas, o Teorema do Eleitor Mediano garante que o consumidor com preferência mediana pelo bem público será decisivo na votação majoritária. A preferência mediana pertence ao consumidor na posição $\dfrac{N+1}{2}$ no *ranking*. Nós rotulamos o consumidor mediano como m e denotamos a quantidade escolhida do bem público por G_m. Uma característica notável do resultado da votação majoritária é que ninguém é capaz de manipular o resultado a seu favor deturpando sua preferência,

[11] A dinâmica e as previsões do teorema mediano dos eleitores apareceram pela primeira vez no artigo *Stability in Competition*, de 1929, do economista Harold Hotelling, no qual Hotelling observa que as plataformas dos candidatos políticos parecem convergir durante as eleições majoritárias. Seu artigo diz respeito ao posicionamento de lojas por dois vendedores ao longo de um segmento de linha, no qual os compradores são distribuídos uniformemente. A previsão de seu modelo, agora conhecido simplesmente como "Lei de Hotelling", é que em muitos mercados é racional que os produtores tornem seus produtos o mais semelhante possível, o chamado "princípio da diferenciação mínima". A análise formal do princípio da Lei de Hotelling nos sistemas de votação majoritária foi fornecida em um artigo relacionado de 1948, intitulado "On the Rationale of Group Decision-making" pelo economista Duncan Black. Anthony Downs, inspirado por Adam Smith, expandiu ainda mais o trabalho de Black em seu livro de 1957, "An Economic Theory of Political Action in Democracy".

Figura 2.2 – *Alocações por meio de Votações*



de modo que a votação sincera é a melhor estratégia. O motivo é que qualquer pessoa à esquerda da mediana só pode afetar o resultado final votando em uma quantidade à direita da mediana que afastaria o resultado da posição preferida e vice-versa para qualquer pessoa à direita da mediana.

O valor G_m é a escolha preferida do consumidor m, por isso resolve

$$\max_G U_m\left(M_m - \frac{G}{N}, G\right) \tag{2.131}$$

em que M_m indica a renda do eleitor mediano que pode diferir da renda mediana com preferências heterogêneas.

Montando o lagrangeano, temos:

$$L(G) = U_m\left(M_m - \frac{G}{N}, G\right) \tag{2.132}$$

A condição de primeira ordem para a maximização pode ser expressa em termos da taxa marginal de substituição para mostrar que o resultado da votação é descrito por

$$\frac{\partial L}{\partial G} = 0 \quad \Longleftrightarrow \quad \frac{\partial U_m}{\partial x_m}\frac{\partial x_m}{\partial G} + \frac{\partial U_m}{\partial G} = 0 \tag{2.133}$$

Reescrevendo essa condição obtemos:

$$\begin{aligned}\frac{\partial U_m}{\partial x_m}\left(-\frac{1}{N}\right)+\frac{\partial U_m}{\partial G}&=0\\ \frac{\partial U_m}{\partial G}&=\frac{\partial U_m}{\partial x_m}\left(\frac{1}{N}\right)\\ \frac{\frac{\partial U_m}{\partial G}}{\frac{\partial U_m}{\partial x_m}}\equiv TMS^m_{G,x_m}&=\frac{1}{N}\end{aligned} \tag{2.134}$$

Por outro lado, como a taxa marginal de transformação é igual a 1, o resultado eficiente satisfaz a regra de Samuelson

$$\sum_{i=1}^{N} TMS^m_{G,x_m}=1 \tag{2.135}$$

Contrastando os valores, o resultado da votação é eficiente apenas se

$$TMS^m_{G,x_m}=\sum_{i=1}^{N}\frac{TMS^m_{G,x_m}}{N} \tag{2.136}$$

Portanto, a votação majoritária leva à provisão eficiente do bem público somente se a TMS do eleitor mediano for igual à TMS média da população de eleitores. Não há razão para esperar que isso aconteça; portanto, deve-se concluir que a maioria dos votos geralmente não alcança um resultado eficiente. Isso ocorre porque o resultado da votação não leva em consideração outras preferências além das do eleitor mediano: alterar todas as preferências, exceto as do eleitor mediano, não afeta o resultado da votação (embora isso afete o nível eficiente de provisão do bem público).

O equilíbrio da votação é único e nenhuma outra alternativa é melhor. Porém, o equilíbrio de votação não existe em geral. Precisamos de duas suposições fortes para garantir a existência de um equilíbrio de votação.

- G é unidimensional (por exemplo, escolha do tamanho de um projeto público único: quão grande deve ser o orçamento de educação).

- as preferências acerca de G são de pico único, ou seja, cada pessoa tem um nível favorito unimodal de G e, em seguida, as preferências por outras alocações são muito pouco relevantes.

Suponha que o governo decidiu cobrar um imposto e fornecer bens públicos com base em alguma regra. Duas complicações surgem quando se tenta chegar a condição de Samuelson (*first-best*):

- existem interações com o setor privado (*crowd-out*).
- o governo não pode financiar bens públicos por meio de tributação *lump-sum*; ele deve usar impostos distorcivos.

Podem ser feitos comentários sobre se a votação majoritária geralmente leva a muito ou pouco bem público? Em geral, a resposta deve ser negativa, já que nenhuma restrição natural pode ser apelada e a TMS do eleitor mediano pode ser menor ou maior que a média. Se for menor, será fornecido muito pouco bem público. O inverso vale se for mais alto. A única abordagem que pode dar uma ideia é observar que a distribuição de renda tem uma cauda direita muito longa. Se a TMS for mais alta para os eleitores de baixa renda, a natureza da distribuição de renda sugere que a mediana da TMS é maior que a média. Assim, a votação levará a uma quantidade excessiva de bem público. Alternativamente, se a TMS estiver aumentando com a renda, a votação levaria a uma subprovisão.

Exemplo 2.8. *Suponha que o custo para produzir um bem público seja* $c = 99$*. E as disposições a pagar de três consumidores sejam, respectivamente,* $r_1 = 90$*,* $r_2 = 30$ *e* $r_3 = 30$*. Claramente,* $r_1 + r_2 + r_3 > c$ *e* $g_i = \frac{99}{3} = 33$*. Portanto, a provisão eficiente do bem público deve ser sim. No entanto, sob a regra da maioria, apenas o consumidor 1 vota "sim", pois recebe um benefício líquido positivo se o bem for fornecido. As pessoas #2 e #3 votam "não" para produzir bem público e, portanto, o bem público não será fornecido, e teremos uma provisão ineficiente do bem público.*

O problema com a regra da maioria é que ela mede apenas o benefício líquido de ter o bem público, enquanto a condição eficiente exige uma comparação da disposição a pagar.

O Paradoxo de Condorcet e o Teorema da Impossibilidade de Arrow são conceitos centrais na teoria da escolha social e na teoria da votação. Ambos abordam as dificuldades associadas à agregação das preferências individuais para formar uma decisão coletiva.

O Paradoxo de Condorcet, nomeado em homenagem ao matemático e filósofo francês Marquês de Condorcet, descreve uma situação onde as preferências individuais são cíclicas, levando a uma inconsistência na escolha coletiva[12]. Em termos simples, o paradoxo ocorre quando não há uma alternativa que seja a vencedora em todas as comparações par a par.

Imagine três candidatos: A, B e C. Suponha que temos três eleitores com as seguintes preferências:

Eleitor 1: $A \succ B \succ C$

Eleitor 2: $B \succ C \succ A$

[12] Em economia definimos a escolha coletiva, como a soma das decisões tomadas pela sociedade como um todo. Tal fato não restringi-se e nem deve ser confundido com decisões de grupo, nem tão pouco a decisões da maioria tão somente, a escolha coletiva até certo ponto deve ser compreendida como o valor agregado da vontade, dos agentes econômicos, uma utilidade geral representada por seus participantes; o governo, instituições e indivíduos. Quando nos debruçamos sobre o estudo da escolha coletiva, e a diferenciamos das decisões tomadas em grupo, o fazemos porque estas assumem um aspecto personalista, em alguns momentos no plano político de maneira emocional, inclusive. O entendimento é de que não é qualquer ajuntamento de pessoas que formam ou produzem uma vontade coletiva em stricto sensu. Nem tão pouco é razoável achar que a escolha coletiva tem em si mesmo uma qualidade altruística, que sempre deverá prevalecer, muito embora este seja um elemento admirável e bem quisto pela maioria das pessoas. A virtuosidade das escolhas coletivas e de suas vontades, não é função dependente de seus participantes, na verdade são apenas função dos critérios axiológicos de uma sociedade, principalmente na política liberal moderna. A teoria democrática baseia-se no conceito de vontade coletiva. No entanto, quando se trata de empreender uma análise detalhada, a teoria da escolha racional individual só encontra dificuldades ao abordar o conceito de comportamento coletivo. É axiomático, para teoria, que o comportamento racional se baseia em motivos de auto-referenciação. O indivíduo calcula o que é aquilo que melhor atende a seus interesses e age de acordo com isso. Este é o fundamento da teoria sobre a qual se baseia a análise econômica e política, e, no entanto ficamos com impressão contrária. Nossa intuição nos diz que os indivíduos contribuem, sim, para o bem público generosamente até mesmo sem hesitações, sem a intenção óbvia de obter benefício próprio.

Eleitor 3: $C \succ A \succ B$

Ao comparar as alternativas par a par:

A vence B (2 votos a 1).

B vence C (2 votos a 1).

C vence A (2 votos a 1).

Neste exemplo, não há uma alternativa que seja a vencedora em todas as comparações, o que significa que há um ciclo entre as preferências. Esse ciclo é o cerne do Paradoxo de Condorcet, demonstrando que não há uma preferência coletiva clara entre as alternativas.

O paradoxo mostra que é possível que o processo de votação não produza uma decisão consistente ou transitiva, revelando uma limitação na forma como as preferências individuais podem ser agregadas para formar uma decisão coletiva.

O Paradoxo de Condorcet é formalizado matematicamente no contexto da teoria da escolha social, onde preferências individuais são agregadas para chegar a uma decisão coletiva. Vamos descrever a formalização do paradoxo em termos matemáticos.

Sejam as seguintes definições:

1. Preferências Individuais:

   Suponha que temos um conjunto de n eleitores e um conjunto de m alternativas $X = \{x_1, x_2, \dots, x_m\}$. Cada eleitor i tem uma preferência sobre essas alternativas, que pode ser representada por uma relação de preferência $\succ_i$ (ou seja, $x \succ_i y$ significa que o eleitor i prefere x a y).

2. Preferência Coletiva:

   A preferência coletiva pode ser representada por uma função de agregação de preferências, que combina as preferências individuais para formar uma preferência social.

A formalização do Paradoxo de Condorcet pode ser escrita como segue:

1. Preferências Individuais e Votação

   Considere um cenário com três alternativas x_1, x_2, x_3 e três eleitores. As preferências dos eleitores podem ser representadas por:

   - Eleitor 1: $x_1 \succ_1 x_2 \succ_1 x_3$
   - Eleitor 2: $x_2 \succ_2 x_3 \succ_2 x_1$
   - Eleitor 3: $x_3 \succ_3 x_1 \succ_3 x_2$

2. Comparações Par a Par

   A ideia central do paradoxo é que, mesmo que uma alternativa não seja claramente preferida por todos, pode ainda assim vencer em todas as comparações diretas contra outras alternativas. Formalmente, definimos a preferência social $\succ_s$ para comparação par a par das alternativas como:

   - $x_1 \succ_s x_2$ se a maioria dos eleitores prefere x_1 a x_2.
   - $x_2 \succ_s x_3$ se a maioria dos eleitores prefere x_2 a x_3.
   - $x_3 \succ_s x_1$ se a maioria dos eleitores prefere x_3 a x_1.

3. Ciclo de Condorcet

   O Paradoxo de Condorcet ocorre quando o resultado da votação revela um ciclo. No exemplo acima, analisamos as comparações par a par:

   - x_1 vence x_2: os votos são 1 para $x_1 \succ x_2$ e 2 para $x_2 \succ x_1$.
   - x_2 vence x_3: os votos são 1 para $x_2 \succ x_3$ e 2 para $x_3 \succ x_2$.
   - x_3 vence x_1: os votos são 1 para $x_3 \succ x_1$ e 2 para $x_1 \succ x_3$.

Assim, o ciclo é $x_1 \succ_s x_2, \quad x_2 \succ_s x_3, \quad x_3 \succ_s x_1$, que mostra que há um ciclo nas preferências coletivas, onde nenhuma alternativa é claramente preferida sobre todas as outras.

Para formalizar matematicamente, podemos usar uma matriz de preferências para representar o ciclo. Se denotarmos as preferências coletivas como uma matriz P onde a entrada P_{ij} representa a comparação entre x_i e x_j, temos:

$$P = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{2.137}$$

Aqui, $P_{ij} = 1$ indica que x_i é preferido a x_j por mais da metade dos eleitores. O ciclo $x_1 \succ x_2 \succ x_3 \succ x_1$ é evidente pela presença de 1s na matriz, indicando que cada alternativa vence a seguinte na sequência.

O Paradoxo de Condorcet revela que mesmo com um sistema de votação que parece razoável, a agregação das preferências individuais pode levar a um resultado incoerente ou cíclico. Isso demonstra a dificuldade em criar um sistema de votação que sempre produza uma decisão clara e consistente a partir das preferências individuais.

O Teorema da Impossibilidade de Arrow, formulado pelo economista Kenneth Arrow, afirma que não há um sistema de votação que satisfaça simultaneamente um conjunto de condições consideradas razoáveis para uma agregação justa e consistente das preferências individuais.

As condições de Arrow são:

- Dominância Pareto: se todos os indivíduos preferem A a B, então a decisão coletiva também deve preferir A a B.

- Não-Dicotomia: o sistema de votação não deve ser afetado pela introdução de alternativas que não sejam preferidas pela maioria.

- Independência das Alternativas Irrelevantes: a escolha entre duas alternativas não deve ser influenciada pela introdução ou remoção de uma terceira alternativa irrelevante.
- Universalidade: o sistema deve ser capaz de lidar com todas as possíveis preferências individuais.
- Transitividade: a decisão coletiva deve ser consistente (se A é preferido a B e B é preferido a C, então A deve ser preferido a C).

Arrow provou que é impossível satisfazer todas essas condições ao mesmo tempo quando há três ou mais alternativas. Em outras palavras, não existe um sistema de votação que possa garantir uma escolha coletiva que seja ao mesmo tempo racional, consistente e justa, conforme as condições acima.

O Teorema da Impossibilidade de Arrow revela limitações fundamentais nos sistemas de votação e na teoria da escolha social. Ele sugere que qualquer método de agregação de preferências será forçado a abrir mão de uma ou mais condições desejáveis, levando a algum tipo de comprometimento ou limitação na representação das preferências coletivas.

Vamos formalizar matematicamente o Teorema da Impossibilidade de Arrow. Sejam as seguintes definições:

1. Conjunto de Alternativas

   Seja $X = \{x_1, x_2, \ldots, x_m\}$ um conjunto de alternativas disponíveis.

2. Preferências Individuais

   Cada eleitor i tem uma relação de preferência $\succ_i$ sobre as alternativas em X. Esta relação pode ser representada por uma função de utilidade u_i tal que para quaisquer $x, y \in X$, $x \succ_i y$ se e somente se $u_i(x) \geq u_i(y)$.

3. Função de Agregação Social

Definimos uma função de agregação social f que mapeia um perfil de preferências individuais $\succ = (\succ_1, \succ_2, \dots, \succ_n)$ para uma relação de preferência social $\succ_s$ sobre X. Formalmente:

$$f : \mathrm{Pref}(X)^n \to \mathrm{Pref}(X)$$

em que $\mathrm{Pref}(X)$ é o conjunto de todas as relações de preferência sobre X.

O Teorema de Arrow considera a função de agregação social f sob os seguintes axiomas:

1. Unanimidade (ou Pareto eficiência): se todos os eleitores preferem a alternativa x à alternativa y, então a sociedade também deve preferir x a y:

$$\forall \succ \text{ tal que } \forall i, x \succ_i y \Rightarrow x \succ_s y$$

2. Não-ditadura: não deve haver um eleitor cuja preferência determine a preferência social final. Ou seja, nenhum eleitor deve ter a capacidade de decidir sozinho a preferência social:

$$\nexists i \text{ tal que } \forall \succ, \forall x, y \in X, x \succ_i y \Rightarrow x \succ_s y$$

3. Independência das Alternativas Irrelevantes (IAI): a preferência social entre duas alternativas x e y deve depender apenas das preferências relativas entre x e y e não das preferências sobre outras alternativas. Formalmente: se $\succ$ e $\succ'$ concordam sobre a relação entre x e y, então $f(\succ)$ e $f(\succ')$ devem concordar sobre a relação entre x e y.

4. Transitividade: a relação de preferência social $\succ_s$ deve ser transitiva. Ou seja,

se $x \succ_s y$ e $y \succ_s z$, então $x \succ_s z$:

$$\text{Se } x \succ_s y \text{ e } y \succ_s z, \text{ então } x \succ_s z$$

O Teorema de Impossibilidade de Arrow afirma que não existe uma função de agregação social f que satisfaça todos os axiomas simultaneamente, quando $n \geq 3$ e $m \geq 3$. Formalmente: não existe uma função de agregação social f que satisfaça os axiomas de unanimidade, não-ditadura, independência das alternativas irrelevantes e transitividade.

Considere três alternativas x, y e z e três eleitores. Vamos trabalhar com o seguinte perfil de preferências:

- Eleitor 1: $x \succ_1 y \succ_1 z$
- Eleitor 2: $y \succ_2 z \succ_2 x$
- Eleitor 3: $z \succ_3 x \succ_3 y$

Esse perfil gera um ciclo nas preferências dos eleitores: $x \succ_1 y \succ_2 z \succ_3 x$. Vamos aplicar os axiomas para provar o teorema. Unanimidade: se todos os eleitores preferem x a y, então x deve ser socialmente preferido a y. No entanto, no perfil dado, não há unanimidade, mas é importante verificar como a função de agregação social se comporta com ciclos. Independência das Alternativas Irrelevantes: considere o seguinte cenário alternativo, onde mudamos a preferência entre x e y, mantendo o resto fixo. Se a função de agregação social deve cumprir a IAI, então a preferência social entre x e y deve ser determinada exclusivamente pelas preferências entre x e y e não pelas preferências sobre z. Transitividade: para a relação social ser transitiva, se $x \succ_s y$ e $y \succ_s z$, então deve ser $x \succ_s z$. Vamos verificar se é possível garantir a transitividade com base no perfil dado.

Considerando a aplicação dos axiomas, verifique a consistência dos resultados. No perfil cíclico acima: se a função de agregação social deve ser transitiva, ela deve respeitar as relações $x \succ_s y$ e $y \succ_s z$ implicando $x \succ_s z$, mas isso não é garantido em um ciclo. Para satisfazer a IAI, o resultado da comparação entre x e y deve depender apenas das comparações entre x e y, e não de z. No entanto, a presença de ciclos pode violar essa condição. A unanimidade pode ser difícil de garantir em ciclos porque mesmo que todos os eleitores concordem em preferir uma alternativa, a preferência social pode não refletir essa unanimidade devido ao ciclo.

Se tentarmos uma função de agregação social que satisfaz todos os axiomas simultaneamente, enfrentamos a dificuldade de lidar com ciclos nas preferências individuais e garantir a transitividade e a IAI. A presença de ciclos e a necessidade de satisfazer unanimidade, não-ditadura, e transitividade tornam impossível atender a todos esses requisitos ao mesmo tempo.

A prova do Teorema da Impossibilidade de Arrow mostra que, com base em um perfil de preferências cíclicas, não é possível encontrar uma função de agregação social que satisfaça simultaneamente todos os axiomas propostos. A presença de ciclos nas preferências dos eleitores e a necessidade de garantir condições como unanimidade, não-ditadura, e transitividade leva a contradições inevitáveis, demonstrando a impossibilidade de satisfazer todos os axiomas simultaneamente em um sistema de escolha social.

Esta formalização mais detalhada ajuda a entender como a violação de qualquer um dos axiomas pode ser inevitável quando lidamos com preferências individuais complexas e a impossibilidade de encontrar uma solução perfeita de agregação social.

O Teorema da Impossibilidade de Arrow demonstra que não há um método perfeito para agregar preferências individuais em uma decisão social coletiva que satisfaça simultaneamente condições razoáveis de justiça e consistência. Isso tem profundas implicações para a teoria da escolha social e a compreensão das limitações dos sistemas de votação e tomada de decisão coletiva.

Em resumo, o Paradoxo de Condorcet reflete a possibilidade de ciclos nas preferências coletivas, onde não há uma alternativa claramente superior em todas as comparações par a par. O Teorema da Impossibilidade de Arrow demonstra que é impossível criar um sistema de votação que satisfaça simultaneamente todas as condições consideradas razoáveis para uma escolha coletiva justa e consistente.

Ambos os conceitos sublinham os desafios e complexidades na tentativa de agregar preferências individuais em decisões coletivas, revelando a dificuldade em encontrar um sistema de votação perfeito.

Críticas da Ciência Política

A teoria do eleitor mediano é aquela que diz que todo candidato deve caminhar ao centro para aumentar suas chances de sucesso eleitoral. Mas essa teoria tem várias limitações que tem sido ignoradas por estrategistas de campanha, especialmente na esquerda. São elas:

1. As preferências dos eleitores não são unidimensionais. Na vida real, as pessoas votam considerando múltiplos temas simultaneamente: economia, segurança, costumes, meio ambiente.

2. A teoria ignora o papel das redes sociais e bolhas informacionais. O ambiente digital fragmentado de hoje impede a formação de consensos e preferências medianas claras.

3. A teoria foi pensada para o bipartidarismo americano e tem aplicação limitada em outros contextos. Sistemas multipartidários complexos, como o do Brasil, não se encaixam perfeitamente no modelo.

4. A teoria assume racionalidade perfeita dos eleitores, ignorando vieses cognitivos, emoções e lealdades partidárias históricas que influenciam o voto.

5. O modelo não considera a abstenção seletiva. Quando certos grupos se abstêm mais que outros, a “mediana” real do eleitorado fica distorcida.

6. O papel do dinheiro e do poder econômico na política distorce as preferências

medianas. Grupos com mais recursos podem influenciar desproporcionalmente o debate público.

7. A dinâmica das redes sociais e da mídia moderna permite que posições minoritárias ganhem mais visibilidade e influência do que seu peso numérico sugeriria.

8. Diversos contextos eleitorais contrariam a tese do eleitor mediano. Candidatos que mobilizam suas bases através de programas e discurso radicais tem sucesso eleitoral e candidatos que se movem ao centro perdem força.

9. A verdade é que não há uma lei sobre sucesso eleitoral. Cada disputa deve ser analisada em sua particularidade para encontrar o melhor caminho para a vitória. Levar a teoria do eleitor mediano como uma lei geral pode virar atalho para derrotas vergonhosas.

Respostas da Ciência Econômica

1. Sobre unidimensionalidade das preferências: é verdade que Downs assume que os eleitores apenas ligam para um elemento econômico – redistribuição. Mas, todavia, porém, isto é apenas uma simplificação, não é papel do modelo descrever a realidade, mas criar alguma simplificação tratável que ajude a pensar a realidade sob determinadas condições. Downs era um economista ortodoxo, modelos ortodoxos não servem para explicar a realidade de propósito, mas eles permitem pensar expedientes que ajudam a compreender a realidade e servem como modelos-base, a partir dos quais se relaxam hipóteses. Então, vamos voltar à questão de análise unidimensional. Ela facilita o cálculo, porque quando inserimos mais dimensões, fica mais difícil analisar o equilíbrio, isso porque cada pauta vai ter saliência diferente e, mais do que isso, políticos não vendem pauta a pauta, mas o pacote inteiro. Agora, o fato de ser mais difícil não significa que é impossível. A grande questão é que agora temos mais hipóteses a levar em conta. Uma primeira solução foi traçar cada dimensão, e aí a seleção varia de

autor a autor[13]. Outros autores propõem que se tire a média ponderada entre os medianos de cada eixo, de forma que há expedientes computacionais a partir de simulações de Monte Carlo para tal[14]. Por sinal, em experimentos de laboratório, há evidências que apenas olhar para o mediano via voto é mais fácil de extrair o resultado[15]. No fim do dia, tudo depende de como você está disposto a agregar as preferências dos votos[16]. Inclusive, temos algumas evidências que elas nem são tão desagregáveis assim, tanto para políticos quanto para eleitores[17]. O que isso significa? Que na realidade as preferências dos eleitores podem nem ser tão complexas assim, porque uma depende da outra. Ele votar em um candidato mais ou menos punitivista é uma decisão de segunda ordem face à posição econômica do candidato, digamos. Ou só se preocupa com questões de direitos humanos/pautas sociais depois que seleciona por pautas econômicas. É o que essa pesquisa do YouGov e Jacobin[18], para ficar em instituições marcadamente progressistas mostram: para o eleitor *working class*, economia é uma pauta fundamental, acabar com racismo é uma pauta secundária. Mas há dissenso na economia, não no racismo. Essa é a parte mais complicada da coisa. Com multidimensionalidade, achar o equilíbrio fica significativamente mais difícil, mas ainda há um equilíbrio de mediano. A diferença é que o Agenda Setter passa a ter mais poder (Teorema de Dougherty-Edward)[19] O que isso tudo quer dizer? Que ainda há um equilíbrio Pareto ótimo mesmo com multidimensiona-

[13] Ver, por exemplo, https://ricardodahis.com/papers/Dahis_Ideology.pdf, que propôs medir a distribuição de preferências em cada eixo e tirar o eleitor médio de todos os eixos.
[14] Ver, por exemplo, https://link.springer.com/chapter/10.1007/978-3-319-05158-1_10.
[15] Ver, por exemplo, https://sciencedirect.com/science/article/abs/pii/S0899825621001159.
[16] Ver, por exemplo, https://cambridge.org/core/books/abs/following-their-leaders/preference-aggregation-through-voting/F1D696A192F7A9F1597017614D4E9ACB.
[17] Ver, por exemplo, https://cambridge.org/core/journals/political-analysis/article/nonseparable-preferences-in-the-statistical-analysis-of-roll-call-votes/6C690059DF08B11EB136EA3AA20D9AE8.
[18] Ver, por exemplo, https://images.jacobinmag.com/wp-content/uploads/2021/11/08095656/CWCPReport_CommonsenseSolidarity.pdf.
[19] Ver, por exemplo, https://spia.uga.edu/faculty_pages/dougherk/irc_multiple_dimensions.pdf, https://journals.sagepub.com/doi/abs/10.1177/0951629809347557 e https://jstor.org/stable/41406947?seq=1.

lidade e ele tende a ser o do mediano. O mediano vai funcionar sempre? Não, mas vai funcionar muito, quase sempre.

2. A segunda afirmação, a saber "a teoria ignora o papel das redes sociais e bolhas informacionais" é equivocada. Ela ignora o papel das redes sociais porque foi formulada em 1957, mas não ignora o papel das bolhas informacionais não, muito pelo contrário. Inclusive é justamente porque há bolhas informacionais que o modelo funciona. Vejam, parte fundamental do modelo de Downs (1957) é que se informar é custoso. Se informar sobre eleições e políticos dá trabalho e tira tempo de lazer, que os eleitores valorizam. No modelo de Downs, o problema fundamental não é que os candidatos não vão ao eleitor mediano, é que os eleitores não possuem informação o suficiente sobre os candidatos. Caso contrário, se os eleitores fossem perfeitamente informados, todos os candidatos sempre iriam exatamente às preferências do mediano e aí o resultado em uma democracia seria o mesmo que jogar uma moeda, os candidatos seriam escolhidos aleatoriamente. É justamente porque os eleitores não são capazes de se informar completamente que partidos existem. Eles existem como atalhos informacionais, isto é, servem para o eleitor não ter que se informar, se ele possui preferências à esquerda ou à direita da distribuição, ele sabe mais ou menos o que um partido à esquerda representa e o quanto se alinha ao candidato. Downs deixa isso bastante claro, o ato de votar é um ato expectacional, ele advém da utilidade esperada que um governo pode trazer a ele. Nesse sentido, o espaço informacional é fundamental, é o que define justamente o voto. A questão é, ninguém, absolutamente ninguém, disse que o único atalho informacional se daria apenas pelos partidos, podem vir por igrejas, empreendedores políticos, formadores de opinião, etc.

3. A teoria foi pensada para o bipartidarismo americano e tem aplicação limitada em outros contextos. Sim e não, de fato ela não funciona perfeitamente para

eleições multipartidárias (apesar de Downs escrever sobre elas) e certamente não para proporcionais como o legislativo. Inclusive Downs dedica dois capítulos para falar do multipartidarismo. Mas é falso que o teorema do eleitor mediano não explica eleições brasileiras. Ele explica sim, explica o segundo turno de eleições para o executivo. Isso porque o segundo turno força um bipartidarismo prático. Isso nem vem do Teorema do Eleitor Mediano, pelo contrário, vem dos estudos da Lei de Duverger para o Brasil[20]. Fernando Guarnieri mostra isso aqui[21]. Olhando de outra forma, Cesar Zucco Jr. e David Samuels mostram que há (ou havia) no Brasil apenas três partidos de fato: petistas, anti-petistas e independentes[22]. Essa tendência efetiva ao bipartidarismo por fatores institucionais faz com que eleições para o executivo se polarizem e virem eleições bipartidárias, seja no primeiro turno quando há voto estratégico, seja no segundo turno. Não tem como fugir do bipartidarismo e aí, o teorema do eleitor mediano se aplica, não tem muito como fugir. O problema inerente do sistema brasileiro é que temos um multipartidarismo no primeiro turno, e nesse caso a base social conta. Isso faz com que, diversas vezes, um partido que não possui chance alguma de ganhar eleições vai para o segundo turno e o candidato que é o preferido pelo mediano dentre as preferências postas não vai[23].

4. A teoria assume racionalidade perfeita dos eleitores, ignorando vieses cognitivos, emoções e lealdades partidárias históricas que influenciam o voto. O problema aqui é a definição errada de racionalidade que a ciência política usa. Citando Becker e Stigler, *de gustibus non disputandum est*, se emoções etc. informam as preferências do eleitor e o eleitor consegue selecionar entre duas alternativas com base em preferências, é racional.

[20] Ver, por exemplo, `https://princeton.edu/~fujiwara/papers/duverger_site.pdf`.
[21] Ver, por exemplo, `https://www.scielo.br/j/rbcsoc/a/WnC9dqbZwQGZwj9yf4ggMZB/`.
[22] Ver, por exemplo, `https://cambridge.org/core/books/partisans-antipartisans-and-nonpartisans/DCB30432C466A8346B0468F52E4F94A4`.
[23] Ver, por exemplo, `https://sciencedirect.com/science/article/pii/S0176268024000508`.

5. O modelo não considera a abstenção seletiva. Quando certos grupos se abstêm mais que outros, a "mediana" real do eleitorado fica distorcida. Sim e não, há pesquisa que mostra que sob determinadas questões (que não necessariamente se aproximam da realidade municipal), abstenção e polarização podem fazer com que se desvie ao centro[24], mas essa pesquisa é minoritária e isso porque Downs deixa explícito que o eleitor mediano se dá sobre aqueles que votam, ou seja, é uma análise do voto efetivo, não do voto total. O que nós percebemos cada vez mais, no entanto, é que a polarização pode sim ser resultado de uma análise downsiana[25]. E percebemos que a abstenção surge justamente dos problemas institucionais que colocam no segundo turno candidatos que não estão exatamente próximos do mediano. Vejam, o Teorema do Eleitor Mediano afirma que quem está mais próximo do centro da distribuição, leva, não que você precisa estar no centro para ganhar.

6. O papel do dinheiro e do poder econômico na política distorce as preferências medianas. Grupos com mais recursos podem influenciar desproporcionalmente o debate público. O teorema do eleitor mediano não diz nada sobre a formação das alternativas políticas ao eleitor, ele diz sobre como os eleitores vão escolher dentre alternativas apresentadas a eles. E isso pode acontecer por diversos fatores, inclusive por inclusão de eleitores.

7. Diversos contextos eleitorais contrariam a tese do eleitor mediano. Candidatos que mobilizam suas bases através de programas e discurso radicais tem sucesso eleitoral e candidatos que se movem ao centro perdem força. Sim, porque o teorema se aplica a um tipo de eleição apenas. O Teorema do Eleitor Mediano não diz que se mover ao centro é receita de sucesso. O que o teorema diz é que o candidato mais próximo do eleitor mediano vai levar a eleição majoritária em que há apenas dois candidatos. O eleitor mediano pode estar muito mais

[24] Ver, por exemplo, `https://nature.com/articles/s41599-022-01056-0`.

[25] Ver, por exemplo, `https://sciencedirect.com/science/article/pii/S0261379423000033`.

à direita do que progressistas acreditam, e muito mais à esquerda que conservadores acreditam, vemos isso em diversas pautas inclusive. Isso significa que seguir o Teorema do Eleitor Mediano à risca funciona? Não, claro que não, há muitos poréns nele. Por exemplo, não adianta nada você adotar uma postura entendida pelo eleitorado como radical na sua carreira política inteira e, em uma eleição, resolver guinar 180 graus. Não vai funcionar, eleitor não é burro e isso não é crível[26]. E, nesse caso, partidos importam negativamente. Instituições importam, e muito.

[26] Ver, por exemplo, https://jstor.org/stable/1811177.

3 Ciclos Políticos-Econômicos

Os piores leitores são aqueles que procedem como soldados saqueadores: escolhem algumas poucas coisas que podem utilizar, corrompem e confundem o restante, e blasfemam o todo.

Nietzsche

3.1 A Teoria dos Ciclos Políticos-Econômicos

Pode-se definir, de maneira simplificada, os ciclos políticos como sendo a sensibilidade das variáveis econômicas perante fatores de ordem política, ou seja, fatores associados ao ambiente político influenciando o comportamento da economia.

Em relação aos políticos, pode-se considerar a hipótese da existência de um incentivo ao governante para que este promova alterações de política econômica, visando elevar sua probabilidade de permanência no poder, ou então, se a probabilidade de reeleição do *policymaker* (ou de outro candidato de seu partido) é reduzida, pode existir um incentivo para que este distorça sua política econômica em seu último período de governo, visando prejudicar o governo do próximo administrador.

Tais fatores poderiam gerar os denominados "ciclos eleitorais", ou seja, a sensibilidade das variáveis econômicas ante a vigência do calendário eleitoral.

Já em relação aos partidos políticos, pode-se considerar a hipótese de que, caso exista um mínimo de consistência ideológica por parte destes e caso haja uma constante alternância no poder entre os mesmos, então a implementação de políticas de gestão pública estaria suscetível aos denominados "ciclos partidários", ou seja, flutuações econômicas associadas às diferenças de postura dos partidos políticos.

Como consequência, políticas de caráter de longo prazo (e provavelmente as de curto prazo também) poderiam apresentar aplicabilidade e retornos limitados.

3.2 Expectativas Adaptativas

Tradicionalmente, a teoria econômica busca explicar a ocorrência dos ciclos econômicos simplesmente através da análise do comportamento das variáveis econômicas, sem se preocupar, portanto, com o possível impacto de fatores políticos sobre este fenômeno.

Dentro desta abordagem, os modelos novo-keynesianos atribuem a existência de flutuações na economia às imperfeições do mercado, que ao enrijecerem os preços e salários possibilitam que as políticas fiscal e monetária tenham efeitos reais sobre as variáveis econômicas no curto prazo.

Por outro lado, a teoria dos ciclos econômicos reais (*real business cycles*) acredita que os preços e salários sejam flexíveis e que, por causa disso, alterações na política econômica afetam apenas as variáveis nominais. De acordo com esses modelos, os ciclos econômicos seriam gerados por choques tecnológicos, que ao elevarem a produtividade do trabalho, aumentam o salário dos trabalhadores e, consequentemente, a oferta de trabalho e o nível de emprego.

Paralelamente ao desenvolvimento destas teorias, em meados dos anos 70, emergiram na literatura econômica diversos artigos relacionando as flutuações econômicas aos eventos políticos.

Baseados na hipótese de que as decisões relativas à condução da política econômica tenham caráter eminentemente político, esses trabalhos propõem uma coincidência entre o calendário eleitoral e os ciclos econômicos.

Essa abordagem dos ciclos econômicos foi batizada de teoria dos ciclos político-econômicos (*political business cycles theory*).

Os modelos tradicionais são assim denominados porque foram desenvolvidos numa época em que prevalecia a crença de que a economia seria caracterizada por uma curva de Phillips de curto prazo, perfeitamente explorável pelos responsáveis pela política econômica.

A existência desse *trade off* permanente entre inflação e desemprego era viabilizada pela hipótese de que os agentes econômicos teriam expectativas adaptativas.

Posteriormente, com o advento das expectativas racionais na teoria econômica, estes modelos tiveram que ser revistos.

Outro aspecto que diferencia os modelos de ciclos político-econômicos é a questão da motivação dos políticos ao escolher a política econômica a ser implementada.

Alguns modelos pressupõem que os políticos sejam oportunistas, enquanto outros acreditam que eles defendem alguma ideologia.

Apesar de se tratar de uma proposta bastante intuitiva, tendo sido constantemente objeto de debates entre os economistas, faltava à relação entre política e economia uma modelagem mais formal.

Esta carência foi suprida com a publicação do artigo clássico de Nordhaus (1975), que impulsionou o estudo dos ciclos político-econômicos.

O modelo desenvolvido por Nordhaus (1975) está calcado na ideia defendida por Downs (1957) de que os políticos são "oportunistas", tendo como única motivação a maximização do número de votos nas eleições.

Seguindo esta linha de raciocínio, acredita que os governantes irão estimular a demanda agregada antes das eleições, através de política monetária expansionista, acelerando o crescimento econômico e reduzindo o desemprego, com o intuito de garantir a vitória do partido governista nas urnas.

Passadas as eleições, para reduzir as pressões inflacionárias geradas pela política anterior, será adotada uma política monetária restritiva, contraindo a atividade econômica e aumentando o desemprego.

Para que isto seja possível, Nordhaus (1975) parte da premissa de que existe um *trade-off* permanente entre inflação e desemprego na economia e que os *policymakers* tenham controle total sobre os instrumentos de política macroeconômica, o que os possibilita escolher o nível que desejam de ambas as variáveis.

Entretanto, propõe que este *trade-off* seja mais acentuado no curto do que no longo prazo, ou seja, uma redução da taxa de desemprego irá gerar maior inflação no longo do que no curto prazo.

O autor atribui esta diferença intertemporal do efeito de uma redução do nível de desemprego a duas razões:

1. devido ao mecanismo de transmissão da inflação (uma redução na taxa de desemprego inicialmente eleva os preços, que somente serão repassados aos salários posteriormente)

2. e porque uma elevação da inflação gera mudanças nas expectativas dos agentes quanto à inflação futura

Em suma, o sistema econômico é representado por uma curva de Phillips com expectativas adaptativas

$$\pi_t = f(u_t) + \lambda v_t, \quad 0 \leq \lambda < 1 \tag{3.1}$$

$$\frac{dv}{dt} = \gamma(\pi_t - v_t) \tag{3.2}$$

em que

- π_t é a taxa de inflação no período t
- u_t é a taxa de desemprego no período t
- $f(u_t)$ é a curva de Phillips de curto prazo t
- v_t é a taxa de inflação esperada no período t
- λ é a constante que capta os erros de previsão das inflações passadas

Quando a expectativa de inflação dos agentes coincidir com a inflação ocorrida no período, $\pi = v$, teremos uma curva de Phillips de longo prazo da seguinte forma

$$\begin{cases} \pi_t = \dfrac{f(u_t)}{1-\lambda}, & \text{se } 0 \leq \lambda < 1 \\ u = \bar{u}, & \text{se } \lambda = 1 \end{cases} \tag{3.3}$$

Como podemos observar, se o parâmetro $\lambda = 1$, a curva de Phillips de longo prazo será vertical, isto é, a economia estará sempre no pleno emprego e as políticas econômicas implementadas não terão efeitos reais sobre a economia.

Para modelar o comportamento dos eleitores, Nordhaus (1975) propõe uma função de utilidade individual $U^i(z)$, na qual estão ordenadas as preferências do eleitor em relação a cada variável macroeconômica $z = (z_1, z_2, \ldots, z_n)$, pressupondo que estes prefiram maior estabilidade dos preços e baixo desemprego, ao invés de elevadas taxas inflação e desemprego.

Portanto, temos

$$
\begin{aligned}
z_1 &= -\pi = -\text{taxa de inflação} \\
z_2 &= -u = -\text{taxa de desemprego} \\
z_3, \ldots, z_n &= \text{outras variáveis macroeconômicas}
\end{aligned}
$$

Os eleitores são considerados racionais nas suas preferências, porém ignorantes em relação ao funcionamento da economia e aos estímulos dos governantes para manipulá-la.

Desta maneira, formam suas expectativas baseados nas experiências passadas (expectativas adaptativas), ficando propensos a erros sistemáticos de previsão, já que são incapazes de antever possíveis alterações na política econômica.

Para simplificar o modelo, admite-se que as expectativas dos eleitores sejam constantes

$$\hat{z}_t = \hat{z}_{t-1} \tag{3.4}$$

Cada indivíduo vai escolher em qual candidato votará, se oposição ou situação, através da comparação da performance da economia no último ano de mandato $U^i(z_i)$ com sua expectativa $U^i\left(\hat{z}_i\right)$, ou seja, o eleitor vota retrospectivamente (*retrospective voting*).

Assim, a função de votação de cada eleitor i é a seguinte

$$V_t^i = \phi^i\left(z_t, \hat{z}_t\right) = \begin{cases} +1, & \text{se } \dfrac{U^i(z_t)}{U^i\left(\hat{z}_t\right)} > 1 \\ 0, & \text{se } \dfrac{U^i(z_t)}{U^i\left(\hat{z}_t\right)} = 1 \\ -1, & \text{se } \dfrac{U^i(z_t)}{U^i\left(\hat{z}_t\right)} < 1 \end{cases} \tag{3.5}$$

O eleitor vota no partido da situação quando a função de voto for igual a $+1$ (o desempenho da economia superou suas expectativas) e na oposição quando o resultado for -1 (o desempenho da economia ficou aquém das suas expectativas).

A função de votação agregada é

$$V_t = \sum_{i=1}^{n} V_t^i = \sum_{i=1}^{n} \phi^i\left(z_t, \hat{z}_t\right) \tag{3.6}$$

Como havíamos proposto anteriormente, a escolha das políticas públicas do governo será realizada visando maximizar esta função de votação agregada.

Antes de mostrarmos como os ciclos econômicos são gerados pelo governo no curto prazo, vamos nos deter na análise de qual o nível ótimo de inflação e desemprego numa economia no longo prazo.

Para obter o ponto de ótimo, Nordhaus (1975) utiliza como função de bem-estar social a própria função de votação agregada

$$V_t = g(u_t, \pi_t) \tag{3.7}$$

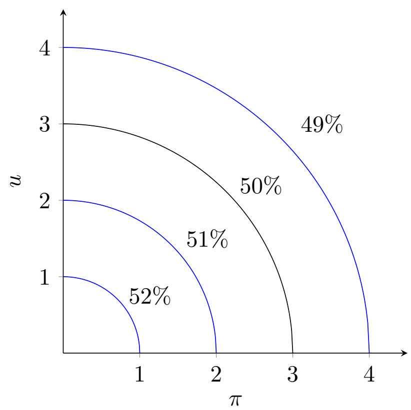
Da mesma forma que as curvas de indiferença representam o contorno da função de utilidade do indivíduo, podemos gerar curvas de "isovoto" a partir da função de bem-estar social.

As curvas de isovoto, apresentadas na Figura 3.1, representam as combinações dos níveis de inflação e desemprego que geram a mesma percentagem de eleitores que votariam no partido do governo.

Entretanto, ao contrário do que ocorre quando analisamos graficamente as curvas de indiferença, à medida que a curva de isovoto se aproxima da origem (taxas de inflação e desemprego menores), maior o bem-estar da sociedade e, consequentemente, maior o número de eleitores que votam pela reeleição.

A curva de isovoto com um traçado mais forte divide o gráfico na região de vitória (à esquerda) e de derrota (à direita) do partido de situação.

Figura 3.1 – *Curvas de Isovoto*



Para encontrar o nível de inflação e desemprego que maximiza a função de bem-estar social no longo prazo, representada pela função de votação agregada com um fator de desconto, resolvemos o seguinte problema de otimização dinâmica por meio de controle ótimo:

$$\max W = \int_0^\infty g(u_t, \pi_t) \mathrm{e}^{\rho t} dt \tag{3.8}$$

$$\text{sujeito a } \pi_t = f(u_t) + \lambda v_t \tag{3.9}$$

$$\frac{dv}{dt} = \gamma(\pi_t - v_t) \tag{3.10}$$

As soluções são[27]:

$$\frac{f'(u_t)}{1-\lambda} = -\frac{g_1}{g_2}\left(\frac{\rho+\gamma(1-\lambda)}{(\rho+\gamma)(1-\lambda)}\right), \quad \text{se } 0 \leq \lambda < 1 \tag{3.11}$$

$$f'(u_t) = -\frac{g_1}{g_2}\left(\frac{\rho}{\rho+\gamma}\right), \quad \text{se } \lambda = 1 \tag{3.12}$$

em que

- $f'(u)$ é a declividade da curva de Phillips de curto prazo
- g_1 é a derivada parcial da função de votação agregada em relação ao desemprego (u)
- g_2 é a derivada parcial da função de votação agregada em relação a inflação (π)

[27] Para resolver um problema de controle ótimo recorremos a Hamiltoniana, como segue:

$$H(u,v) = \mathrm{e}^{-\rho t}\left[g(u, f(u)+\lambda v)\right] + \psi_t\gamma[f(u)+\lambda v - v]$$

que deve também satisfazer a seguinte equação diferencial $\dot{\psi}(t) = [\rho+\gamma(1-\lambda)]\,\psi(t) - g_2\lambda$. Uma simples diferenciação mostra que H é maximizado quando

$$g_1 + g_2 f'(u) + \psi\gamma f' = 0$$

Procuramos soluções estacionárias, isto é, $\dot{\psi} = 0$, o que resulta em

$$\psi = \frac{g_2\lambda}{\rho+\gamma(1-\lambda)}$$

Assim, combinando as equações, obtemos:

$$\begin{aligned} g_1 + g_2 f'(u) + \psi\gamma f' &= 0 \\ g_1 + g_2 f'(u) + \left(\frac{g_2\lambda}{\rho+\gamma(1-\lambda)}\right)\gamma f' &= 0 \end{aligned}$$

Analisando as equações podemos verificar que $\frac{g_1}{g_2}$ é a declividade das curvas de isovoto, significando a taxa marginal de substituição entre desemprego e inflação.

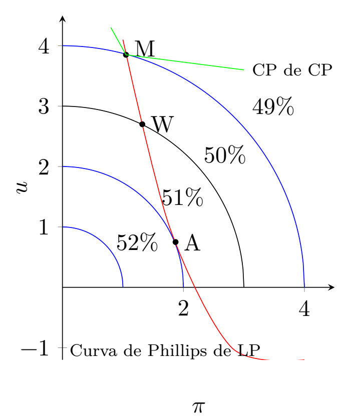
Considerando que $0 \leq \lambda < 1$, Nordhaus (1975) mostra três soluções de longo prazo supondo taxas de desconto intertemporal (ρ) diferentes para cada uma.

Caso as gerações presente e futura tenham o mesmo peso nas decisões dos responsáveis pela política econômica, a taxa de desconto intertemporal será $\rho = 0$. Então, temos como solução de longo prazo

$$\frac{f'(u_t)}{1-\lambda} = -\frac{g_1}{g_2} \tag{3.13}$$

Essa solução foi denominada política da regra de ouro (*golden rule policy*), representada pelo ponto A na Figura 3.2, no qual a curva de Phillips de longo prazo tangencia a função de votação agregada.

Figura 3.2 – *Resultados de Políticas Econômicas no Longo Prazo*



Caso as gerações futuras não tenham a menor importância para os responsáveis pela política econômica, a taxa de desconto intertemporal ρ tenderá ao infinito e a solução de longo prazo será

$$f'(u) = -\frac{g_1}{g_2} \tag{3.14}$$

Essa solução foi chamada de política puramente míope (*purely miopic policy*), representada pelo ponto M na Figura 3.2, no qual a curva de Phillips de curto prazo é tangente à função de votação agregada.

Se a taxa de desconto intertemporal estiver entre os dois extremos, $0 < \rho < \infty$, teremos soluções geralmente ótimas de bem-estar (*general welfare optimum*), que graficamente estarão entre os pontos A e M (por exemplo, o ponto W).

Neste caso, nos pontos de ótimo a inclinação da curva de isovoto será um valor entre a inclinação da curva de Phillips de curto e de longo prazo.

A proposta do modelo desenvolvido por Nordhaus é que numa democracia a taxa de desconto intertemporal tenda ao infinito, pois o horizonte temporal do *policymaker* é o fim do mandato, e que a solução estável de longo prazo do sistema econômico seja puramente míope, com menor taxa de desemprego e maior inflação do que o socialmente ótimo.

Após termos encontrado a solução de longo prazo do modelo, vamos centrar nossa análise no comportamento do governo no curto prazo, mostrando como as decisões dos governantes geram os ciclos político-econômicos.

Dados os pressupostos básicos do modelo apresentados anteriormente, sabendo que o período eleitoral é determinado exogenamente e supondo que o governante tenha perfeito conhecimento das preferências dos eleitores, este irá adotar um programa "politicamente ótimo" a fim de garantir a reeleição.

Para fazer isto, se defronta com uma função de votação multiplicada por um fator de decaimento da memória do eleitor, pois acredita-se que os eventos econômicos mais recentes tenham maior influência do que os passados na decisão do eleitor, ou seja, que o eleitor tenha memória curta.

A seguir apresentamos o modelo, no qual o *policymaker* escolhe a trajetória da taxa de desemprego ao longo do mandato que maximiza o número de votos na próxima eleição

$$\max V_\theta = \int_0^\theta g(u_t, \pi_t)\mathrm{e}^{\mu t} dt \tag{3.15}$$

$$\text{sujeito a } \pi_t = f(u_t) + \lambda v_t \tag{3.16}$$

$$\frac{dv}{dt} = \gamma(\pi_t - v_t) \tag{3.17}$$

em que θ é o mandato do governante, μ é a taxa de decaimento da memória do eleitor e $g(u_t, \pi_t)$ é a função de votação.

Supondo que as funções $g(u_t, \pi_t)$ e $f(u_t)$ sejam dados por

$$g(u_t, \pi_t) = -u^2 - \beta\pi, \quad \pi \geq 0, \beta > 0 \tag{3.18}$$

$$f(u_t) = \alpha_0 - \alpha_1 u, \quad \alpha_0 > 0, \alpha_1 > 0 \tag{3.19}$$

Asim, o problema de maximização se torna

$$\max V_\theta = \int_0^\theta \left[-\beta\alpha_0 - u^2 + \beta\alpha_1 u - \beta\lambda v\right] \mathrm{e}^{\mu t} dt \tag{3.20}$$

$$\text{sujeito a } \frac{dv}{dt} = \gamma\left(\alpha_0 - \alpha_1 u - (1-\lambda)v\right) \tag{3.21}$$

$$v(0) = v_0, v(\theta) \text{ livre}$$

Da equação (3.19) podemos observar que na função de votação do eleitor representativo, a desutilidade marginal da taxa de desemprego, $\left(\frac{\partial g}{\partial u}\right) = -2u$, é maior do que a da taxa de inflação, $\left(\frac{\partial g}{\partial \pi}\right) = -\beta$.

Resolvendo o problema obtemos a trajetória do desemprego que maximiza a quantidade de votos na eleição

$$u^*(t) = \frac{\beta\alpha_1}{2B}\left[(\mu-\gamma)+\gamma\lambda e^{B(\theta-t)}\right] \tag{3.22}$$

com $B = \mu - \gamma + \lambda\gamma$ e $B \neq 0$.

Assim,

$$\frac{du^*(t)}{dt} = -\frac{1}{2}\beta\alpha_1\lambda\gamma e^{B(\theta-t)} < 0 \tag{3.23}$$

No início do mandato, temos $t = 0$ e a taxa de desemprego

$$u^*(0) = \frac{\beta\alpha_1}{2B}\left[(\mu-\gamma)+\gamma\lambda e^{B(\theta)}\right] \tag{3.24}$$

No final do mandato, temos $t = \theta$ e a taxa de desemprego

$$u^*(0) = \frac{\beta\alpha_1}{2B}\left[(\mu-\gamma)+\gamma\lambda\right] = \frac{\beta\alpha_1}{2} > 0 \tag{3.25}$$

Analisando as Figuras 3.3 e 3.4 podemos comprovar a existência do ciclo político econômico proposto por Nordhaus.

A taxa de desemprego deve ser decrescente ao longo de todo o mandato, subindo a um nível relativamente elevado imediatamente após a eleição, para conter o aumento do nível de preços e modificar as expectativas inflacionárias dos agentes.

Observe que mesmo que no instante da eleição ($t = \theta$) a inflação seja bastante elevada, o governante conseguirá se reeleger, pois o eleitor representativo tem preferência pela redução do desemprego, conforme a equação (3.19).

Figura 3.3 – *Dinâmica da Taxa de Desemprego*

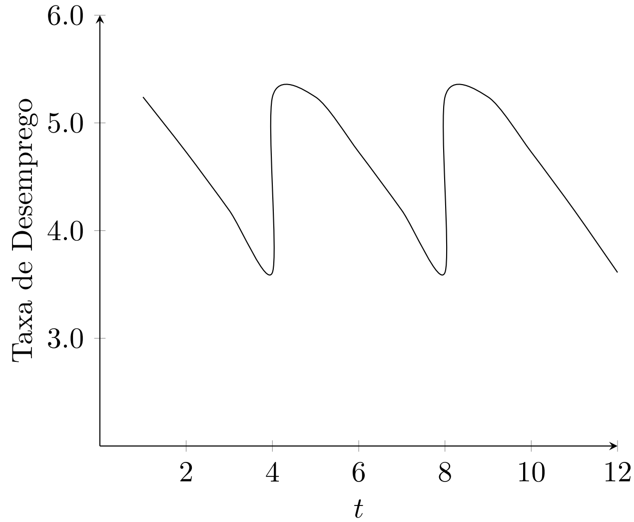
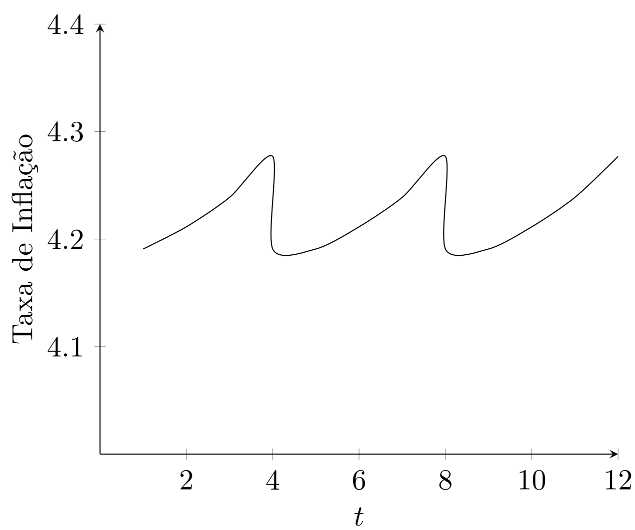


Figura 3.4 – *Dinâmica da Taxa de Inflação*



A velocidade com que a taxa de desemprego decresce ao longo do mandato, irá depender do sinal da segunda derivada da trajetória ótima. Se esta for positiva, a taxa de desemprego deverá cair com maior velocidade no início do que no final do mandato, se for negativa ocorrerá o oposto, isto é,

$$\frac{d^2u^*(t)}{dt^2} = \frac{1}{2}\beta\alpha_1\lambda\gamma B\mathrm{e}^{B(\theta-t)} \tag{3.26}$$

Na mesma época em que Nordhaus publicava seu artigo, de uma forma independente, MacRae (1977) preparava um modelo de ciclos político-econômicos semelhante.

A diferença principal entre os modelos está no fato de que MacRae (1977) não utilizou a taxa de decaimento da memória do eleitor, o que não o impediu de provar que o governo gera um ciclo econômico igual ao período eleitoral.

Outra diferença sutil está relacionada à função objetivo dos políticos.

Enquanto no modelo de Nordhaus estes procuram maximizar a função de votação agregada, no modelo de MacRae buscam minimizar a função de perda de votos (*vote-loss function*).

Os modelos oportunistas tradicionais ao pressuporem que os políticos eram idênticos, tendo como única motivação a permanência no poder e que isto guiaria as escolhas relativas à política econômica, ignoraram a influência da ideologia partidária sobre essas decisões.

Como os partidos representam a união de indivíduos que comungam dos mesmos interesses e preferências, é de se esperar que as suas ideologias se reflitam na maneira com a qual cada um conduz a política econômica quando estiver no poder.

Acreditando nessa hipótese, foram desenvolvidos modelos que incorporaram o componente de ideologia partidária à teoria dos ciclos político-econômicos, denominados modelos partidários tradicionais (*traditional partisan models*).

O autor conclui que ao observarmos o comportamento das variáveis macroeconômicas, notaremos a presença de um ciclo gerado por diferentes partidos alternando-se no poder.

Supondo que inicialmente a economia encontre-se no pleno emprego.

Caso um partido de direita vença as eleições, a prioridade será a redução da inflação, a política econômica será restritiva, gerando um aumento do desemprego.

Por outro lado, se uma partido de esquerda assumir o poder, a política econômica será conduzida de uma maneira mais "frouxa", pois o objetivo é a redução do desemprego, gerando pressões inflacionárias.

3.3 Expectativas Racionais

A partir da publicação da crítica de Lucas em meados dos anos 70, sugerindo que a racionalidade dos agentes deveria ser estendida à formação das expectativas, houve uma verdadeira revolução na teoria macroeconômica.

Ao invés de se basearem apenas nos fatos passados, como postulava a hipótese de expectativas adaptativas, Lucas propôs que os agentes otimizam todas as informações disponíveis para formar suas expectativas, incluindo-se aí as informações relativas às políticas econômicas vigentes.

Consequentemente, a capacidade dos *policymakers* de reduzirem o desemprego além da sua taxa natural[28] fica limitada, pois as políticas econômicas implementadas

[28] A taxa natural de desemprego é um conceito fundamental na economia que se refere à taxa de desemprego que persiste em uma economia quando está em equilíbrio a longo prazo, sem pressões inflacionárias ou deflações. Essa taxa é considerada natural porque reflete as condições normais de mercado, onde todos os trabalhadores que estão dispostos a trabalhar encontram emprego, e as empresas encontram trabalhadores adequados. A taxa natural de desemprego é composta principalmente por dois tipos de desemprego:

- Desemprego Friccional: Esse tipo de desemprego ocorre quando indivíduos estão temporariamente desempregados enquanto procuram um novo emprego ou estão se movendo entre empregos. É uma parte inevitável do mercado de trabalho, pois as pessoas podem estar procurando trabalho que melhor se adeque às suas habilidades ou preferências.
- Desemprego Estrutural: Este é o desemprego que resulta de mudanças na estrutura da economia. Por exemplo, mudanças tecnológicas podem tornar certas habilidades obsoletas, e trabalhadores podem precisar de tempo para adquirir novas habilidades ou encontrar novos setores onde suas habilidades sejam demandadas.

somente serão eficazes caso surpreendam os agentes.

Sob a hipótese de expectativas racionais, a suposição de que os eleitores são ingênuos, não sendo capazes de aprender e propensos a erros sistemáticos, torna-se insatisfatória.

Uma vez que o eleitor tenha presenciado um ciclo político-econômico, não será mais ludibriado por políticos oportunistas, pois sabe que um aumento da atividade econômica no período pré-eleitoral será seguido por uma elevação do nível de preços e recessão após as eleições.

Assim, é de se supor que ao invés de premiar com a reeleição o candidato que manipula a economia, o eleitor deve puni-lo votando no seu oponente.

Entre as características da Taxa Natural de Desemprego podemos citar:

- Não é zero: mesmo em uma economia saudável e eficiente, sempre haverá algum nível de desemprego devido às fricções e mudanças estruturais.
- Mudança ao longo do tempo: a taxa natural de desemprego pode variar ao longo do tempo devido a mudanças nas condições econômicas, políticas de emprego, estrutura educacional, e avanços tecnológicos.
- Não é influenciada pela política monetária no longo prazo: no longo prazo, a taxa natural de desemprego não é afetada pela política monetária, pois os esforços para reduzir o desemprego abaixo da taxa natural podem levar a inflação sem reduzir significativamente a taxa de desemprego.

Da importância da Taxa Natural de Desemprego:

- Política econômica: conhecer a taxa natural de desemprego ajuda os formuladores de políticas a entender quando a economia está operando em pleno emprego ou quando pode estar sobrecarregada.
- Expectativas de inflação: a taxa natural de desemprego está relacionada às expectativas de inflação. Uma taxa de desemprego muito abaixo da taxa natural pode levar a aumentos inflacionários, enquanto uma taxa muito acima pode indicar uma economia fraca.
- Análise econômica: economistas utilizam a taxa natural de desemprego para avaliar a eficácia das políticas de emprego e para fazer previsões sobre o futuro do mercado de trabalho.

Entre as implicações práticas podemos citar políticas que visam reduzir o desemprego friccional e estrutural, como melhoramento na formação profissional e na educação, podem ajudar a reduzir a taxa natural de desemprego. Também as expectativas e comportamento econômico:, isto é, mudanças na percepção da taxa natural de desemprego podem afetar decisões de consumo e investimento, bem como as políticas econômicas. Em resumo, a taxa natural de desemprego é um conceito crucial para entender a dinâmica do mercado de trabalho e a eficácia das políticas econômicas. Ela representa o nível de desemprego que se pode esperar na economia em condições normais e é um ponto de referência para analisar a saúde econômica e a eficácia das políticas governamentais.

A partir dos anos 80, diversos autores procuraram conciliar os modelos de ciclos político-econômicos tradicionais à hipótese de expectativas racionais.

Esses modelos oportunistas com expectativas racionais atribuem a existência de ciclos político-econômicos à assimetria de informação entre os eleitores e os políticos, quanto à competência desses no comando do governo.

Cada governante conduz a economia de forma mais ou menos competente.

Entretanto, a competência é uma informação privada, somente o próprio político sabe do seu potencial.

Os eleitores observam os resultados das variáveis macroeconômicas para avaliar a competência do governante.

Em virtude disso, os políticos tendem a agir oportunisticamente para parecer o mais competente possível a cada eleição, gerando os ciclos político-econômicos propostos por Nordhaus.

Vamos discutir a curva de Phillips num modelo de competência.

O modelo de Persson e Tabellini (1990) sugere que a existência dos ciclos político-econômicos seja o resultado de um problema de seleção adversa.

Em virtude da assimetria de informação entre os políticos e os eleitores, o político que estiver no poder procurará obter os melhores resultados macroeconômicos possíveis para sinalizar sua competência aos votantes, criando os ciclos econômicos.

Este modelo foi desenvolvido num ambiente keynesiano, no qual os preços não são totalmente flexíveis, possibilitando a existência de um *trade-off* de curto prazo entre inflação e desemprego.

O governante tenta mostrar-se competente buscando reduzir o desemprego além da sua taxa natural sem elevar o nível de preços através da política monetária, controlada diretamente pelos *policymakers*.

Contudo, nem todos os políticos terão sucesso nesta investida, apenas os verdadeiramente competentes.

Essas hipóteses podem ser sintetizadas pela equação da curva de Phillips

$$y_t = \bar{y} + \pi_t - \pi_t^e + \varepsilon_t \tag{3.27}$$

em que

- π_t é a taxa de inflação no período t
- π_t^e é a taxa de inflação esperada no período t
- $\bar{y}$ é a taxa de crescimento natural do produto t
- y_t é a taxa de crescimento do produto no período t
- ε_t é a medida de competência do governante no período t

Como se pressupõe que as expectativas sejam racionais, os agentes econômicos determinarão a taxa de inflação esperada para o período t a partir das informações disponíveis até o período $t-1(I_{t-1})$ conforme a equação

$$\pi_t^e = \mathbb{E}\left(\pi_t^e | I_{t-1}\right) \tag{3.28}$$

Verificamos que o componente da curva de Phillips de curto prazo que representava os choques exógenos na economia foi modificado para captar a competência do governante.

Através dessa equação, podemos observar que quanto mais competente o governante maior a taxa de crescimento do produto além da taxa natural, para dados níveis de inflação e inflação esperada.

Além do incentivo de permanecer no poder, os autores dotaram os políticos de certa dose de benevolência ao incluírem o bem-estar social na sua função objetivo.

Os governantes competentes terão maior preocupação com o resultado do pleito do que os demais, pois sabem que poderão fazer um melhor trabalho para a sociedade.

Persson e Tabellini (1990) postularam que a competência é persistente, ou seja, um governante não deixa de ser (in)competente de um ano para o outro.

Assim, definiram a competência do governante como um processo média móvel de ordem 1, MA (1), conforme a equação (3.29), enquanto que a competência do candidato de oposição foi normalizada em zero no ano da eleição

$$\varepsilon_t = \mu_t + \mu_{t-1} \tag{3.29}$$

Para simplificar o modelo a competência pode assumir dois valores

$$\mu_t = \bar{\mu} > 0 \quad \text{alta competência com probabilidade } \rho \tag{3.30}$$

$$\mu_t = \underline{\mu} < 0 \quad \text{baixa competência com probabilidade } 1 - \rho \tag{3.31}$$

O valor esperado de μ_t é, portanto,

$$\mathbb{E}(\mu_t) = \rho\bar{\mu} + (1 - \rho)\underline{\mu} = 0 \tag{3.32}$$

Mesmo sabendo a distribuição dos valores de μ_t e que o valor esperado de μ_t é zero, os eleitores desconhecem a competência do atual governante, que será verificada com defasagem de um período.

Para que isto seja possível, os autores impõem como condição necessária que no período t o eleitor observe apenas a taxa de crescimento do produto y_t, identificando a taxa de inflação π_t somente no período seguinte $t + 1$.

Os ciclos político-econômicos serão gerados por causa desta assimetria de informação.

Os governantes tentarão parecer competentes, estimulando a atividade econômica via política monetária, perto das eleições.

Os realmente competentes conseguirão elevar a taxa de crescimento do produto além do seu nível natural ($y_t > \bar{y}$), porém os incompetentes não.

Como os votantes não têm informação de qual o tipo de político está no poder, formam suas expectativas de inflação a partir da média da inflação do competente (elevada) e do incompetente (baixa), ponderada pela probabilidade de ambos ocorrerem.

Consequentemente, a inflação verificada no ano da eleição será acima das expectativas dos eleitores se o político for do tipo competente, e abaixo em caso contrário.

O modelo de Persson e Tabellini (1990) traz uma conclusão um tanto inconveniente.

Justamente os governantes mais competentes proporcionam maior instabilidade à economia.

Além disso, como os próprios autores sugerem, a proposição de que os eleitores conseguem verificar a taxa de crescimento do produto imediatamente, mas não a variação do nível de preços é difícil de ser aceita.

Os autores justificam que na realidade os *policymakers* não têm controle sobre a taxa de inflação, somente sobre os instrumentos de política monetária, raramente rastreados pelo eleitor médio.

Vamos analisar agora modelos de ciclos políticos orçamentários.

Os modelos apresentados até o presente momento deram um tratamento keynesiano aos ciclos político-econômicos, ao basearem suas análises em curvas de Phillips de curto prazo e na eficácia de políticas monetárias expansionistas.

Por discordarem destes undamentos, Rogoff e Sibert (1988) e Rogoff (1990) elaboraram propostas alternativas, sugerindo que sob a hipótese de expectativas racionais os ciclos político-econômicos devem ser observados nos instrumentos de política fiscal.

De acordo com os autores, durante anos de eleição os políticos tendem a reduzir os impostos, elevar os gastos públicos e transferir recursos para projetos mais perceptíveis (obras públicas, saúde e educação), criando o que denominaram de ciclos políticos orçamentários (*political budget cycles*).

Assim como no trabalho de Persson e Tabellini (1990), o modelo de ciclos políticos orçamentários desenvolvido por Rogoff (1990) é viabilizado pela existência de assimetria de informação entre políticos e eleitores.

Os cidadãos desconhecem a capacidade do governante em administrar eficientemente os recursos públicos.

Para sinalizar sua competência, em épocas de eleição, o governante procura prover a população com a maior quantidade de bens públicos possível, a fim de elevar o nível de utilidade dos eleitores.

A função de utilidade do votante representativo é

$$\Gamma_t = \sum_{s=t}^{T} \left[U(c_s, g_s) + V(k_s) + \eta_s \right] \beta^{s-t} \tag{3.33}$$

em que

- c_s é o consumo de bens privados no período s
- g_s é o bem de "consumo" público *per capita* no período s
- k_s é o bem "investimento" público no período s
- β é a taxa de desconto intertemporal, com $\beta < 1$
- η_s é o termo do choque aleatório, representado por um processo média móvel de ordem 1[29], MA(1), tal que $\eta_s = q_s + q_{s-1}$, com q sendo uma variável aleatória contínua

[29] Um processo de média móvel (MA) é um modelo estatístico amplamente utilizado em séries tem-

Pressupõe-se que no início de cada período todos os indivíduos recebam exogenamente y unidades de bens não estocáveis (perecíveis), que poderão ser consumidos privadamente pelos próprios cidadãos ou servir de insumo para a produção de bens públicos.

Como existem impostos *lump-sum* no período t (τ_t), a dotação inicial de cada indivíduo será

porais, onde o valor presente de uma série é explicado como uma função linear de erros passados (ruídos brancos). A ordem do processo indica quantos termos passados de erro (resíduos) são utilizados na formulação do modelo. Um processo de média móvel de ordem q, denotado por MA(q), é representado pela equação:

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

em que

Y_t é o valor da série no tempo t;

μ é a média da série (pode ser zero se a série for estacionária em torno de zero);

ϵ_t é o erro (ou ruído branco) no tempo t, que é uma variável aleatória com média zero e variância constante;

$\theta_1, \theta_2, \ldots, \theta_q$ são os parâmetros do modelo que multiplicam os erros passados.

Neste modelo, o valor atual da série (Y_t) depende dos q erros passados, além do erro atual. O número q indica quantos termos passados estão influenciando o valor atual da série. Um processo de média móvel de ordem 1, denotado por MA(1), é um caso particular do MA(q), onde apenas o erro imediatamente anterior influencia o valor presente da série. A equação do modelo MA(1) é:

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$$

em que

Y_t é o valor da série no tempo t;

μ é a média da série;

ϵ_t é o erro no tempo t;

θ_1 é o único parâmetro do modelo que multiplica o erro no tempo $t-1$.

Neste caso, o valor atual da série (Y_t) depende apenas do erro atual (ϵ_t) e do erro do período imediatamente anterior (ϵ_{t-1}). Esse tipo de modelo é útil para capturar correlações de curto prazo em séries temporais.

$$c_t = y - \tau_t \tag{3.34}$$

Além dos impostos *lump-sum* arrecadados, a produção de bens públicos depende da competência administrativa do governante, ε_t.

Um administrador competente produzirá um certo nível de bens públicos com menor tributação do que um incompetente, ou seja, competência é sinônimo de eficiência.

A função de produção de bens públicos toma a forma

$$g_t + k_{t+1} = \tau_t + \varepsilon_t \tag{3.35}$$

Observe que há uma diferença temporal na produção dos bens de "consumo" (g_t) e de "investimento" público (k_t).

Para ofertar bens de "investimento" público no período $t+1$, o governo necessita investir na sua produção no período t.

A competência do governante segue um processo MA (1)

$$\varepsilon_t = \alpha_t + \alpha_{t-1} \tag{3.36}$$

No qual as probabilidades de α seguem uma distribuição Bernoulli

$$\Pr\left(\alpha = \alpha^H\right) = \rho \tag{3.37}$$

$$\Pr\left(\alpha = \alpha^L\right) = 1 - \rho \tag{3.38}$$

em que $\alpha^H > \alpha^L > 0$.

A proposição de que a competência do governante varia ao longo do tempo é bastante realista.

O político pode aperfeiçoar a competência administrativa ao enfrentar crises durante o seu mandato, quando precisará desenvolver habilidades para ofertar a mesma quantidade de bens públicos com um menor quantidade de recursos.

Também poderíamos imaginar situações nas quais o governante “desaprende” a administrar, como no caso de um político honesto que devido à influência do seu meio de trabalho tornou-se corrupto, não resistindo a tentação de desviar recursos públicos.

A função de utilidade do governante é semelhante a dos demais indivíduos.

Entretanto, como o cargo de administrador público é uma honra para qualquer cidadão, somente o fato de estar no poder gera um acréscimo à utilidade do político, o que Rogoff e Sibert (1988) convencionaram chamar de *ego rents*.

A função de utilidade do governante é representada por

$$\Gamma_t^G = \sum_{s=t}^{T} \left[U(c_s, g_s) + V(k_s) + \eta_s\right] \beta^{s-t} + \sum_{s=t}^{T} \beta^{s-t} X \pi_{s,t} \tag{3.39}$$

em que

- Γ_t é a função de utilidade do indivíduo “comum” no período t, pois o governante também é um consumidor de bens públicos
- $\pi_{s,t}$ é a estimativa da probabilidade do governante estar no poder no período s
- X é o ganho de utilidade por estar no poder ou *ego rents*
- β é a taxa de desconto intertemporal, com $\beta < 1$

Como podemos constatar através da função de utilidade do governante, ele não possui qualquer dose de altruísmo.

A busca pela maximização da utilidade do político, implica a tentativa de aumentar sua probabilidade de se manter no poder, que será refletida na condução da política fiscal.

Os ciclos políticos orçamentários ocorrem porque o governo possui uma vantagem temporária sobre os eleitores em relação às informações concernentes a produção de bens públicos e, portanto, da competência do administrador.

No período t, os eleitores conseguem observar os impostos *lump sum* e os gastos do governo em bens de consumo, inferindo a partir das informações disponíveis a respeito dos gastos em investimento público e da competência do governante, observáveis apenas no período seguinte.

Esta defasagem de tempo acontece porque há um custo elevado para os indivíduos monitorarem e avaliarem as ações orçamentárias do governo, além de não se sentirem motivados para tomar tal atitude, já que podem estimar a competência do governante sem nenhum custo.

O eleitor representativo decide em qual candidato votar a partir da comparação das utilidades esperadas sob o governo de ambos, escolhendo aquele que lhe proporcionar maior benefício.

Considerando que $v = 1$ represente o voto para o candidato da situação e que $v = 0$ signifique o voto para a oposição, teremos

$$v_t = \begin{cases} 1, & \text{se } \mathbb{E}^G\left(\Gamma_{t+1}\right) \geq \mathbb{E}^O\left(\Gamma_{t+1}\right) \\ 0, & \text{caso contrário} \end{cases} \tag{3.40}$$

em que

- $\mathbb{E}^G\left(\Gamma_{t+1}\right)$ é a utilidade esperada no período $t+1$ sob o governo do candidato da situação
- $\mathbb{E}^O\left(\Gamma_{t+1}\right)$ é a utilidade esperada no período $t+1$ sob o governo do candidato

da oposição

Os resultados do modelo nos fazem concluir que num ambiente com assimetria de informação, apenas o governante competente conseguirá aumentar os gastos públicos e reduzir os impostos antes das eleições, sinalizando seu tipo.

O administrador incompetente não terá êxito na tentativa de mascarar a sua capacidade, adotando a política fiscal em condições de informação perfeita.

Assim, chegamos novamente a incomoda conclusão que apenas o governante competente gera os ciclos políticos orçamentários.

Rogoff (1990) sugere que os ciclos políticos orçamentários possam ser aliviados com a realização de modificações constitucionais que restrinjam o uso dos instrumentos de política fiscal nos anos de eleição, como a adoção de planos orçamentários bienais.

Outra proposta apresentada, diz respeito à mudança na estrutura eleitoral.

O autor propõe que as eleições não sejam estabelecidas exogenamente.

Ao invés disso, acredita que deva haver um sistema no qual o governante pode convocar as eleições antes do final do mandato, a exemplo do que acontece em países como Japão, Canadá e Reino Unido.

Com a inclusão de expectativas racionais aos modelos partidários tradicionais, a proposição de que diferentes partidos no poder escolhem os níveis de inflação e desemprego condizentes com as suas ideologias, já que a curva de Phillips de curto prazo é estável, não sobrevive.

Sob essa nova hipótese, as políticas econômicas adotadas somente terão efeitos reais sobre a economia se surpreenderem os agentes econômicos.

Ao reformular os modelos partidários tradicionais, Alesina (1987) partiu do princípio de que processos eleitorais justos têm por característica a incerteza quanto ao resultado da apuração.

Nos anos de eleição, os eleitores, mesmo que possuam expectativas racionais, não saberão ao certo qual partido estará no poder no próximo ano e, consequentemente, a política econômica a ser adotada.

Desta forma, qualquer que seja o resultado do pleito, os agentes serão surpreendidos.

O modelo elaborado por Alesina (1987) pressupõe que os salários sejam previamente estabelecidos por um período de um ano.

Assim, no período imediatamente anterior à eleição, os trabalhadores irão negociar seus salários com os empregadores observando as probabilidades de vitória dos dois partidos, esquerda ou direita.

O sistema econômico é caracterizado pela curva de Phillips com expectativas racionais

$$y_t = \bar{y} + \gamma(\pi_t - w_t) \tag{3.41}$$

em que w_t é a taxa de crescimento dos salários nominais no período t.

Os trabalhadores não sofrem de ilusão monetária e negociam o crescimento dos seus salários nominais conforme suas expectativas de inflação.

Então, o salário acordado para o próximo período será baseado nas informações disponíveis até o instante da assinatura do contrato.

Isto é

$$w_t = \pi_t^e = \mathbb{E}\left(\pi_t^e | I_{t-1}\right) \tag{3.42}$$

Assim,

$$y_t = \bar{y} + \gamma(\pi_t - \pi_t^e) \tag{3.43}$$

Conforme a equação (3.43), a taxa de crescimento do produto somente será diferente da sua taxa natural se as expectativas dos agentes não se confirmarem.

O autor representa as funções objetivo dos dois partidos através de funções de custo.

No caso do partido de esquerda (L), como sua prioridade é o crescimento do produto (y_t), este incorrerá num custo adicional (b') se houver uma redução do nível do produto.

Além disso, supõe-se que o partido estabeleça uma meta de inflação (c).

O custo será aumentado à medida que a taxa de inflação se afasta deste parâmetro.

Assim, a função de custo do partido de esquerda será

$$Z^L = \sum_{t=0}^{\infty} q^t \left[\frac{1}{2} (\pi_t - c)^2 - b' y_t \right], \quad b' > 0, c > 0, 0 < q < 1 \tag{3.44}$$

em que q é o fator de desconto intertemporal igual para ambos partidos.

Já como o partido de direita é avesso à inflação, não tem interesse em gerar surpresas inflacionárias, estabelecendo, assim, uma meta de inflação nula ($c = 0$).

A taxa de crescimento do produto não entra na sua função de custo, já que combater o desemprego não é uma prioridade do partido.

A função de custo do partido de direita é

$$Z^R = \sum_{t=0}^{\infty} q^t \left[\frac{1}{2} (\pi_t)^2 \right] \tag{3.45}$$

Para simplificar vamos assumir que $y = 0$ e substituir a equação (3.43) em (3.44)

$$Z^L = \sum_{t=0}^{\infty} q^t \left[\frac{1}{2} \pi_t^2 - \frac{1}{2} c^2 - c\pi_t - b' \gamma (\pi_t - \pi_t^e) \right] \tag{3.46}$$

Supondo que $b = b'\gamma$ e rescrevendo o somatório infinito de q^t como $\dfrac{1}{1-q}$, podemos manipular a equação (3.46) para que $Z^{L'} = Z^L - \left(\dfrac{c^2}{2(1-q)}\right)$.

Então, a função de custo do partido de esquerda fica transformada em

$$Z^{L'} = \sum_{t=0}^{\infty} q^t \left(\frac{1}{2}\pi_t^2 - b\left(\pi_t - \pi_t^e\right) - c\pi_t\right) \tag{3.47}$$

Alesina (1987) supõe que o governante tenha plenos poderes sobre os instrumentos de política macroeconômica, o que o possibilita escolher o nível de inflação que bem desejar.

Logo, o partido que saiu vitorioso da eleição colocará em prática, imediatamente após a posse, sua ideologia determinando a taxa de inflação de sua preferência, ou seja, que minimiza a sua função de custo.

A taxa de inflação esperada de cada partido é dada pela condição de primeira ordem da função de custo

$$\pi_t^L = b + c \tag{3.48}$$

$$\pi_t^R = 0 \tag{3.49}$$

Nos anos de eleição, como há a incerteza quanto ao futuro político, os trabalhadores irão acertar seus contratos salariais para o próximo período levando em consideração apenas as probabilidades de vitória de cada partido, perfeitamente observáveis através das pesquisas realizadas pelos órgãos especializados.

As informações sobre o desempenho da economia no passado serão totalmente irrelevantes.

Portanto, a taxa de crescimento do salário nominal antes das eleições, igual à taxa de inflação esperada, será dada pelas taxas de inflação esperadas dos partidos, ponderadas pelas chances de vitória de cada um

$$w_t = \pi_t^e = p\mathbb{E}\left(\pi_t^L\right) + (1-p)\mathbb{E}\left(\pi_t^R\right) = p(b+c), \quad 0 \leq p \leq 1 \tag{3.50}$$

em que onde p é a probabilidade de vitória do partido de esquerda e $1-p$ é a probabilidade de vitória do partido de direita.

Substituindo a taxa de inflação esperada pelos eleitores e a taxa de inflação que os partidos irão determinar caso sejam eleitos na equação (3.43), podemos estimar o nível do produto após a eleição sob a administração de ambos os partidos.

Assim, se o partido de esquerda (L) vencer teremos uma expansão do produto

$$y^L = \gamma\left(\pi_t^L - \pi_t^e\right) = \gamma(1-p)(b+c) \tag{3.51}$$

Por outro lado, caso o partido de direita (R) vença a eleição, haverá uma recessão

$$y^R = \gamma\left(\pi_t^R - \pi_t^e\right) = -\gamma p(b+c) \tag{3.52}$$

Paldam (1996): podemos concluir que quanto mais surpreendente for o resultado da eleição maior será a flutuação do produto na economia.

Digamos que antes da data da eleição as pesquisas apontem o candidato do partido de direita como vitorioso (p muito baixo).

Conhecendo a ideologia do partido, os agentes econômicos formarão uma expectativa de inflação baixa no ano seguinte.

Entretanto, o resultado das urnas contradiz as pesquisas e indica a vitória do candidato do partido de esquerda.

Consequentemente, o partido de esquerda conseguirá expandir o produto (y) e reduzir a taxa de desemprego (u) sem gerar um aumento elevado do nível de preços no primeiro ano de mandato.

Posteriormente, à medida que os agentes econômicos reformulam suas expectativas de inflação, a taxa de inflação (π) aumentará e a economia será reconduzida a sua taxa natural de desemprego.

Por outro lado, se o resultado da próxima eleição seguir a tendência das pesquisas e o partido de esquerda vencer (p elevado), a expectativa de inflação dos agentes econômicos será elevada e o impacto das políticas expansionistas sobre o produto será menor.

A importância da teoria dos ciclos político-econômicos pode ser comprovada pela tendência atual dos países em modificar seus arranjos institucionais a fim de "amarrar as mãos" dos governantes.

Ao longo das últimas décadas, diversos bancos centrais adquiriram independência legal para diminuir a pressão política sobre a condução da política monetária.

Neste sentido, conforme sugerem Giavazzi e Pagano (1988), outro fato importante foi a constituição do Sistema Monetário Europeu. Países que historicamente possuíam uma política monetária com viés inflacionário, como Itália e França, ao se filiarem ao sistema e se submeterem as regras por ele impostas, retomaram a credibilidade perdida.

Em relação à política fiscal, um exemplo disso é a recente elaboração da Lei de Responsabilidade Fiscal, que visa disciplinar os gastos dos governos estaduais no Brasil.

Resolução via Controle Ótimo

Queremos resolver o seguinte problema de controle ótimo intertemporal:

$$\max_{u_t} \quad W = \int_0^\infty g(u_t, \pi_t)\, e^{\rho t} dt \tag{3.53}$$

$$\text{sujeito a:} \quad \pi_t = f(u_t) + \lambda v_t \tag{3.54}$$

$$\frac{dv_t}{dt} = \gamma(\pi_t - v_t) = \gamma(f(u_t) + (\lambda - 1)v_t) \tag{3.55}$$

onde u_t é a taxa de desemprego, π_t é a inflação, v_t é a inflação esperada e $g(u, \pi)$ representa a função de bem-estar (ou votação agregada), com $\rho > 0$ como a taxa de desconto intertemporal.

Definimos a Hamiltoniana de valor corrente como:

$$\mathcal{H} = g(u, \pi) + \mu(t) \cdot \gamma(f(u) + (\lambda - 1)v) \tag{3.56}$$

onde $\mu(t)$ é o multiplicador de valor corrente associado à equação de estado.

2. Condições de Primeira Ordem

(a) Condição de ótimo para o controle u:

Sabendo que $\pi = f(u) + \lambda v$, temos:

$$\frac{\partial \mathcal{H}}{\partial u} = g_1(u, \pi) + g_2(u, \pi) \cdot f'(u) + \mu(t) \cdot \gamma f'(u) = 0 \tag{3.57}$$

com $g_1 = \frac{\partial g}{\partial u}$ e $g_2 = \frac{\partial g}{\partial \pi}$.

(b) Dinâmica do coestado:

$$\dot{\mu}(t) = \rho\mu(t) - \frac{\partial \mathcal{H}}{\partial v} \tag{3.58}$$

Sabendo que $\frac{\partial \pi}{\partial v} = \lambda$, temos:

$$\frac{\partial \mathcal{H}}{\partial v} = g_2 \cdot \lambda + \mu \cdot \gamma(\lambda - 1) \tag{3.59}$$

$$\Rightarrow \dot{\mu} = \rho\mu - g_2\lambda - \mu\gamma(\lambda - 1) \tag{3.60}$$

(c) Condição de transversalidade:

$$\lim_{t\to\infty} \mu(t)v(t) = 0$$

3. Solução de estado estacionário

Assumimos que no longo prazo as variáveis são constantes: $\dot{v} = 0$, $\dot{\mu} = 0$.

- Da equação (3.55), temos:

$$\dot{v} = 0 \Rightarrow f(u^*) + (\lambda - 1)v^* = 0 \Rightarrow v^* = -\frac{f(u^*)}{\lambda - 1}, \quad \text{se } \lambda \neq 1$$

- Da equação da coestado:

$$0 = \rho\mu - g_2\lambda - \mu\gamma(\lambda - 1) \Rightarrow \mu = \frac{g_2\lambda}{\rho + \gamma(1 - \lambda)}$$

- Substituímos na FOC:

$$g_1 + g_2 f'(u) + \mu\gamma f'(u) = 0$$

$$\Rightarrow g_1 + f'(u)\,[g_2 + \mu\gamma] = 0$$

Substituindo o valor de μ:

$$f'(u) = -\frac{g_1}{g_2 + \frac{g_2\lambda\gamma}{\rho+\gamma(1-\lambda)}} = -\frac{g_1}{g_2} \cdot \left(\frac{1}{1 + \frac{\lambda\gamma}{\rho+\gamma(1-\lambda)}} \right)$$

Essa expressão pode ser reorganizada como:

$$\frac{f'(u)}{1-\lambda} = -\frac{g_1}{g_2} \cdot \left(\frac{\rho + \gamma(1-\lambda)}{(\rho+\gamma)(1-\lambda)} \right), \quad \text{se } 0 \leq \lambda < 1$$

$$f'(u) = -\frac{g_1}{g_2} \cdot \left(\frac{\rho}{\rho+\gamma} \right), \quad \text{se } \lambda = 1$$

4. Interpretação econômica

- $f'(u)$ é a inclinação da curva de Phillips de curto prazo.
- g_1 e g_2 representam as preferências sociais sobre desemprego e inflação, respectivamente.
- O termo com λ mostra o papel das expectativas: quanto maior λ, mais as expectativas se ajustam, o que afeta o trade-off inflação-desemprego.

4 Externalidades

[...] anything that has been so important and has survived for so long is too valuable to be put aside. Every effort should be made to preserve its centrality, even if that means revising the way one thinks about the concept in light of recent challenges.

Louis Makowsky e Joseph Ostroy, 2001

4.1 Introdução

Uma externalidade surge sempre que a utilidade ou possibilidade de produção de um agente depende diretamente das ações de outro agente (firma ou indivíduo). Dito de outra forma, significa que o efeito não é transmitido através dos preços (ou seja, através de um mecanismo de mercado).

Pigou, em *The Economics of Welfare*, define externalidades como:

> *Here the essence of the matter is that one person A, in the course of rendering some service, for which payment is made, to a second person B, incidentally also renders services or disservices to other persons (not producers of like services), of such a sort that payment cannot be extracted from the benefited parties or compensation enforced on behalf of the injured parties* (p. 159).

A definição de externalidades pecuniárias versus não-pecuniárias não é uma definição tecnológica exógena. Depende fundamentalmente dos mercados que estão em vigor. A presença de externalidades depende dos detalhes do arranjo como definição de mercadorias e direitos de propriedade. Exemplo: uma firma polui o rio e a outra firma é uma fazenda de peixes naquele rio que sofre poluição da primeira firma. Se as duas firmas se fundem ou se uma é dona do rio e pode cobrar a outra pela poluição, então o efeito é internalizado e não há mais uma externalidade.

A velha escola de Chicago (Coase) afirma que é possível converter todas as externalidades em externalidades pecuniárias com mercados apropriados. Aqui há uma conexão com desenho de mecanismo. Até o momento estivemos tomando o arranjo institucional onde ocorrem as transações como um dado: o mercado. Note, porém, que mesmo nesse caso, permitíamos que algumas transações ocorressem dentro de outros arranjos institucionais, como é o caso do que ocorre no interior das firmas. A teoria de desenho de mecanismos permite um arcabouço teórico capaz de tratar de forma unificada e coerente todos esses temas. A teoria de desenho de mecanismos define instituição como jogos não cooperativos e compara essas instituições em termos dos resultados do equilíbrio desses jogos, permitindo ao cientista social analisar performance de cada arranjo institucional relativamente ao ótimo teórico. A teoria de desenho de mecanismos começa com Hurwicz (1960). De acordo com Hurwicz, um mecanismo é um sistema de comunicação em que os participantes trocam mensagens uns com os outros de forma a determinar uma escolha social. Essas mensagens podem comunicar informações privadas dos agentes, como por exemplo, a sua disposição a pagar por um bem público. O mecanismo age, então, como um computador central que compila e processa as mensagens dos agentes, agregando as informações disponíveis na economia e recomendando uma ação a partir do conjunto de informações recebidas de acordo com um conjunto pre-especificado de regras que mapeiam mensagens em escolhas.

Inicialmente, a maior parte da investigação tinha por foco a complexidade (medido pelos custos informacional e computacional dos mecanismos), desconsiderando os aspectos de incentivos. A teoria, porém, só tornou-se relevante na prática depois de mais um artigo de Hurwicz (1972) em que os incentivos dos participantes eram incorporados de forma tratável com a noção de compatibilidade em incentivos. Isto permitiu uma análise rigorosa do fato fundamental de as pessoas só revelarem as suas informações privadas quando lhes for conveniente. Em outras palavras, como os indivíduos otimizam conhecendo as regras que levam o conjunto de mensagens às escolhas sociais, só vão revelar as informações que lhes convier. Esta observação leva à noção de implementação (no sentido fraco) de resultados como um equilíbrio do jogo de mensagens, em que o mecanismo é o conjunto de regras governando esse jogo de mensagens. Mecanismos podem, então ser comparados pela comparação dos equilíbrios induzidos por cada um desses jogos. Note a complexidade do problema de determinar o desenho institucional ótimo. Temos que descobrir qual jogo de mensagens tem o melhor equilíbrio dentre todos os jogos de mensagens concebíveis. Um avanço fundamental ocorrido na década de 1970 para tornar o problema tratável foi a formulação do princípio da revelação; uma ideia simples que permite uma simplificação enorme do problema de desenho de mecanismo. Formulado inicialmente por Gibbard (1971) e em sua maior generalidade por Myerson (1986), o princípio estabelece que, em sua busca do melhor mecanismo possível, o pesquisador pode restringir-se aos chamados mecanismos reveladores diretos verdadeiros (ou compatíveis em incentivos). A estrutura matemática do problema de desenho de mecanismo reformulado pela aplicação do princípio da revelação é relativamente simples. De fato, o problema de otimização no espaço de mecanismos diretos reveladores é um problema matemático bem definido, o que permite que se resolva um problema que de outra forma não seria passível de solução prática: o problema de encontrar o desenho institucional ótimo.

Note a conexão com a teoria de bens públicos. Bens públicos são bens que têm externalidades produtivas em larga escala.

Há dois tipos de externalidades:

1. Externalidades no consumo

$$\text{Sem preferência por externalidades} \implies u_i(x_i) \tag{4.1}$$

$$\text{Com preferência por externalidades} \implies u_i(x_1, \dots, x_i, \dots, x_n) \tag{4.2}$$

   ou seja, o consumo de um indivíduo afeta a utilidade de outros.

2. Externalidades na produção

   A função de produção inclui outros argumentos do que somente os insumos.

Isso leva a um exame de várias sugestões/maneiras alternativas de alocar recursos que podem levar a resultados eficientes. Alcançar uma alocação eficiente na presença de externalidades envolve, essencialmente, garantir que os agentes enfrentem os preços corretos por suas ações. As formas de resolver problemas de externalidade incluem tributação, regulação, direitos de propriedade, fusões etc.

Exemplos de externalidades negativas de produção incluem:

- Poluição do ar pela queima de combustíveis fósseis. Essa atividade causa danos às plantações, edifícios (históricos) e saúde pública.

- O clima muda como consequência das emissões de gases do efeito estufa da queima de petróleo, gás e carvão. As mudanças climáticas apresentam um desafio único para a economia: é o maior exemplo de falha de mercado que já vimos.

- Poluição da água por indústrias que adicionam efluentes, que prejudicam plantas, animais e seres humanos.

- Poluição sonora durante o processo de produção, que pode ser mental e psicologicamente perturbadora.

- Risco sistêmico: os riscos para a economia em geral decorrentes dos riscos assumidos pelo sistema bancário. Uma condição de risco moral pode ocorrer na ausência de regulamentação bancária bem projetada ou na presença de regulamentação mal projetada.

- Efeitos negativos da produção industrial de animais de criação, incluindo o aumento do *pool* de bactérias resistentes a antibióticos por causa do uso excessivo de antibióticos; problemas de qualidade do ar; contaminação de rios, córregos e águas costeiras com resíduos concentrados de animais; problemas de bem-estar animal.

- O esgotamento do estoque de peixes no oceano devido à sobrepesca. Este é um exemplo de um recurso de propriedade comum, vulnerável à tragédia dos bens comuns na ausência de uma governança ambiental apropriada.

- O armazenamento de resíduos nucleares de usinas nucleares.

Exemplos de externalidades negativas de consumo incluem:

- Privação do sono devido ao vizinho ouvir música alta tarde da noite.

- Resistência a antibióticos, causada pelo aumento do uso de antibióticos. Os indivíduos não consideram esse custo de eficácia ao tomar decisões de uso. As políticas do governo propostas para preservar a eficácia futura dos antibióticos incluem campanhas educacionais, regulamentação, impostos pigouvianos e patentes.

- Custos compartilhados de saúde causados pelo fumo e/ou abuso de álcool. Aqui, o "custo" é o de proporcionar bem-estar social mínimo. Os economistas atribuem com mais frequência esse problema à categoria de riscos morais, a perspectiva de que as partes isoladas do risco possam se comportar de maneira diferente da que teriam se estivessem totalmente expostas ao risco. Por exemplo, indivíduos com seguro contra roubo de automóveis podem ser menos vigilantes ao trancar seus carros, porque as consequências negativas do roubo de automóvel são (parcialmente) suportadas pela companhia de seguros.

- Maiores custos de congestionamento e maiores riscos de acidentes quando as pessoas usam as vias públicas.

- O consumo de um consumidor faz com que os preços subam e, portanto, piora outros consumidores, talvez reduzindo seu consumo. Esses efeitos são chamados de "externalidades pecuniárias" e se distinguem de "externalidades reais" ou "externalidades tecnológicas". As externalidades pecuniárias parecem ser externalidades, mas ocorrem dentro do mecanismo de mercado e não são consideradas uma fonte de falha ou ineficiência de mercado, embora ainda possam resultar em danos substanciais a terceiros.

Exemplos de externalidades positivas de produção incluem:

- Um apicultor que guarda as abelhas pelo mel. Um efeito colateral ou externalidade associado a essa atividade é a polinização das culturas vizinhas pelas abelhas. O valor gerado pela polinização pode ser mais importante que o valor do mel colhido.

- A construção e operação de um aeroporto. Isso beneficiará as empresas locais, devido ao aumento da acessibilidade.
- Uma empresa industrial que oferece aulas de primeiros socorros para que os funcionários aumentem a segurança do emprego. Isso também pode salvar vidas fora da fábrica.
- Uma empresa estrangeira que apresenta tecnologias atualizadas para empresas locais e melhora sua produtividade.

Exemplos de externalidades positivas de consumo incluem:

- Um indivíduo que mantém uma casa atraente pode conferir benefícios aos vizinhos na forma de maiores valores de mercado para suas propriedades.
- Um indivíduo que recebe uma vacinação contra uma doença transmissível não apenas diminui a probabilidade de infecção do próprio indivíduo, mas também diminui a probabilidade de outros serem infectados pelo contato com o indivíduo.
- Dirigir um veículo elétrico, reduzindo as emissões de gases de efeito estufa e melhorando a qualidade do ar local e a saúde pública. Embora isso possa aumentar as emissões de usinas de energia que queimam combustíveis fósseis, isso geralmente é mais do que compensado pelas emissões reduzidas de veículos, especialmente onde as fontes hidrelétricas, nucleares e renováveis são predominantes.
- Maior educação das pessoas, pois isso pode levar a benefícios mais amplos da sociedade, na forma de maior produtividade econômica, menor taxa de desemprego, maior mobilidade das famílias e maiores taxas de participação política.

- Um indivíduo que compra um produto que está interconectado em uma rede (*smartphone*). Isso aumentará a utilidade desses telefones para outras pessoas que possuem um celular com vídeo. Quando cada novo usuário de um produto aumenta o valor do mesmo produto de propriedade de outros, o fenômeno é chamado de externalidade da rede ou efeito de rede. As externalidades de rede geralmente têm "pontos de inflexão", onde, de repente, o produto atinge aceitação geral e uso quase universal.

- Em uma área que não possui corpo de bombeiros; os proprietários que adquirem serviços de proteção contra incêndios particulares fornecem uma externalidade positiva às propriedades vizinhas, que têm menos risco de o fogo do vizinho protegido se espalhar para sua casa (desprotegida).

4.2 Externalidades no Consumo

Quando não há externalidades no consumo, a função de utilidade do agente i é uma função de apenas seu próprio consumo:

$$u_i(x_i, y_i) \tag{4.3}$$

Neste caso, supondo dois agentes (A e B), as condições de primeira ordem para o equilíbrio competitivo são dadas por

$$TMgS^A_{x,y} = \frac{p_x}{p_y} = TMgS^B_{x,y} \tag{4.4}$$

e as condições de primeira ordem para a eficiência de Pareto são dadas por:

$$TMgS^A_{x,y} = TMgS^B_{x,y} \tag{4.5}$$

Assim, por causa do comportamento de tomador de preços, todo equilíbrio competitivo implica eficiência de Pareto se as funções de utilidade forem quase-côncavas.

O objetivo principal desta seção é mostrar que uma alocação de equilíbrio competitivo não é, em geral, eficiente quando existe uma externalidade no consumo. Mostramos isso observando que as condições de primeira ordem para um equilíbrio competitivo não são em geral as mesmas condições de primeira ordem para alocações eficientes de Pareto na presença de externalidades no consumo. Aqui estamos seguindo Tian e Yang (2009).

Considere a seguinte economia de troca com dois agentes e dois bens:

$$u_A(x_A, x_B, y_A) \tag{4.6}$$

$$u_B(x_A, x_B, y_B) \tag{4.7}$$

que são consideradas estritamente crescentes em seu próprio consumo de bens, quase-côncavas, e satisfazem as condições de Inada, $\dfrac{\partial u(\cdot)}{\partial x_i}(0) = +\infty$ e $\lim\limits_{x_i \to 0} \dfrac{\partial u(\cdot)}{\partial x_i} x_i = 0$, resultando em soluções interiores. Além disso, assumimos que o gradiente[30] de $u_i(\cdot)$ é diferente de zero nas alocações eficientes de Pareto. Aqui a externalidade recai sob o bem x.

As condições de primeira ordem para o equilíbrio competitivo são as mesmas

$$TMgS^A_{x,y} = \frac{p_x}{p_y} = TMgS^B_{x,y} \tag{4.8}$$

Agora encontramos as condições de primeira ordem para alocações eficientes de Pareto em economias de troca com externalidades. Assim, alocações eficientes de

[30] O gradiente de uma função escalar f de várias variáveis é um vetor que aponta na direção de máxima taxa de variação de f e cuja magnitude é a taxa de variação nessa direção. Seja $f : \mathbb{R} \to \mathbb{R}$ uma função escalar de n variáveis, $f(x_1, x_2, \dots, x_n)$. O gradiente de f, denotado por ∇f é um vetor definido como $\nabla f = \left(\dfrac{\partial f}{\partial x_1}, \dfrac{\partial f}{\partial x_2}, \dots, \dfrac{\partial f}{\partial x_n}\right)$, em que $\dfrac{\partial f}{\partial x_i}$ é a derivada parcial de f em relação à variável x_i. O vetor gradiente aponta na direção em que a função f aumenta mais rapidamente. Se você se mover na direção do gradiente a partir de um ponto x, a função f aumentará mais rapidamente do que se você se mover em qualquer outra direção. A magnitude do gradiente, $\|\nabla f\|$, representa a taxa de variação máxima de f naquela direção. Uma grande magnitude indica uma variação rápida da função, enquanto uma magnitude pequena indica uma variação lenta. Por fim, o gradiente é perpendicular às linhas de nível (ou superfícies de nível) da função f. As linhas de nível são as curvas (ou superfícies) ao longo das quais a função f tem um valor constante.

Pareto x^* podem ser completamente determinadas pelas condições de primeira ordem do seguinte problema de otimização é

$$
\begin{aligned}
& \max_{x \in \mathbb{R}_{++}} u_B(x_A, x_B, y_B) && (4.9)\\
\text{sujeito a} \quad & x_A + x_B \leq w_x && (4.10)\\
& y_A + y_B \leq w_y && (4.11)\\
& u_A(x_A, x_B, y_A) \geq u_A(x_A^*, x_B^*, y_A^*) && (4.12)
\end{aligned}
$$

O lagrangeano é

$$
\begin{aligned}
L(x_A, y_A, x_B, y_B, \lambda_x, \lambda_y, \mu) = {} & u_B(x_A, x_B, y_B) + \lambda_x(w_x - x_A - x_B)\\
& + \lambda_y(w_y - y_A - y_B)\\
& + \mu\left[u_A(x_A, x_B, y_A) - u_A(x_A^*, x_B^*, y_A^*)\right] \qquad (4.13)
\end{aligned}
$$

As condições de primeira ordem são:

$$
\begin{aligned}
\frac{\partial L}{\partial x_A} = 0 \quad & \Longleftrightarrow & \frac{\partial u_B}{\partial x_A} - \lambda_x + \mu \frac{\partial u_A}{\partial x_A} = 0 \qquad (4.14)\\
\frac{\partial L}{\partial y_A} = 0 \quad & \Longleftrightarrow & -\lambda_y + \mu \frac{\partial u_A}{\partial y_A} = 0 \qquad (4.15)\\
\frac{\partial L}{\partial x_B} = 0 \quad & \Longleftrightarrow & \frac{\partial u_B}{\partial x_B} - \lambda_x + \mu \frac{\partial u_A}{\partial x_B} = 0 \qquad (4.16)\\
\frac{\partial L}{\partial y_B} = 0 \quad & \Longleftrightarrow & \frac{\partial u_B}{\partial y_B} - \lambda_y = 0 \qquad (4.17)\\
\frac{\partial L}{\partial \lambda_x} = 0 \quad & \Longleftrightarrow & w_x - x_A - x_B \geq 0 \qquad (4.18)\\
& & \lambda_x \geq 0 \qquad (4.19)\\
& & \lambda_x(w_x - x_A - x_B) = 0 \qquad (4.20)\\
\frac{\partial L}{\partial \lambda_y} = 0 \quad & \Longleftrightarrow & w_y - y_A - y_B \geq 0 \qquad (4.21)
\end{aligned}
$$

$$\lambda_y \geq 0 \tag{4.22}$$

$$\lambda_y(w_y - y_A - y_B) = 0 \tag{4.23}$$

$$\frac{\partial L}{\partial \mu} = 0 \quad \Longleftrightarrow \quad u_A(x_A, x_B, y_A) - u_A(x_A^*, x_B^*, y_A^*) \geq 0 \tag{4.24}$$

$$\mu \geq 0 \tag{4.25}$$

$$\mu\left[u_A(x_A, x_B, y_A) - u_A(x_A^*, x_B^*, y_A^*)\right] = 0 \tag{4.26}$$

Por (4.17), temos que $\lambda_y = \dfrac{\partial u_B}{\partial y_B} > 0$ e, portanto, por (4.23),

$$y_A + y_B = w_y \tag{4.27}$$

o que significa que nunca há dilapidação do bem que não exibe uma externalidade negativa.

De (4.15) e (4.17), temos:

$$\begin{aligned}
-\lambda_y + \mu\frac{\partial u_A}{\partial y_A} &= 0 \\
-\frac{\partial u_B}{\partial y_B} + \mu\frac{\partial u_A}{\partial y_A} &= 0 \\
\mu\frac{\partial u_A}{\partial y_A} &= \frac{\partial u_B}{\partial y_B} \\
\mu &= \frac{\dfrac{\partial u_B}{\partial y_B}}{\dfrac{\partial u_A}{\partial y_A}}
\end{aligned} \tag{4.28}$$

Então, por (4.14) e (4.15), chegamos a:

$$\begin{aligned}
\frac{\partial u_B}{\partial x_A} - \lambda_x + \mu\frac{\partial u_A}{\partial x_A} &= 0 \\
\frac{\partial u_B}{\partial x_A} - \lambda_x + \frac{\lambda_y}{\dfrac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_A} &= 0
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{\lambda_y}\frac{\partial u_B}{\partial x_A} - \frac{1}{\lambda_y}\lambda_x + \frac{1}{\lambda_y}\frac{\lambda_y}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_A} = 0 \\
\frac{1}{\lambda_y}\frac{\partial u_B}{\partial x_A} - \frac{\lambda_x}{\lambda_y} + \frac{1}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_A} = 0 \\
-\frac{\lambda_x}{\lambda_y} &= -\frac{1}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_A} - \frac{1}{\lambda_y}\frac{\partial u_B}{\partial x_A} \\
-\frac{\lambda_x}{\lambda_y} &= -\frac{1}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_A} - \frac{1}{\frac{\partial u_B}{\partial y_B}}\frac{\partial u_B}{\partial x_A} \\
\frac{\lambda_x}{\lambda_y} &= \frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} \qquad (4.29)
\end{aligned}
$$

E por (4.15), (4.16) e (4.17), chegamos a:

$$
\begin{aligned}
\frac{\partial u_B}{\partial x_B} - \lambda_x + \mu\frac{\partial u_A}{\partial x_B} = 0 \\
\frac{\partial u_B}{\partial x_B} - \lambda_x + \frac{\lambda_y}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_B} = 0 \\
\frac{\partial u_B}{\partial x_B} - \lambda_x + \frac{\frac{\partial u_B}{\partial y_B}}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_B} = 0 \\
-\lambda_x &= -\frac{\partial u_B}{\partial x_B} - \frac{\frac{\partial u_B}{\partial y_B}}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_B} \\
-\frac{\lambda_x}{\lambda_y} &= -\frac{1}{\lambda_y}\frac{\partial u_B}{\partial x_B} - \frac{1}{\lambda_y}\frac{\frac{\partial u_B}{\partial y_B}}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_B}
\end{aligned}
$$

$$
-\frac{\lambda_x}{\lambda_y} = -\frac{1}{\frac{\partial u_B}{\partial y_B}}\frac{\partial u_B}{\partial x_B} - \frac{1}{\frac{\partial u_B}{\partial y_B}}\frac{\frac{\partial u_B}{\partial y_B}}{\frac{\partial u_A}{\partial y_A}}\frac{\partial u_A}{\partial x_B}
$$

$$
\frac{\lambda_x}{\lambda_y} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \tag{4.30}
$$

Igualando (4.29) e (4.30), obtemos:

$$
\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \tag{4.31}
$$

que expressa a igualdade das taxas marginais sociais de substituição para os dois consumidores em pontos eficientes de Pareto. Assim, temos imediatamente a seguinte conclusão.

Uma alocação de equilíbrio competitivo pode não ser um ótimo de Pareto porque as condições de primeira ordem para o equilíbrio competitivo e a otimização de Pareto não são as mesmas. A partir da condição de igualdade marginal acima, sabemos que, para avaliar as taxas marginais de substituição relevantes para as condições de otimalidade, devemos levar em conta os efeitos diretos e indiretos das atividades de consumo na presença de externalidades. Ou seja, para atingir a otimização de Pareto, quando um consumidor aumenta o seu consumo do bem x, não apenas o consumo de y precisa mudar, mas o consumo do bem y pelo outro consumidor também deve ser alterado. Portanto, a taxa marginal social de substituição do bem x pelo bem y para o consumidor i é igual a $\frac{\frac{\partial u_i}{\partial x_i}}{\frac{\partial u_i}{\partial y_i}} + \frac{\frac{\partial u_j}{\partial x_i}}{\frac{\partial u_j}{\partial y_j}}$.

Resolvendo (4.14) e (4.16) para μ e λ_x, obtemos:

$$\begin{aligned}
\frac{\partial u_B}{\partial x_A} + \mu\frac{\partial u_A}{\partial x_A} &= \frac{\partial u_B}{\partial x_B} + \mu\frac{\partial u_A}{\partial x_B} \\
\mu\frac{\partial u_A}{\partial x_A} - \mu\frac{\partial u_A}{\partial x_B} &= \frac{\partial u_B}{\partial x_B} - \frac{\partial u_B}{\partial x_A} \\
\mu\left(\frac{\partial u_A}{\partial x_A} - \frac{\partial u_A}{\partial x_B}\right) &= \frac{\partial u_B}{\partial x_B} - \frac{\partial u_B}{\partial x_A} \\
\mu &= \frac{\dfrac{\partial u_B}{\partial x_B} - \dfrac{\partial u_B}{\partial x_A}}{\dfrac{\partial u_A}{\partial x_A} - \dfrac{\partial u_A}{\partial x_B}} > 0 \qquad (4.32)
\end{aligned}$$

e

$$\begin{aligned}
\frac{\lambda_x}{\dfrac{\partial u_A}{\partial x_A}} - \frac{\dfrac{\partial u_B}{\partial x_A}}{\dfrac{\partial u_A}{\partial x_A}} &= \frac{\lambda_x}{\dfrac{\partial u_A}{\partial x_B}} - \frac{\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_A}{\partial x_B}} \\
\frac{\lambda_x}{\dfrac{\partial u_A}{\partial x_A}} - \frac{\lambda_x}{\dfrac{\partial u_A}{\partial x_B}} &= \frac{\dfrac{\partial u_B}{\partial x_A}}{\dfrac{\partial u_A}{\partial x_A}} - \frac{\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_A}{\partial x_B}} \\
\lambda_x\left(\frac{1}{\dfrac{\partial u_A}{\partial x_A}} - \frac{1}{\dfrac{\partial u_A}{\partial x_B}}\right) &= \frac{\dfrac{\partial u_B}{\partial x_A}}{\dfrac{\partial u_A}{\partial x_A}} - \frac{\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_A}{\partial x_B}} \\
\lambda_x\left(\frac{\dfrac{\partial u_A}{\partial x_A} - \dfrac{\partial u_A}{\partial x_B}}{\dfrac{\partial u_A}{\partial x_A}\dfrac{\partial u_A}{\partial x_B}}\right) &= \frac{\dfrac{\partial u_B}{\partial x_A}\dfrac{\partial u_A}{\partial x_B} - \dfrac{\partial u_A}{\partial x_A}\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_A}{\partial x_A}\dfrac{\partial u_A}{\partial x_B}} \\
\lambda_x &= \frac{\dfrac{\partial u_A}{\partial x_B}\dfrac{\partial u_B}{\partial x_A} - \dfrac{\partial u_A}{\partial x_A}\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_A}{\partial x_A} - \dfrac{\partial u_A}{\partial x_B}} \qquad (4.33)
\end{aligned}$$

Quando a externalidade no consumo é positiva, a partir de (4.29) ou (4.30), podemos facilmente ver que λ_x é sempre positivo, dado que $\lambda_y = \dfrac{\partial u_B}{\partial y_B} > 0$. Além disso, quando não existe externalidade ou uma externalidade unilateral (somente um consumidor impõe uma externalidade sobre o outro), por (4.29) ou (4.30), λ_x é positivo. Assim, a condição de igualdade marginal (4.31) e as condições de ajustamento determinam completamente todas as alocações eficientes de Pareto para esses casos. No entanto, quando há externalidades negativas para ambos os consumidores, o multiplicador de Kuhn-Tucker λ_x dado diretamente por (4.33) ou indiretamente dado por (4.29) ou (4.30) é a soma de um termo negativo e positivo e, portanto, o sinal de λ_x pode ser indeterminado.

Para garantir que uma alocação é Pareto-eficiente na presença de externalidades negativas, devemos exigir $\lambda_x \geq 0$ em pontos eficientes, o que, por sua vez, requer que as taxas marginais de substituição sejam não-negativas, isto é,

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \geq 0 \tag{4.34}$$

e

$$\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} \geq \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} \tag{4.35}$$

isto é, o benefício marginal conjunto deve ser maior ou igual ao custo marginal conjunto para todos os pontos Pareto-eficientes.

Para consumir os bens eficientemente, uma condição necessária é que o benefício marginal conjunto de consumir o bem x não seja menor do que o custo marginal de consumir o bem x.

Assim, as seguintes condições

$$\begin{cases} \dfrac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \dfrac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \dfrac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \dfrac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \geq 0 \\ y_A + y_B = w_y \\ x_A + x_B \leq w_x \\ \left(\dfrac{\partial u_A}{\partial x_A}\dfrac{\partial u_B}{\partial x_B} - \dfrac{\partial u_A}{\partial x_B}\dfrac{\partial u_B}{\partial x_A}\right)(w_x - x_A - x_B) = 0 \end{cases} \tag{4.36}$$

constituem um sistema que permite obter todas as alocações eficientes de Pareto. Podemos fazer isso considerando três casos.

1. Quando $\dfrac{\partial u_A}{\partial x_A}\dfrac{\partial u_B}{\partial x_B} > \dfrac{\partial u_A}{\partial x_B}\dfrac{\partial u_B}{\partial x_A}$ ou $\dfrac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \dfrac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \dfrac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \dfrac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} > 0$, $\lambda_x > 0$ e as duas últimas condições no sistema se reduzem a $x_A + x_B = w_x$. Neste caso não há dilapidação. Substituindo $x_A + x_B = w_x$ e $y_A + y_B = w_y$ na condição de igualdade marginal (4.31), isso nos daria uma relação entre x_A e y_A, que define exatamente as alocações eficientes de Pareto.

2. Quando o benefício marginal conjunto é igual ao custo marginal conjunto:

$$\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} = \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} \tag{4.37}$$

então,

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} = 0 \tag{4.38}$$

e, portanto, $\lambda_x = 0$. Neste caso, quando $x_A + x_B \leq w_x$, a depleção é indeterminada. No entanto, mesmo quando a depleção ocorre, ainda podemos determinar o conjunto de alocações eficientes de Pareto usando $y_A + y_B = w_y$ e as condições (4.38). De fato, depois de substituir $y_A + y_B = w_y$ em (4.38), podemos resolver para x_A em termos y_A.

3. Quando $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} < \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}$, para quaisquer alocações que satisfaçam $x_A + x_B = w_x$, $y_A + y_B = w_y$ e a condição de igualdade marginal (4.31), as taxas marginais sociais de substituição devem ser negativas. Assim, a alocação não será Pareto eficiente. Nesse caso, deve haver uma depleção do bem x para atingir a eficiência de Pareto e uma alocação eficiente de Pareto irá satisfazer (4.38).

Resumindo, temos a seguinte proposição que fornece condições suficientes para caracterizar se deve haver ou não dilapidação da dotação w_x na obtenção de alocações eficientes de Pareto.

Se $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} > \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}$, as condições suficientes para as alocações Pareto-ótima são:

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \tag{4.39}$$

$$x_A + x_B = w_x \tag{4.40}$$

$$y_A + y_B = w_y \tag{4.41}$$

Se $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} < \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}$, as condições suficientes para as alocações Pareto-ótima são:

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \tag{4.42}$$

$$x_A + x_B \leq w_x \tag{4.43}$$

$$y_A + y_B = w_y \tag{4.44}$$

A questão da depleção não é apenas importante na teoria, mas também é relevante para a realidade. Pode ser usado para explicar um enigma conhecido da relação felicidade-renda nas literaturas econômica e psicológica: a felicidade aumenta com a renda até certo ponto, mas não além dela. Por exemplo, o bem-estar diminuiu nos últimos 25 anos nos EUA, e a satisfação com a vida ficou praticamente estável ao mesmo tempo na Grã-Bretanha. Se interpretarmos a renda como um bem, e se os consumidores se invejarem em termos de níveis de consumo, pelo nosso resultado, quando a renda atingir um determinado nível, talvez seja necessário dispor livremente da riqueza para obter alocações eficientes em Pareto; caso contrário, as alocações resultantes serão Pareto ineficientes. Portanto, o crescimento econômico não aumenta o bem-estar indexado por nenhuma função de bem-estar social, uma vez atingido o nível crítico de renda.

Exemplo 4.1. *Seja a seguinte função de utilidade:*

$$u_i(x_A, x_B, y_i) = \sqrt{x_i y_i} - x_j, \quad i \in \{A, B\}, j \in \{A, B\}, j \neq i \tag{4.45}$$

Pela condição (4.31), temos que

$$\frac{y_A}{x_A} = \frac{y_B}{x_B} \tag{4.46}$$

Vamos supor que $x_A + x_B \equiv \overline{x}$. Substituindo $\overline{x}$ e $y_A + y_B = w_y$ na condição acima, obtemos:

$$\frac{y_A}{x_A} = \frac{y_B}{x_B}$$
$$\frac{y_A}{x_A} = \frac{w_y - y_A}{\overline{x} - x_A}$$
$$y_A\overline{x} - y_A x_A = w_y x_A - y_A x_A$$

$$\frac{y_A}{x_A} = \frac{w_y}{\overline{x}} \tag{4.47}$$

Temos que

$$\begin{aligned}
\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} &= \frac{1}{4}\sqrt{\frac{y_A}{x_A}}\sqrt{\frac{y_B}{x_B}} \\
&= \frac{1}{4}\sqrt{\frac{y_A}{x_A}}\sqrt{\frac{y_A}{x_A}} \qquad [por\ (4.46)] \\
&= \frac{1}{4}\frac{y_A}{x_A} \\
&= \frac{1}{4}\frac{w_y}{\overline{x}}
\end{aligned} \tag{4.48}$$

Como $\frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} = 1$, *temos que*

$$\overline{x} = \frac{w_y}{4} \tag{4.49}$$

Portanto, $\overline{x} = \frac{w_y}{4}$ *é o ponto crítico que faz com que* $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} = 0$, *ou, de forma equivalente,* $\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} = 0$. *Assim, se* $w_x > \frac{w_y}{4}$, *então* $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} < 0$. *Isso implica que qualquer alocação eficiente de Pareto requer que toda a dotação seja alocada em consumo. Se* $w_x < \frac{w_y}{4}$, *então* $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} > 0$. *Isso implica que nenhuma alocação ótima de Pareto requer que toda a dotação seja alocada em consumo. E de forma semelhante para* $w_x = \frac{w_y}{4}$.

Exemplo 4.2. *Uma economia é constituída por dois indivíduos cujas utilidades são* $u_A(f, m_A) = \frac{4}{3}\sqrt{f} + m_A$ *e* $u_B(f, m_B) = \ln(1-f) + m_B$, *em que* f *é representa a poluição gerada pelo consumo de cigarro por parte do indivíduo* A *e* m *representa o gasto com a aquisição de outros bens. Suponha que o indivíduo* B *tenha direito a todo* *o ar puro, mas que possa vender, ao preço unitário p, o direito de poluir parte do ar para A. Se no equilíbrio o indivíduo A paga G unidades monetárias ao indivíduo B para poluir parte do ar, quanto deve ser G? Suponha que $f \in [0, 1]$.*

O indivíduo A escolhe o nível de poluição gerada pelo consumo de cigarro que resolva o seguinte problema:

$$\max_f \frac{4}{3}\sqrt{f} + m_A - pf \tag{4.50}$$

cuja condição de primeira ordem será dada por

$$\frac{\partial u_A}{\partial f} = 0 \quad \Longleftrightarrow \quad \frac{4}{6} f^{-1/2} - p = 0 \quad \Longrightarrow \quad p = \frac{4}{6\sqrt{f}} \tag{4.51}$$

Por sua vez, o indivíduo B resolve o seguinte problema de otimização:

$$\max_f \ln(1-f) + m_B + pf \tag{4.52}$$

cuja condição de primeira ordem será dada por

$$\frac{\partial u_B}{\partial f} = 0 \quad \Longleftrightarrow \quad -\frac{1}{1-f} + p = 0 \quad \Longrightarrow \quad p = \frac{1}{1-f} \tag{4.53}$$

Logo,

$$\begin{aligned} \frac{1}{1-f} &= \frac{4}{6\sqrt{f}} \\ 6\sqrt{f} &= 4(1-f) \\ \frac{9}{4} &= 1 - 2f + f^2 \\ f^2 - \frac{17}{4}f + 1 &= 0 \\ f &= \frac{\frac{17}{4} \pm \sqrt{\frac{225}{16}}}{2} \end{aligned}$$

$$f = \frac{1}{4} \; ou \; f = 4 \tag{4.54}$$

Assim, ele vai pagar $G = pf = \left(\frac{1}{1-f}\right) f = \left(\frac{1}{1-\frac{1}{4}}\right)\frac{1}{4} = \frac{1}{3}.$

Exemplo 4.3. *Considere um grupo de n estudantes. Suponha que cada estudante i estude h_i horas o que gera uma desutilidade de $\frac{h_i^2}{2}$ para ele. O benefício gerado pelo estudo para i depende da sua performance frente à de seus colegas da sala e tem a seguinte forma: $u\left(\frac{h_i}{\bar{h}}\right)$ para todo i, em que $\bar{h} = \frac{1}{n}\sum_i h_i$ denota a média de horas estudadas por todos os alunos da sala e $u(\cdot)$ é uma função crescente e côncava. Encontre o equilíbrio de Nash simétrico.*

Num Equilíbrio de Nash a escolha de cada indivíduo i maximiza sua utilidade dadas as ações dos outros

$$\max_{h_i} u\left(\frac{h_i}{\frac{1}{n}\sum_i h_i}\right) - \frac{h_i^2}{2} \quad \Longrightarrow \quad \max_{h_i} u\left(\frac{nh_i}{h_i + \sum_{j\neq i} h_j}\right) - \frac{h_i^2}{2} \tag{4.55}$$

A condição de primeira ordem é

$$\begin{aligned} \frac{\left(h_i + \sum_{j\neq i} h_j\right) n - nh_i}{\left(h_i + \sum_{j\neq i} h_j\right)^2} u'\left(\frac{nh_i}{h_i + \sum_{j\neq i} h_j}\right) - h_i = 0 \\ \frac{n\sum_{j\neq i} h_j}{\left(h_i + \sum_{j\neq i} h_j\right)^2} u'\left(\frac{nh_i}{h_i + \sum_{j\neq i} h_j}\right) - h_i = 0 \end{aligned} \tag{4.56}$$

Em um equilíbrio simétrico,

$$\frac{n(n-1)h}{(nh)^2}u'\left(\frac{nh}{nh}\right) - h = 0 \tag{4.57}$$

Assim,

$$h^* = \left[\frac{n-1}{n}u'(1)\right]^{1/2} > 0 \tag{4.58}$$

4.3 Externalidades na Produção

Mostramos agora que a alocação de recursos pode não ser eficiente também para o caso da externalidade na produção. Para mostrar isso, considere uma economia simples com duas empresas. A empresa 1 produz o bem x que será vendido em um mercado competitivo. No entanto, a produção de x impõe um custo de externalidade denotado por $e(x)$ para firma 2, que é assumido como sendo convexo e estritamente crescente. Seja y o bem produzido pela firma 2, que é vendido no mercado competitivo. Sejam $c_x(x)$ e $c_y(y)$ as funções de custo das firmas 1 e 2, ambas convexas e estritamente crescentes. Sejam p_x e p_y os preços de x e de y, respectivamente.

As funções objetivo são

$$\max_x \pi_1 = p_x x - c_x(x) \tag{4.59}$$

$$\max_y \pi_2 = p_y y - c_y(y) - e(x) \tag{4.60}$$

Então, pelas condições de primeira ordem, temos quantidades positivas de produtos:

$$\frac{d\pi_1}{dx} = 0 \quad \Longleftrightarrow \quad p_x = c_x'(x) \tag{4.61}$$

$$\frac{d\pi_2}{dy} = 0 \quad \Longleftrightarrow \quad p_y = c_y'(y) \tag{4.62}$$

No entanto, o produto maximizador de lucro, x, obtido da condição de primeira ordem é muito grande do ponto de vista social. A primeira firma só leva em conta o custo privado - o custo que é imposto a si mesmo - mas ignora o custo social - o custo privado mais o custo que impõe à outra firma. Qual é o resultado social eficiente?

O lucro social, $\pi_1 + \pi_2$, não é maximizado em x e y que satisfazem as duas condições de primeira ordem. Se as duas empresas se fundiram de modo a internalizar a externalidade, podemos escrever a função lucro como

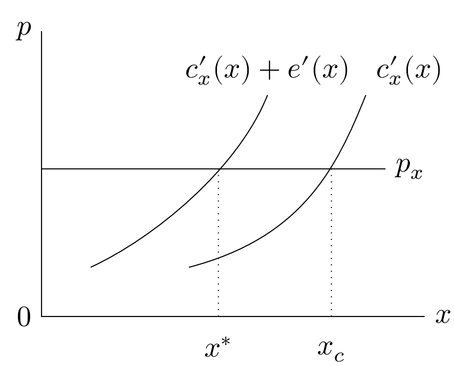
$$\max_{x,y} \pi = p_x x + p_y y - c_x(x) - c_y(y) - e(x) \tag{4.63}$$

que resulta nas seguintes condições de primeira ordem:

$$\frac{\partial \pi}{\partial x} = 0 \quad \Longleftrightarrow \quad p_x = c_x'(x^*) + e'(x^*) \tag{4.64}$$
$$\frac{\partial \pi}{\partial y} = 0 \quad \Longleftrightarrow \quad p_y = c_y'(y^*) \tag{4.65}$$

em que x^* é uma quantidade eficiente de produto; é caracterizado pelo preço ser igual ao custo social marginal. Assim, a produção de x^* é menor que a produção competitiva no caso de externalidade pela convexidade de $e(x)$ e $c_x(x)$.

Figura 4.1 – *O Produto Eficiente x^* é Menor do que o Produto em um Ambiente Competitivo*



4.4 Recurso de Uso Comum

Um recurso de uso comum é um bem em que o consumo é rival e não excludente. Alguns dos exemplos prototípicos são pesqueiros locais, pastagens comuns ou sistemas de irrigação. Nessas situações, os indivíduos tenderão a usar em excesso o recurso, pois escolherão o nível de uso em que a utilidade marginal individual é zero, mas o Pareto ótimo é o nível em que o benefício marginal total é zero, que geralmente é um nível mais baixo de consumo do que o equilíbrio do mercado.

Considere o exemplo a seguir de um recurso de uso comum. O número total de peixes capturados em uma pesca local não excludente é dado por $f(k)$, em que k é o número total de barcos de pesca que trabalham no local de pesca. Assuma que $f'(k) > 0$ e $f''(k) < 0$ e que $f(0) = 0$. Ou seja, assumimos que a produção total de peixes é uma função côncava crescente do número de barcos que trabalham no local de pesca. Observe também que, como consequência da concavidade de $f(\cdot)$, $\frac{f(k)}{k} > f'(k)$[31]. Ou seja, o número de peixes capturados por barco é sempre maior

[31] Uma função de produção $f(k)$ é côncava se, para todos k_1 e k_2 e para qualquer $\lambda \in [0, 1]$, temos:

$$f(\lambda k_1 + (1 - \lambda)k_2) \geq \lambda f(k_1) + (1 - \lambda)f(k_2)$$

Essa desigualdade define a propriedade de concavidade. Intuitivamente, uma função côncava apresenta retornos decrescentes, ou seja, o incremento adicional na produção diminui à medida que aumentamos o capital. O valor médio do produto do capital é dado por:

$$\text{PMe}(k) = \frac{f(k)}{k}$$

Isso representa a produção média por unidade de capital.

O valor marginal do produto do capital é dado por:

$$\text{PMg}(k) = f'(k)$$

Isso representa a variação na produção resultante de uma variação infinitesimal no capital.

Para demonstrar que, em uma função de produção côncava, o valor médio é maior que o valor marginal, consideramos o seguinte. A concavidade de $f(k)$ implica que $f'(k)$ é uma função decrescente, ou seja, à medida que aumentamos K, o valor marginal do produto diminui. Para verificar a relação entre o valor médio $\text{PMe}(k)$ e o valor marginal $\text{PMg}(k)$, podemos diferenciar a função $\text{PMe}(k)$:

que o produto marginal da adição de outro barco. Isto resulta da observação de que o produto médio está diminuindo para uma função de produção côncava. Defina o produto médio como $\mathrm{PMe}(k) = \frac{f(k)}{k}$. Então, $\mathrm{PMe}'(k) = \frac{1}{k}\left(f'(k) - \mathrm{PMe}(k)\right) < 0$. Portanto, $\mathrm{PMe}(k) > f'(k)$.

Os barcos de pesca são produzidos a um custo $c(k)$, em que k é o número total de barcos e $c(\cdot)$ é uma função estritamente crescente e estritamente convexa. O preço do peixe é normalizado para 1.

O número eficiente de barcos de Pareto é encontrado resolvendo

$$\max_k f(k) - c(k) \tag{4.66}$$

o que implica que a condição de primeira ordem para o número ideal de barcos k^0 é:

$$f'\left(k^0\right) = c'\left(k^0\right) \tag{4.67}$$

Seja k_i o número de barcos que o trabalhador i emprega e assuma que há I

$$\mathrm{PMe}'(k) = \frac{d}{dk}\left(\frac{f(k)}{k}\right) = \frac{f'(k)k - f(k)}{k^2}$$

Rearranjando, obtemos:

$$\mathrm{PMe}'(k) = \frac{kf'(k) - f(k)}{k^2}$$

Para que o valor médio esteja diminuindo, é necessário que $\mathrm{PMe}'(k) \leq 0$. Isso ocorre quando:

$$kf'(k) \leq f(k)$$

Ou seja,

$$f'(k) \leq \frac{f(k)}{k} = \mathrm{PMe}(k)$$

Isso demonstra que, para uma função de produção côncava, o valor marginal $\mathrm{PMg}(k) = f'(k)$ é menor ou igual ao valor médio $\mathrm{PMe}(k) = \frac{f(k)}{k}$.

Portanto, se a função de produção é estritamente côncava, o valor médio será estritamente maior que o valor marginal, exceto em $k = 0$.

pescadores. Portanto, $k = \sum_i k_i$. Se p é o preço de um barco de pesca, os produtores do barco resolvem

$$\max_k pk - c(k) \tag{4.68}$$

cuja condição ótima é

$$c'(k) = p \tag{4.69}$$

Pressupondo que cada barco pesqueiro captura o mesmo número de peixes, cada pescador resolve o problema

$$\max_{k_i} \frac{k_i}{k_i + k_{-i}} f(k) - pk_i \tag{4.70}$$

em que $k_{-i} = \sum_{j \neq i} k_j$. A condição ótima deste problema é

$$f'(k^*) \frac{k_i^*}{k^*} + \frac{f(k^*)}{k^*} \left(\frac{k_{-i}^*}{k^*} \right) = p \tag{4.71}$$

A compensação do mercado implica então que:

$$f'(k^*) \frac{k_i^*}{k^*} + \frac{f(k^*)}{k^*} \left(\frac{k_{-i}^*}{k^*} \right) = c'(k^*) \tag{4.72}$$

Como todos os nossos produtores são idênticos, o ideal envolverá $k_i^* = k_j^* \ \forall i, j$. Ou seja, todos os pescadores escolherão o mesmo número de barcos. Se houver um total de n pescadores, podemos reescrever essa condição como:

$$f'(k^*) \frac{1}{n} + \frac{f(k^*)}{k^*} \left(\frac{n-1}{n} \right) = c'(k^*) \tag{4.73}$$

Assim, o lado esquerdo é uma combinação convexa do produto marginal, $f'(k)$, e do produto médio, $\frac{f(k)}{k}$. E dado que $\frac{f(k)}{k} > f'(k)$, isto implica que $f'(k^*)\frac{1}{n} + \frac{f(k^*)}{k^*}\left(\frac{n-1}{n}\right) > f'(k^*)$. Finalmente, como $c'(k)$ é crescente em k isso implica que $k^* > k^0$, ou seja, o mercado "exagera" na pesca.

O que faz com que as pessoas usem de forma demasiada o recurso comum? Considere mais uma vez a condição de otimização do pescador:

$$f'(k^*)\frac{k_i^*}{k^*} + \frac{f(k^*)}{k^*}\left(\frac{k_{-i}^*}{k^*}\right) = p \tag{4.74}$$

Os termos no lado esquerdo correspondem a dois fenômenos diferentes. Primeiro, se o pescador compra outro barco, isso aumenta a captura total e o pescador i obtém um aumento de $\frac{k_i^*}{k^*}$, o termo $f'(k^*)\frac{k_i^*}{k^*}$. Segundo, como o pescador agora tem mais um barco, ele tem uma proporção maior do total de barcos e, portanto, ganha com o fato de que uma proporção maior da captura total é dada a ele. Esse segundo efeito, que corresponde ao termo $\frac{f(k^*)}{k^*}\left(\frac{k_{-i}^*}{k^*}\right)$, que pode ser pensado como um efeito de "roubo de mercado".

Observe que esses dois efeitos podem levar o mercado a agir ineficientemente. Uma vez que o pescador recebe apenas $\frac{k_i^*}{k^*}$ do aumento do número de peixes devido à adição de outro barco, isso tenderá a fazer com que o mercado comporte poucos barcos. No entanto, como adicionar outro barco também resulta em roubar parte das capturas dos outros pescadores, e isso é lucrativo, isso tende a fazer com que os pescadores comprem muitos barcos. Quando $f(\cdot)$ é côncavo, sabemos que o último efeito domina o primeiro e haverá muitos barcos.

O fenômeno anterior, de que o mercado tenderá a usar em excesso recursos comuns, é conhecido como a tragédia dos bens comuns. Há muitas soluções. Por exemplo, colocar uma cota no número de barcos que cada pescador pode possuir ou no número de peixes que cada barco pode pegar ajudaria a resolver o problema. Além disso, a tributação de barcos ou peixes também ajudaria a resolver o problema. O imposto do barco apropriado seria

$$t^* = \frac{k^*_{-i}}{k^*}\left(\frac{f(k^*)}{k^*} - f'(k)\right) \tag{4.75}$$

Além das ferramentas que eu já mencionei, é importante observar que, embora a tragédia dos bens comuns seja um problema real, as pessoas a resolvem há centenas de anos, pelo menos em parte. Frequentemente, quando as pessoas que têm acesso a um recurso comum (como pastagens comuns, um sistema de irrigação ou uma pesca de acesso aberto) fazem parte de uma comunidade próxima, como os moradores de uma vila, encontram maneiras informais de cooperar um com o outro. Como os moradores se conhecem, eles podem encontrar maneiras informais de punir as pessoas que usam demais os recursos.

Considere-se, para exemplo, o trecho abaixo que remete ao Brasil Colônia:

> *Antes da guerra de 1801, os colonos portugueses influíam pouco na vida da terra, mas sonhavam muito com seu progresso. Imaginavam a riqueza que acumulariam com o aproveitamento das oportunidades perdidas pela gente primitiva do lugar, que não sabia o potencial que tinha nas mãos; civilizar a região significava colocar o gado em propriedades demarcadas, dar destino à carne jogada fora, à terra mal utilizada e ao tempo ocioso. Agora essas possibilidades de aproveitamento pareciam realizadas. Chegaram os charqueadores para transformar em mercadoria valiosa a carne. As terras foram rapidamente divididas entre proprietários, que queriam aumentar a produção de reses. Surgiram as primeiras plantações.* (Caldeira, 1995: p. 46)

O trecho ilustra uma solução histórica ao problema do uso de recursos comuns/bens de uso comum que foi a do estabelecimento dos direitos de propriedade privada sobre uma região não aproveitada de forma economicamente eficiente – como o autor parece afirmar. O uso da terra por vários charqueadores poderia ser estabelecido de duas formas. Usando-se o critério de uso comum – no sentido de que não haveria um proprietário específico; ou através do uso da propriedade privada. Neste caso, a alocação pode ser feita de duas formas: através do livre comércio ou através da determinação governamental.

Suponha o primeiro caso, no qual cada charqueador sabe que quanto mais gado colocar na terra, melhor para ele. Cada unidade de gado produz uma unidade de carne e, adicionalmente, suponha que as unidades de gado são homogêneas, i.e., a qualidade da carne gerada é a mesma. Caso cada charqueador esteja interessado no seu próprio lucro, o que tenderá a ocorrer? A teoria econômica prevê que o gado deixará de ser incorporado ao pasto momento em que o lucro econômico for igual a zero, i.e., quando a receita total gerada for igual ao custo total.

A solução para a tragédia dos comuns citada no item anterior, como apontam a maioria dos livros-textos, é que a terra seja dividida entre os charqueadores – o que é o mesmo que dizer que a terra deverá ser privatizada – garantindo a cada um deles o direito de excluir outros de seu uso e também garantindo que cada qual possa ter direito autônomo de uso do próprio lote de terra.

Os direitos de propriedade privada têm 3 características:

1. o direito de tomar decisões sobre as condições físicas e usos de específicos bens;
2. o direito de vender os direitos de propriedade a outras pessoas;
3. o direito de usufruir das rendas e de arcar com as perdas resultantes das decisões de uso.

Se você vende algo de acordo com a lei, interpreta-se que você vende os direitos de propriedade privada, não a coisa propriamente dita.

O direito de propriedade é um direito da pessoa. Mais que isso, Armen Alchian afirma peremptoriamente em um artigo intitulado “Property Rights”:

> *For decades social critics in the United States and throughout the Western world have complained that "property" rights too often take precedence over "human" rights, with the result that people are treated unequally and have unequal opportunities. Inequality exists in any society. But the purported conflict between property rights and human rights is a mirage. Property rights are human rights.*

Assim, Armen Alchian esclarece que o propósito fundamental dos direitos de propriedade é justamente a eliminação da competição destrutiva sobre o controle de recursos econômicos. Uma boa economia de mercado alicerçada sobre a garantia dos direitos de propriedade é precisamente o que promove a competição pacífica.

O direito de propriedade protege o seu bem contra ameaças e danos físicos causados por outras pessoas. Eu não posso, por exemplo, atirar uma pedra na sua janela. Ele não o protege, porém – e nem deve – de efeitos de valor-de-mercado causados por variações de demanda. Se o preço do frete caiu significativamente, o seu direito de propriedade sobre o caminhão não lhe permite fazer uma greve exigindo que o governo fixe um preço máximo para o litro de diesel abaixo do de mercado, apenas para que você não precise arcar com a perda decorrente de uma má decisão que você tomou no passado, mesmo que essa decisão tenha sido induzida por uma política criminosa de "incentivos" de um governo ruim. Lembre-se da condição (3).

4.5 Soluções para as Externalidades

A partir da discussão acima, sabemos que um mercado competitivo em geral pode não resultar em um resultado eficiente de Pareto na presença de externalidades, e é necessário buscar outros mecanismos alternativos para resolver o problema da falha de mercado. Nesta seção, apresentamos agora algumas remédios para essa falha de mercado decorrente da presença de externalidades, como:

- Imposto Pigouviano

- Negociação voluntária (abordagem do Coase)
- Imposto compensatório/subsídio
- Criar mercados com direitos de propriedade
- Intervenção direta
- Fusões
- Desenho de mecanismo

Qualquer uma das soluções acima pode resultar em resultados eficientes de Pareto, mas pode levar a diferentes distribuições de renda. Além disso, é importante saber que tipo de informação é necessário para implementar uma solução listada acima. A maioria das soluções propostas acima precisa fazer as seguintes suposições:

- A fonte e o grau da externalidade são identificáveis.
- Os destinatários da externalidade são identificáveis.
- A relação causal da externalidade pode ser estabelecida objetivamente.
- O custo de prevenir (por diferentes métodos) uma externalidade é perfeitamente conhecido de todos.
- O custo da implementação de impostos e subsídios é insignificante.
- O custo da negociação voluntária é insignificante.

4.5.1 Imposto Pigouviano

Defina uma taxa de imposto, t, tal que $t = e'(x^*)$. Esta taxa de imposto é dada para a firma 1 internalizar a externalidade.

A firma maximiza seu lucro dado o imposto:

$$\max_x \pi_1 = p_x x - c_x(x) - tx \tag{4.76}$$

A condição de primeira ordem é

$$\frac{\partial \pi_1}{\partial x} = 0 \quad \Longleftrightarrow \quad p_x = c'_x(x) + t = c'_x(x) + e'(x^*) \tag{4.77}$$

que é a mesma que a da otimização social. Ou seja, quando a empresa 1 enfrenta o preço errado de sua ação, um imposto $t = e'(x^*)$ deve ser imposto para cada unidade de produção da empresa 1. Isso levará a um resultado social ótimo, menor do que o resultado do equilíbrio competitivo. Tais taxas de correção são chamadas de impostos pigouvianos. O problema com essa solução é que ela exige que a autoridade tributária conheça o custo da externalidade $e(x)$. Mas, como a autoridade conhece a externalidade e como ela estima o valor da externalidade no mundo real? Se a autoridade conhece essa informação, pode também dizer à firma quanto produzir, em primeiro lugar. Então, na maioria dos casos, não funcionará bem.

4.5.2 Negociação Voluntária em Coase e Direitos de Propriedade

Uma abordagem diferente do problema da externalidade depende das partes negociarem uma solução para o problema. A maior novidade da contribuição do Prêmio Nobel Ronald Coase foi o tratamento sistemático da negociação de direitos de propriedade. Para resolver o problema da externalidade, Coase, em um artigo famoso, "*The Problem of Social Cost*", em 1960, argumenta que o sucesso de tal sistema depende de garantir que os direitos de propriedade sejam claramente atribuídos. O chamado Teorema de Coase avalia que, desde que os direitos de propriedade sejam garantidos, as duas partes negociarão de tal forma que o nível ideal da atividade produtora de externalidade seja implementado. Como uma implicação política, um governo deveria simplesmente reorganizar os direitos de propriedade e garanti-los. O mercado, então, poderia cuidar das externalidades sem intervenção direta do governo.

O teorema de Coase contém duas afirmações. Uma é que o nível da externalidade será o mesmo, independentemente da atribuição de direitos de propriedade, o que é chamado Teorema da Neutralidade de Coase. A segunda é que negociações voluntárias sobre externalidades levarão a um resultado ótimo de Pareto, que é chamado de Teorema da Eficiência de Coase. Coase mostra suas proposições principalmente usando exemplos de economia com apenas dois agentes e uma externalidade negativa. Coase fez uma observação de que, na presença de externalidades, a vítima tem um incentivo para pagar à firma para parar a produção se a vítima puder compensar a empresa pagando $p_x - c_x'(x^*)$. O exemplo a seguir captura os argumentos de Coase.

Suponha o seguinte exemplo. Duas empresas: uma é uma fábrica de produtos químicos que descarta produtos químicos em um lago e a outra é um pescador. Suponha que o lago permita a obtenção de um valor de R\$ 50.000. Se os produtos químicos poluem o lago, o peixe não pode ser comido. Como se resolve a externalidade? A proposta de Coase afirma que, desde que os direitos de propriedade do lago sejam claramente atribuídos, resultará em resultados eficientes. Ou seja, o governo deveria dar a propriedade do lago tanto para a empresa química quanto para o pescador, então ele produziria uma produção eficiente. Para ver isso, assuma que o custo de um filtro é denotado por c_f.

1. O lago é entregue à fábrica.

    (a) Seja $c_f =$ R\$ 50.000. O pescador está disposto a comprar um filtro para a fábrica. O pescador pagará pelo filtro para que o produto químico não polua o lago.

    (b) Seja $c_f >$ R\$ 50.000. O produto químico é descartado no lago. O pescador não quer instalar nenhum filtro.

2. O lago é dado ao pescador e a receita líquida da empresa é maior que R\$ 50.000.

(a) Seja $c_f =$ R\$ 50.000. A fábrica compra o filtro para que o produto químico não possa poluir o lago.

(b) $c_f >$ R\$ 50.000. A firma paga R\$ 50.000 ao pescador e o produto químico é descartado no lago.

Como no exemplo acima, os exemplos de Coase apoiam suas afirmações com base em negociações entre firmas ou empresas, e não entre indivíduos. Essa diferença é importante, uma vez que as empresas maximizam os lucros em vez da utilidade e agem como fiduciárias para os indivíduos. Temos que fazer algumas suposições sobre as funções de utilidade dos consumidores para fazer o Teorema de Coase ser mantido.

Agora considere uma economia com dois consumidores e L bens. Assuma que o consumidor i tem a dotação inicial w_i. Cada consumidor tem preferências sobre as commodities que ele consome e sobre alguma ação h que é tomada pelo consumidor 1. Isto é,

$$u_i\left(x_i^1, \ldots, x_i^L, h\right) \tag{4.78}$$

A atividade h é algo que não tem custo monetário direto para a pessoa 1. Por exemplo, h é a quantidade de música alta escutada pela pessoa 1. Para ouvir música alta, o consumidor deve comprar eletricidade, mas a eletricidade pode ser capturada por meio de um dos componentes de x_i. Do ponto de vista do consumidor 2, h representa um efeito externo da ação do consumidor 1. No modelo, assumimos que

$$\frac{\partial u_2}{\partial h} \neq 0 \tag{4.79}$$

Assim, a externalidade neste modelo está no fato de que h afeta a utilidade do consumidor 2, mas não é precificada pelo mercado.

Seja $v_i(p, w_i, h)$ a função de utilidade indireta do consumidor i. O problema de maximização nesse contexto é dado por:

$$v_i(w_i, h) = \max_{x_i} u_i(x_i, h) \tag{4.80}$$

$$\text{sujeito a } px_i \leq w_i \tag{4.81}$$

Assumimos que as preferências são quase-lineares em relação a algum bem numerário. Assim, a função de utilidade indireta do consumidor assume a forma:

$$v_i(w_i, h) = \phi_i(h) + w_i \tag{4.82}$$

Além disso, assumimos que a utilidade é estritamente côncava em h: $\phi_i''(h) < 0$. Novamente, o resultado do equilíbrio competitivo em geral não é Pareto ótimo. O consumidor 1 deve escolher h para maximizar v_1, de modo que a solução interior satisfaça $\phi_1'(h^*) = 0$. Embora a utilidade do consumidor 2 dependa de h, ela não pode afetar a escolha de h.

Por outro lado, o nível socialmente ótimo de h maximizará a soma das utilidades dos consumidores:

$$\max_h \phi_1(h) + \phi_2(h) \tag{4.83}$$

A condição de primeira ordem para um máximo interno é:

$$\phi_1'\left(h^{**}\right) + \phi_2'\left(h^{**}\right) = 0 \tag{4.84}$$

em que h^{**} é a quantidade ótima de Pareto de h. Assim, o ótimo social é quando a soma do benefício marginal dos dois consumidores é igual a zero. No caso em que a externalidade é ruim para o consumidor 2 (música alta), temos que $h^* > h^{**}$. Ou seja, h é produzido em excesso. No caso em que a externalidade é boa para o consumidor 2, muito pouco será fornecido, $h^* < h^{**}$.

Agora mostramos que, desde que os direitos de propriedade sejam claramente atribuídos, as duas partes negociarão de tal forma que o nível ótimo da atividade produtora de externalidade seja implementado. Em primeiro lugar, consideramos o caso em que o consumidor 2 tem o direito de proibir o consumidor 1 de realizar atividades h. Mas esse direito é contratado. O consumidor 2 pode vender ao consumidor 1 o direito de realizar h_2 unidades de atividade h em troca de alguma transferência, T_2. Os dois consumidores irão negociar tanto sobre o tamanho da transferência T_2 quanto sobre o número de unidades da externalidade boa produzida, h_2. Para determinar o resultado da negociação, primeiro especificamos o mecanismo de barganha da seguinte forma:

1. O consumidor 2 oferece ao consumidor 1 um contrato do tipo aceite-recuse especificando um pagamento T_2 e um nível de atividade h_2.

2. Se o consumidor 1 aceitar a oferta, esse resultado será implementado. Se o consumidor 1 não aceitar a oferta, o consumidor 1 não pode produzir nenhuma das externalidades boas, isto é, $h = 0$.

Para analisar isso, comece considerando quais ofertas (h, T) serão aceitas pelo consumidor 1. Como na ausência de acordo o consumidor 1 deve produzir $h = 0$, o consumidor 1 aceitará (h_2, T_2) se e somente se isso oferecer maior utilidade que $h = 0$. Ou seja, o consumidor 1 aceita se e somente se:

$$\phi_1(h_2) - T_2 \geqq \phi(0) \tag{4.85}$$

Dada essa restrição no conjunto de ofertas aceitáveis, o consumidor 2 escolherá (h_2, T_2) para resolver o seguinte problema:

$$\max_{h_2, T_2} \phi_2(h_2) + T_2 \tag{4.86}$$

$$\text{sujeito a } \phi_1(h_2) - T_2 \geqq \phi_1(0) \tag{4.87}$$

Como o consumidor 2 prefere T_2 mais alto, a restrição será *binding* no ótimo (ou seja, a igualdade irá prevalecer). Assim, o problema se torna:

$$\max_{h_2} \phi_1(h_2) + \phi_2(h_2) - \phi_1(0) \tag{4.88}$$

A condição de primeira ordem para este problema é dada por

$$\phi_1'(h_2) + \phi_2'(h_2) = 0 \tag{4.89}$$

Mas essa é a mesma condição que define o nível socialmente ótimo de h_2. Assim, o consumidor 2 escolhe $h_2 = h^{**}$, e usa a restrição, $T_2 = \phi\left(h^{**}\right) - \phi_1(0)$. A oferta (h_2, T_2) é aceita pelo consumidor 1. Assim, esse processo de barganha implementa o ótimo social.

Agora, consideramos o caso em que o consumidor 1 tem o direito de produzir o máximo da externalidade que deseja. Nós mantemos o mesmo mecanismo de barganha. O consumidor 2 faz ao consumidor 1 uma oferta aceite ou recuse (h_1, T_1), em que o índice indica que o consumidor 1 tem o direito de propriedade nessa situação. No entanto, agora, no caso de 1 rejeitar a oferta, o consumidor 1 pode optar por produzir o máximo de externalidade que desejar, o que significa que ela optará por produzir h^*. Assim, a única mudança entre esta situação e o primeiro caso é o que acontece no caso de nenhum acordo ser alcançado. Nesse caso, o problema do consumidor 1 é:

$$\max_{h_1, T_1} \phi_1(h_1) - T_1 \tag{4.90}$$

$$\text{sujeito a } \phi_1(h_1) + T_1 \geq \phi_1\left(h^*\right) \tag{4.91}$$

Mais uma vez, sabemos que a restrição será *binding*, e assim o consumidor 1 escolhe h_1 e T_1 para maximizar

$$\max_{h_1} \phi_1(h_1) + \phi_2(h_1) - \phi_1\left(h^*\right) \tag{4.92}$$

A condição de primeira ordem para este problema é dada por

$$\phi_1'(h_1) + \phi_2'(h_1) = 0 \tag{4.93}$$

que é ótima em $h_1 = h^{**}$. A única diferença aqui é que $T_1 = \phi_1\left(h^*\right) - \phi_1\left(h^{**}\right)$.

Embora ambas as alocações de direitos de propriedade implementem h^{**}, elas têm consequências distributivas diferentes. O pagamento da transferência é positivo no caso em que o consumidor 2 tem os direitos de propriedade, enquanto é negativo quando o consumidor 1 tem os direitos de propriedade. A razão para isto é que o consumidor 2 está em melhor posição de barganha quando o resultado de não barganhar é que o consumidor 1 é forçado a produzir 0 unidades da externalidade boa.

No entanto, observe que, na estrutura quase-linear, a redistribuição do numerário não tem efeito sobre o bem-estar social. O fato de que, independentemente de como os direitos de propriedade são alocados, a barganha leva a uma alocação ótima de Pareto é um exemplo do Teorema de Coase: se a negociação da externalidade pode ocorrer, a negociação levará a um resultado eficiente, independentemente de como os direitos de propriedade são alocados (desde que sejam claramente atribuídos). Observe que direitos de propriedade executáveis e bem definidos são essenciais para a barganha funcionar. Se houver uma disputa sobre quem tem o direito de poluir (ou não poluir), a negociação pode não levar à eficiência. Um requisito adicional para a eficiência é que o próprio processo de barganha seja gratuito. Observe que o governo não precisa conhecer os consumidores individuais aqui - só precisa definir os direitos de propriedade. No entanto, é fundamental que isso seja feito claramente. Assim, o Teorema de Coase fornece um argumento em favor de ter leis claras e tribunais bem desenvolvidos.

O Teorema da Neutralidade de Coase está baseado no fato de que os custos de transação são nulos e os efeitos renda são zero (as funções de utilidade são quase-lineares). O problema deste teorema de Coase é que, os custos de negociação e organização, em geral, não são desprezíveis, e o efeito de renda pode não ser zero. Assim, uma privatização é ótima apenas no caso de custo de transação zero, sem efeito de renda e ambientes econômicos perfeitos.

O problema do Teorema de Eficiência de Coase é mais sério. Primeiro, como apontou Arrow (1979, p. 24), o postulado básico subjacente à teoria de Coase parece ser que o processo de negociação sobre os direitos de propriedade pode ser modelado como um jogo cooperativo, e isso requer a suposição de que cada jogador conhece as preferências ou funções de produção de cada um dos outros jogadores. Quando a informação não é completa ou assimétrica, em geral não temos um resultado ótimo de Pareto. Por exemplo, quando há um poluidor e há muitos indivíduos atingidos pela poluição, surge um problema de "parasitismo" (*free-rider*) e existe um incentivo para que os atingidos pela poluição deturpem suas preferências. Se o poluidor é responsável ou não, pode-se esperar que os atingidos exagerem a quantia necessária para compensar a externalidade. Assim, podemos precisar desenhar um mecanismo de incentivo para resolver o problema do *free-rider*.

Assim, a hipótese de que as negociações sobre externalidades imitarão o ambiente em um equilíbrio competitivo é, como o próprio Coase admitiu, uma hipótese que deve ser considerada como uma conjectura empírica que pode ou não ser confirmada pelos dados. Muitos trabalhos teóricos, portanto, ainda surgem, a fim de fornecer à economia Coasiana uma base rigorosa.

Os direitos de propriedade privada têm 3 características:

1. o direito de tomar decisões sobre as condições físicas e usos de específicos bens;

2. o direito de vender os direitos de propriedade a outras pessoas;

3. o direito de usufruir das rendas e de arcar com as perdas resultantes das decisões de uso.

Se você vende algo de acordo com a lei, interpreta-se que você vende os direitos de propriedade privada, não a coisa propriamente dita.

Algumas pessoas pensam que a defesa dos direitos de propriedade privada coloca o direito sobre bens acima dos direitos das pessoas ou mesmo de direitos humanos. Isso é um erro, pois o direito de propriedade é um direito da pessoa. Mais que isso, Armen Alchian afirma peremptoriamente em um artigo intitulado "*Property Rights*":

"*For decades social critics in the United States and throughout the Western world have complained that "property" rights too often take precedence over "human" rights, with the result that people are treated unequally and have unequal opportunities. Inequality exists in any society. But the purported conflict between property rights and human rights is a mirage. Property rights are human rights*".

Há alguns problemas práticos concernentes ao teorema de Coase. O primeiro deles diz respeito à própria atribuição dos direitos de propriedade. Enquanto nos mercados convencionais está sempre claro quem é o comprador e quem é o vendedor, no caso das externalidades a definição é um tanto arbitrária. Afinal, deve a casa ter direito ao ar puro ou a fábrica direito a produzir fumaça?

Em segundo lugar, e talvez seja essa a maior limitação prática do teorema de Coase, estão os custos de transação. Estes podem surgir por várias razões, principalmente quando o número de pessoas envolvidas é grande.

Em terceiro está o tamanho do mercado, no caso em que o número de pessoas envolvidas é pequeno. Neste caso a hipótese competitiva capaz de garantir eficiência desaparece e ficamos envolvidos com o problema da barganha que pode ou não gerar uma solução eficiente.

Finalmente, cabe chamar a atenção para o fato de que a invariância à determinação dos direitos de propriedade depende da inexistência de efeitos renda, já que a distribuição de ganhos depende da alocação de direitos. Quando há efeito renda, diferentes atribuições de direito de propriedade geram diferentes disposições a pagar por determinado bem (ou pela redução dele). A invariância pode ser perdida neste caso.

Exemplo 4.4. *O indivíduo A cria cabras em seu sítio. A quantidade de cabras existentes é a quantidade máxima que a terra pode suportar. Ele está pensando em comprar cabras adicionais. Porém, nesse caso, as cabras adicionais irão invadir o sítio vizinho para pastar. Se o indivíduo A adquirir uma cabra e ela invadir o sítio vizinho, ele aufere benefícios líquidos positivos, que são medidos em termos monetários. O indivíduo B, proprietário do sítio vizinho, planta alfaces. As cabras que invadem seu sítio comem as alfaces, causando-lhe danos que também são medidos em termos monetários. De acordo com o número de cabras de A que invadem a plantação de B, a tabela de benefícios totais para A e danos totais para B é como a seguir:*

Cabras	Benefício total de A	Custo total de B	Benefício marginal	Custo marginal	Excedente da transação	Excedente total
0	0	0				
1	50	10	50	10	40	40
2	90	30	40	20	20	60
3	120	59	30	29	1	61
4	140	99	20	40	-20	41
5	150	148	10	49	-39	2
6	152	205	2	57	-55	-53
7	145	240	-7	45	-38	-91

Existe um número socialmente ótimo de cabras que o indivíduo A pode adquirir, apesar das eventuais perdas que essas cabras adicionais causem ao indivíduo B. Se A adquirir a primeira cabra, ele ganha \$50, mas o indivíduo B perde \$10. Em termos líquidos, há um ganho de \$40 = \$50 - \$10. Socialmente vale a pena. Logo, o indivíduo A deveria adquirir a primeira cabra adicional.

Será que ele deve adquirir a segunda cabra? Se fizer isso, A ganha mais \$40. Assim, o benefício total da aquisição de duas cabras é \$90, que é o número mostrado na tabela. Note que aqui estamos calculando os benefícios marginais: o quanto cada *unidade a mais acrescenta ao benefício total. Ora, o indivíduo B, por outro lado, terá um custo adicional de \$20. De fato, como a primeira cabra lhe causa um dano de \$10 e duas lhe causam um dano de \$30, então o dano causado pela segunda cabra é a diferença \$30 - \$10 = \$20. Assim, a aquisição da segunda cabra fornece um ganho de \$40 para o indivíduo A e uma perda de \$20 para o indivíduo B. Em termos líquidos, há um ganho de \$20 = \$40 - \$20. Vale a pena, do ponto de vista social, adquirir a segunda cabra.*

Vale a pena adquirir a terceira cabra? Para o indivíduo B, a terceira cabra lhe dá um benefício adicional de \$30. Para o indivíduo B, o custo adicional causado pela terceira cabra é de \$29. Há um ganho social de \$1. Vale a pena. A terceira cabra é então adquirida.

E a quarta cabra? A aquisição da quarta cabra dá ao indivíduo A um benefício adicional de \$20, mas causa ao indivíduo B um custo adicional de \$40. Não vale a pena. Logo a quarta cabra não deve ser adquirida. Daí para a frente as coisas só pioram.

Em suma, é socialmente ótimo que o indivíduo A adquira mais três cabras. Elas invadirão o sítio vizinho e comerão as alfaces, mas os benefícios superarão os custos. O ganho social é \$61 = \$120 - \$59, e este é o maior valor possível. Alto lá! Há algo estranho. Com que direito o proprietário A deixa suas cabras invadirem o sítio do vizinho B?

Suponha que o dono das cabras é, por lei, responsável por quaisquer danos que venham a ser causados ao proprietário do sítio invadido. Em outras palavras, se uma cabra invadir o sítio de B, o dono das cabras deverá ressarcir ao indivíduo B o montante exato dos danos causados. O plantador de alfaces, o indivíduo B, possui o direito de propriedade sobre seu sítio e sua produção. Se A lhe causar danos, basta que B vá a corte e reclame. A lei lhe garantirá o ressarcimento. O indivíduo A sabe disso. Portanto, se suas cabras pularem a cerca, ele naturalmente ressarcirá o vizinho, pois, se não o fizer, perderá na corte de qualquer jeito.

O que o criador de cabras deve fazer? Vale a pena adquirir três cabras, pois para cada uma delas o benefício adicional supera o custo adicional que ele tem que pagar ao vizinho. Logo, se o direito de propriedade pertence ao plantador de alfaces, a decisão privada do criador de cabras coincide com a decisão socialmente ótima.

A outra possibilidade é bem mais interessante. E se o direito de propriedade pertencesse ao criador de cabras, ao indivíduo A? O que o indivíduo B faria? Ora, por lei, ele não pode evitar que o criador de cabras adquira mais cabras e, por conseguinte, que elas comam suas alfaces. A solução é pagar ao indivíduo A para que ele não adquira mais cabras. Como isso funciona?

Bem, o indivíduo B sabe que A quer adquirir mais uma cabra. Essa cabra causará a B um dano de $10. Já que ele vai perder mesmo esses $10, ele decide pagar até $10 ao indivíduo A para ele não adquirir a cabra. Para A, a primeira cabra adicional vale $50. Obviamente, ele preferirá adquirir a cabra, mesmo que B faça sua oferta máxima de $10. A negociação não deu certo: o indivíduo A adquiriu a primeira cabra.

Agora A está cogitando de adquirir a segunda cabra. O indivíduo B bate à porta de A mais uma vez oferecendo-lhe dinheiro para não adquirir a segunda cabra. Como o custo adicional da segunda cabra para B é $20, ele está disposto a pagar até $20. Para A, o benefício adicional da segunda cabra é de $40, mesmo que B faça sua oferta máxima, é mais vantajoso para A não aceitá-la. Ele compra enfim a segunda cabra.

Finalmente, A quer comprar a terceira cabra. Agora B oferece a A $29 para que este não compre a cabra. Como para A a terceira cabra lhe dá um benefício adicional de $30, ele não aceita a oferta e compra a terceira cabra.

Se A quiser comprar a quarta cabra, o indivíduo B decidirá pagar-lhe $40. Como para A a quarta cabra lhe dá um benefício adicional de $20, ele aceita a oferta e não compra a quarta cabra.

Logo, mesmo neste outro caso, alcançou-se a solução socialmente desejada.

Os indivíduos podem resolver a disputa privadamente. Desde que eles possam *negociar sem custos de outra natureza que não os custos privados (é o que chamamos de hipótese dos custos de transação nulos) e desde que os direitos de propriedade estejam bem definidos (não importando como esses direitos estão de fato alocados entre os indivíduos), então o mercado pode resolver essas disputas causadas por externalidades sem qualquer intervenção do governo. O único papel do governo, nesse caso, seria assegurar que os direitos de propriedade estivessem bem definidos e que a livre negociação ocorresse sem empecilhos.*

Tal resultado simples, conhecido como Teorema de Coase, mostra como o papel das instituições, principalmente as jurídicas, são importantes para o bom funcionamento da economia de mercado.

4.5.3 Criação de um Mercado

Podemos considerar a externalidade como uma falta de mercado para uma "externalidade". Para o exemplo acima, um mercado faltante é um mercado para a poluição. Adicionar um mercado para a empresa 2 para expressar sua demanda por poluição – ou por uma redução da poluição - fornecerá um mecanismo para alocações eficientes. Ao adicionar esse mercado, a empresa 1 pode decidir quanta poluição deseja vender, e a empresa 2 pode decidir quanta poluição deseja comprar.

Seja r o preço da poluição, x_1 as unidades de poluição que a empresa 1 quer vender, x_2 as unidades de poluição da empresa 2 querem comprar. Normalizamos o produto da empresa 1 por x_1. Os problemas de maximização do lucro se tornam:

$$\max_x \pi_1 = p_x x_1 + r x_1 - c_1(x_1) \tag{4.94}$$

$$\max_y \pi_2 = p_y y - r x_2 - e_2(x_2) - c_y(y) \tag{4.95}$$

As respectivas condições de primeira ordem são:

$$\frac{\partial \pi_1}{\partial x} = 0 \quad \Longleftrightarrow \quad p_x + r = c_1'(x_1) \tag{4.96}$$

$$\frac{\partial \pi_2}{\partial y} = 0 \quad \Longleftrightarrow \quad p_y = c_y'(y) \quad \text{ou} \quad -r = e'(x_2) \tag{4.97}$$

No equilíbrio de mercado, $x_1^* = x_2^* = x^*$. Logo,

$$p_x = c_1'(x^*) + e'(x^*) \tag{4.98}$$

o que resulta em um resultado social ótimo.

4.5.4 Mecanismos de Compensação

Os impostos pigouvianos não são adequados em geral para resolver externalidades devido ao problema de informação: a autoridade fiscal não pode conhecer o custo imposto pela externalidade. Como alguém pode resolver este problema de informação incompleta? Varian (1994)[32] propôs um mecanismo de incentivo que incentiva as empresas a revelarem corretamente os custos que elas impõem aos outros. Aqui, discutimos esse mecanismo. Em resumo, um mecanismo consiste em um espaço de mensagem e uma função de resultado (regras de um jogo).

Espaço de estratégia (espaço de mensagem): $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ com $\mathcal{M}_1 = \{(t_1, x_1)\}$ e $\mathcal{M}_2 = \{(t_2, x_2)\}$, onde t_1 é interpretado como um imposto Pigouviano proposto pela empresa 1 e x_1 é o nível proposto de produto pela empresa 1 e t_2 é interpretado como um imposto Pigouviano proposto pela firma 2 e y_2 é o nível proposto de produto pela firma 2.

O mecanismo tem dois estágios:

1. Fase de anúncio: as empresas 1 e 2 dizem as taxas de imposto Pigouvianas,

[32] Varian, H. "*A Solution to the Problem of Externalities when Agents Are Well Informed*". American Economic Review, 84(1994), 1278–1293.

$t_i, i = 1, 2$, que podem ou não ser o nível eficiente de tal taxa de imposto.

2. Estágio de escolha: se a empresa 1 produz x unidades de poluição, a empresa 1 deve pagar t_2x para firmar 2. Assim, cada firma considera a taxa de imposto como dada. A empresa 2 recebe as unidades t_1x como compensação. Cada empresa paga uma multa, $(t_1 - t_2)^2$, se anunciar taxas de impostos diferentes.

Os lucros das duas firmas são dados por:

$$\max_x \pi_1 = p_x x - c_x(x) - t_2 x - (t_1 - t_2)^2 \tag{4.99}$$

$$\max_y \pi_2 = p_y y - c_y(y) + t_1 x - e(x) - (t_1 - t_2)^2 \tag{4.100}$$

Como esse é um jogo de dois estágios, podemos usar o equilíbrio perfeito do subjogo, ou seja, um equilíbrio no qual cada empresa leva em conta as repercussões de sua escolha do primeiro estágio nos resultados do segundo estágio. Como de costume, resolvemos esse jogo olhando primeiro para o estágio 2.

- **Estágio** 2

A empresa 1 escolherá $x(t_2)$ para satisfazer a condição de primeira ordem:

$$\frac{\partial \pi_1}{\partial x(t_2)} = 0 \quad \Longleftrightarrow \quad p_x - c_x'(x) - t_2 = 0 \tag{4.101}$$

Note que, pela convexidade de c_x, $c_x''(x) > 0$, temos que

$$x'(t_2) = -\frac{1}{c''(x)} < 0 \tag{4.102}$$

A firma 2 escolherá y para satisfazer a condição de primeira ordem:

$$\frac{\partial \pi_2}{\partial y} = 0 \quad \Longleftrightarrow \quad p_y - c_y'(y) = 0 \tag{4.103}$$

- **Estágio** 1

  Cada empresa escolherá a alíquota t_1 e t_2 que maximize seus lucros.

  Para a firma 1,

$$\max_{t_1} p_x x - c_x(x) - t_2 x(t_2) - (t_1 - t_2)^2 \tag{4.104}$$

  o que nos dá a condição de primeira ordem:

$$\frac{\partial \pi_1}{\partial t_1} = 0 \quad \Longleftrightarrow \quad 2(t_1 - t_2) = 0 \tag{4.105}$$

  Então a solução ótima é $t_1^* = t_2$.

  Para a firma 2,

$$\max_{t_2} p_y y - c_y(y) - t_1 x(t_2) - e(x(t_2)) - (t_1 - t_2)^2 \tag{4.106}$$

  o que nos dá a condição de primeira ordem:

$$\begin{aligned} \frac{\partial \pi_2}{\partial t_2} = 0 \quad \Longleftrightarrow \quad & t_1 x'(t_2) - e'(x(t_2))x'(t_2) + 2(t_1 - t_2) = 0 \\ & [t_1 - e'(x(t_2))]x'(t_2) + 2(t_1 - t_2) = 0 \end{aligned} \tag{4.107}$$

  Disso decorre que

$$t^* = e'(x(t^*)) \quad \text{com} \quad t^* = t_1^* = t_2^* \tag{4.108}$$

  Substituindo a taxa de imposto de equilíbrio, $t^* = e'(x(t^*))$, que maximiza o lucro da firma 1 no segundo estágio, obtemos:

$$p_x = c_x'(x^*) + e'(x^*) \tag{4.109}$$

que é a condição para a eficiência social da produção.

Esse mecanismo funciona estabelecendo incentivos opostos para dois agentes. A empresa 1 sempre tem um incentivo para igualar seu resultado ao anunciado pela empresa 2. Mas considere o incentivo da empresa 2. Se a firma 2 achar que a firma 1 proporá uma grande taxa de compensação t_1 para ela, ela irá querer que a empresa 1 seja taxada o mínimo possível, de modo que a firma 1 produza o máximo possível. Por outro lado, se a empresa 2 achar que a empresa 1 proporá um pequeno t_1, ela irá querer que a empresa 1 seja taxada o máximo possível. Assim, o único ponto em que a empresa 2 é indiferente quanto ao nível de produção da empresa 1 é onde a empresa 2 é exatamente compensada pelo custo da externalidade. Em geral, o objetivo do indivíduo é diferente do objetivo social. No entanto, podemos ser capazes de construir um mecanismo adequado para que a meta de maximização de lucro do indivíduo seja consistente com a meta social, como alocações eficientes.

4.5.5 Fusões

Vamos entender como uma fusão pode resolver o problema das externalidades por meio de um exemplo.

Exemplo 4.5. *Considere a seguinte situação:*

- *Uma siderúrgica produz aço e poluição em conjunto.*
- *A poluição afeta adversamente uma pesca nas proximidades.*
- *Ambas as empresas são competitivas.*
- *Seja p_s o preço de mercado de aço.*
- *Seja p_f o preço de mercado de peixe.*

Suponha que $s_s(s,x)$ *é o custo da empresa siderúrgica de produzir* s *unidades de aço em conjunto com* x *unidades de poluição. Se a empresa siderúrgica não enfrenta nenhum dos custos externos de sua produção de poluição, sua função de lucro é*

$$\pi_s(s,x) = p_s s - c_s(s,x) \tag{4.110}$$

e o problema da empresa é

$$\max_{s,x} \pi_s(s,x) = p_s s - c_s(s,x) \tag{4.111}$$

cujas condições de primeira ordem são:

$$\frac{\partial \pi_s(s,x)}{\partial s} = 0 \quad \Longleftrightarrow \quad p_s - \frac{\partial c_s(s,x)}{\partial s} = 0 \tag{4.112}$$

$$\frac{\partial \pi_s(s,x)}{\partial x} = 0 \quad \Longleftrightarrow \quad \frac{\partial c_s(s,x)}{\partial x} = 0 \tag{4.113}$$

Suponha que $c_s(s,x) = s^2 + (x-4)^2$ *e que* $p_s = 12$*. Então,*

$$\pi_s(s,x) = 12s - s^2(x-4)^2. \tag{4.114}$$

Logo, $s^* = 6$ *e* $x^* = 4$*, o que implica que* $\pi_s(s^*, x^*) = 36$*.*

O custo para a pesca de capturar f *unidades de peixe quando a siderúrgica emite* x *unidades de poluição é* $c_f(f,x)$*. Dado* f*,* $c_f(f,x)$ *aumenta com* x*; isto é, a empresa siderúrgica inflige uma externalidade negativa na pesca.*

Suponha que $c_f(f,x) = f^2 + xf$ *e que* $p_f = 10$*. Então,*

$$\pi_f(f,x) = 10f - f^2 - xf. \tag{4.115}$$

Logo, $f^* = 5 - \dfrac{x}{2}$*. Como* $x^* = 4$*, temos que* $f^* = 3$*, dando um lucro máximo de* $\pi_f(f,x) = 9$*. Note que o custo externo é de 12. Essas escolhas das duas empresas são* *eficientes? Quando a empresa siderúrgica ignora os custos externos de suas escolhas, a soma dos lucros da empresa é de* $R\$36 + R\$9 = R\$45$. *R\$ 45 é o maior lucro total possível que pode ser alcançado? Suponha que as duas empresas se fundam para se tornar uma. Qual é o maior lucro que essa nova empresa pode alcançar?*

$$\pi_m(s, f, x) = 12s + 10f - s^2 - (x-4)^2 - f^2 - xf \tag{4.116}$$

Quais as escolhas de d, f e x que maximizam o novo lucro? Para tanto, é só calcular as condições de primeira ordem:

$$\frac{\partial \pi_s(s,x)}{\partial s} = 0 \iff 12 - 2s = 0 \tag{4.117}$$
$$\frac{\partial \pi_s(s,x)}{\partial f} = 0 \iff 10 - 2f - x = 0 \tag{4.118}$$
$$\frac{\partial \pi_s(s,x)}{\partial x} = 0 \iff -2(x-4) - f = 0 \tag{4.119}$$

A solução é $s^* = 6$, $f^* = 4$ *e* $x^* = 2$. *O lucro da firma é, portanto,* $\pi_m(s^*, f^*, x^*) = 48$.

A fusão melhorou a eficiência. Por si só, a empresa siderúrgica produziu $x^* = 4$ *unidades de poluição. Dentro da empresa resultante da fusão, a produção de poluição é de apenas* $x^* = 2$ *unidades. Portanto, a fusão causou uma melhoria na eficiência e menos produção de poluição. Por quê? Quando não precisa enfrentar os custos externos de sua poluição, a empresa siderúrgica aumenta a poluição até que esse custo marginal seja zero; portanto* $x^* = 4$. *O custo marginal de poluição da empresa resultante da fusão é maior porque enfrenta o custo total de sua própria poluição através do aumento dos custos de produção na pesca, portanto, menos poluição é produzida pela empresa resultante da fusão. Portanto, a fusão internaliza uma externalidade e induz a eficiência econômica.*

Há dois problemas, porém, com essa solução. O primeiro é que gera uma tendência à concentração na indústria e o segundo é que sua aplicação tende a ser muito limitada no caso de externalidade dos consumos (é pouco provável que você vá se casar com sua colega do lado só para aliviar a externalidade nos consumos de ambos, principalmente no caso de se tratar de externalidade negativa).

4.6 Teorema de Coase: Panorama

O teorema de Coase é simultaneamente uma das ideias mais influentes e das mais controversas na história da economia pós-Segunda Guerra Mundial. Originado a partir da teoria econômica das externalidades, seu alcance agora se estende a virtualmente todos os subcampos da economia e do direito e, de fato, a campos de estudo em todo o espectro acadêmico e literaturas ao redor do mundo. No entanto, sua validade como uma proposição na lógica econômica foi por muitos anos um ponto de significativa contenda e, mesmo hoje, não é de maneira alguma universalmente aceita. A relevância do teorema para problemas do mundo real também é altamente contestada. Alguns sugeriram que ele deveria ser o ponto de partida para a análise de políticas de externalidades (por exemplo, Turvey 1963), enquanto outros restringiriam sua aplicabilidade a um “mundo de contabilidade sem custos de transação” (Randall 1975). Foi o colega de Coase na Universidade de Chicago, George Stigler, quem deu o nome pelo qual o resultado da negociação de Coase (1960) passou a ser conhecido – curiosamente, em seu livro-texto, *The Theory of Price* (1966).

Seria padrão, neste ponto, fazer uma declaração do teorema de Coase, mas isso é bastante problemático. Embora seja difícil encontrar um economista que não consiga fornecer uma declaração do teorema, reunir uma coleção de tais declarações revelaria uma ampla variedade de opiniões sobre o conteúdo do teorema – especificamente, as suposições subjacentes e as alegações feitas por ele. De fato, alguns economistas adotam versões do teorema que outros consideram demonstravelmente falsas. O mesmo não pode ser dito sobre os outros famosos teoremas da economia – teoremas que, como acontece, aparecem muito menos proeminentemente na literatura do que o que leva o nome de Coase.

Para aqueles impacientes para saber como a história termina, faremos a declaração do teorema de Coase aqui antes de passar para uma análise de como chegamos a essa delimitação particular dele.

> Se os agentes são racionais e os custos de transação são zero, os recursos serão alocados de forma eficiente, independentemente de como os direitos sobre esses recursos são inicialmente distribuídos. Além disso, se as funções de utilidade são uniformemente afins em bens privados e o registro de valores subjetivos não é limitado pela riqueza, essa alocação é independente da estrutura inicial dos direitos.

Quando Ronald Coase, então membro da faculdade de economia da Universidade da Virgínia, escreveu "*The Problem of Social Cost*" (1960), oferecendo uma crítica da teoria das externalidades recebida, ele não pretendia oferecer ao mundo um teorema. Ele nem sequer considerou a proposição que agora conhecemos como o teorema de Coase como a principal percepção do artigo. Sua discussão sobre soluções negociadas para externalidades era pouco mais do que uma ficção conveniente destinada a mostrar o erro da igualmente fictícia (em sua mente) tradição pigouviana e a apontar o caminho para uma abordagem muito diferente de pensar sobre a teoria e a política das externalidades – uma abordagem institucional comparativa baseada na natureza recíproca das externalidades e no custo da coordenação, à qual ele dedicou aproximadamente dois terços do seu artigo. Na verdade, Coase não escreveu mais nada sobre seu resultado de negociação por duas décadas.

O que agora conhecemos como o teorema de Coase é muito mais uma criação da comunidade de economistas e estudiosos do direito que se empenharam em analisar e aplicar a percepção de Coase. O teorema é, por qualquer número de medidas, um dos resultados mais curiosos na história das ideias econômicas. Seu desenvolvimento tem sido envolto em lembranças imprecisas, controvérsias políticas e toda sorte de confusões pessoais e comunitárias, servindo como um excelente exemplo do processo bagunçado pelo qual novas ideias se tornam conhecimento científico. Não há uma declaração única do teorema de Coase; há literalmente dezenas de declarações diferentes sobre ele, muitas das quais são inconsistentes com outras e parecem marcar partidas significativas do que Coase havia argumentado em 1960. Entre elas, podemos citar:

Calabresi (1968) ... se se assume racionalidade, ausência de custos de transação e ausência de impedimentos legais para a negociação, todas as má alocações de recursos seriam totalmente corrigidas no mercado por meio de acordos. [A ênfase em "todas" é de Calabresi.]

Buchanan (1972) ... na ausência de custos de transação e efeitos de renda, a atribuição de direitos de propriedade ou reivindicações não afeta a alocação de recursos.

Regan (1972) ... em um mundo de competição perfeita, informação perfeita e custos de transação zero, a alocação de recursos na economia será eficiente e não será afetada por regras legais relacionadas ao impacto inicial dos custos resultantes das externalidades.

Miller (1978) sempre que a contratação e a aplicação dos direitos de propriedade forem relativamente sem custo, os custos sociais e os custos privados tenderão a ser um e o mesmo.

Greenwood e Ingene (1978) desde que as negociações (transações de mercado) sejam sem custo, a alocação de recursos ao final do processo de negociação é socialmente ótima porque possui as mesmas características da posição de equilíbrio obtida por uma fusão das partes afetadas em uma única empresa que internaliza completamente as externalidades.

Hoffman e Spitzer (1982) ... uma mudança em uma regra de responsabilidade deixará as decisões de produção e consumo dos agentes inalteradas e economicamente eficientes dentro do seguinte (implícito) *framework*: (a) dois agentes para cada negociação de externalidade, (b) conhecimento perfeito das funções de produção e lucro ou utilidade (convexas) um do outro, (c) mercados competitivos, (d) custos de transação zero, (e) sistema judicial sem custos, (f) produtores maximizadores de lucro e consumidores maximizadores de utilidade esperada, (g) sem efeitos de riqueza, (h) os agentes farão acordos mutuamente vantajosos na ausência de custos de transação.

Cooter e Ulen (1988) ... quando as partes podem negociar juntas e resolver suas divergências por cooperação, seu comportamento será eficiente independentemente da regra de lei subjacente.

Schwab (1988) ... uma mudança na [lei] não afeta nem a eficiência dos contratos nem a distribuição de riqueza entre as partes.

Hurwicz (1995) ... o nível de uma externalidade (por exemplo, poluição) é independente de fatores institucionais (em particular, a atribuição de responsabilidade por danos), exceto na presença de custos de transação.

Russell (1995) ... o nível de uma externalidade produzido no equilíbrio competitivo de uma economia não é afetado por uma realocação de direitos de propriedade negociáveis na atividade que causa a externalidade.

Allen (1999) na ausência de custos de transação, a alocação de recursos é independente da distribuição dos direitos de propriedade.

Dixit e Olson (2000) se os custos de transação são zero, as partes racionais alcançarão necessariamente uma alocação Pareto-eficiente por meio de transações voluntárias ou negociações.

Dixit e Olson (2000) ... se o Teorema de Coase é verdadeiro, então também o é um “super Teorema de Coase”, ou seja, “partes racionais necessariamente alcançarão uma alocação Pareto-eficiente por meio de transações voluntárias ou negociações, não importa quão altos sejam os custos de transação."

Foss e Foss (2005) o teorema de Coase afirma que todo valor que pode ser criado a partir da troca e uso dos bens disponíveis em uma economia será, de fato, criado quando os custos de transação estão ausentes.

Rochet e Tirole (2006) o teorema de Coase afirma que, se os direitos de propriedade estão claramente estabelecidos e são negociáveis, e se não há custos de transação nem informações assimétricas, o resultado da negociação entre duas (ou várias) partes será Pareto eficiente, mesmo na presença de externalidades.

Foros e Hansen (2011) sempre que há ganhos com o comércio, ... existem contratos tais que ambas as [partes] estão em uma posição melhor ao assinarem um acordo.

O teorema nunca recebeu uma prova formal amplamente aceita; no entanto, tem sido o objeto de inúmeras tentativas de refutá-lo em uma sequência de análises e debates que continua até hoje. Foi rotulado como "tautologia" e a "lei de Say da economia do bem-estar" (Calabresi 1968), uma "falsidade iluminadora" (Cooter 1982), e até mesmo um "preceito religioso" (Posin 1993). Halpin (2007) chama o teorema de "teoricamente degenerado e ideologicamente carregado". Usher (1998) agrupa essas várias acusações, alegando que o teorema é "tautológico, incoerente ou errado", com o veredicto específico dependendo de qual versão do teorema se adota. O ceticismo sobre seu status como teorema é refletido nos vários rótulos alternativos atribuídos a ele na literatura: a "conjectura de Coase" (Stiglitz 2000; Chipman e Tian 2011), a "proposição de Coase" (Samuelson 1995), a "hipótese de Coase" (Conley e Smith 2005a) e a "fábula de Coase" (Ackerman 1982).

A natureza das suposições subjacentes do teorema é frequentemente dita tornar seu domínio de aplicabilidade direto nulo; no entanto, ele foi invocado, criticado e aplicado a questões de política legal-econômica em milhares de artigos de periódicos e livros de economia e direito, assim como em periódicos abrangendo áreas desde a filosofia (Hale 2008) até a literatura (Minda 2001) e a biologia (Frech 1973). De fato, o teorema de Coase pode ser o único conceito econômico cujo uso é mais extenso fora da economia do que dentro dela. Embora seja uma declaração positiva sem implicações normativas diretas, foi tanto usado como justificativa para a aplicação de princípios econômicos na tomada de decisões judiciais quanto visto como um tiro inicial no que muitos percebem como uma virada neoliberal impulsionada pela escola de Chicago na economia – a última, apesar do fato de que a difusão do teorema para a literatura jurídica, pelo menos, originou-se bem fora (e, poderia-se argumentar, à esquerda daquilo comumente associado) de Chicago e quase uma década antes do surgimento da análise econômica do direito de Chicago (Medema 2014d). Foi ridicularizado de um lado como ideologia conservadora e do outro como ideologia liberal. Semelhante à proposição da mão invisível de Adam Smith (Smith 1776), foi um ponto relativamente menor no trabalho do autor, mas ganhou vida própria nas mãos dos comentaristas subsequentes.

Entender o lugar que o teorema de Coase ocupa dentro da análise econômica hoje exige que primeiro voltemos nossa atenção para o passado.

Uma das características definidoras da literatura sobre o teorema de Coase é a ausência de uma declaração singular e geralmente aceita do teorema. A multiplicidade de teoremas de Coase alimentou a controvérsia sobre a correção do teorema como uma proposição na lógica econômica, bem como disputas sobre o domínio de sua aplicabilidade no mundo real. As raízes dessas perspectivas divergentes residem em visões diferentes quanto ao contexto dentro do qual a atividade contemplada pelo teorema se desenrola, as suposições subjacentes ao teorema (incluindo o conteúdo atribuído a elas) e os resultados reivindicados para ele. Como observou Usher (1983), o teorema de Coase é "o único teorema com um nome estabelecido, mas sem conteúdo universalmente reconhecido". Não é pequena a responsabilidade que recai sobre o próprio Coase. Sendo ele nem um economista moderno no que diz respeito aos métodos formais nem ciente de que estava apresentando uma ideia que seria rotulada como um teorema sua análise exibe uma flexibilidade que a torna sujeita a múltiplas interpretações e, como veremos, a uma ampla gama de críticas.

Utilizando a agora famosa ilustração do fazendeiro cujas vacas pisoteiam as culturas de um agricultor vizinho, Coase demonstrou que os dois agentes negociariam para alcançar o resultado que maximiza o valor de sua produção conjunta, independentemente de a quem os direitos relevantes fossem atribuídos. Ao concluir essa análise, Coase fez a seguinte observação:

> *É necessário saber se o negócio prejudicial é responsável ou não pelos danos causados, pois sem o estabelecimento dessa delimitação inicial de direitos não pode haver transações de mercado para transferi-los e recombiná-los. Mas o resultado final (que maximiza o valor da produção) é independente da posição legal se o sistema de preços for assumido como funcionando sem custo.*

Isso é o mais próximo que Coase chegou de uma declaração do que agora chamamos de teorema de Coase.

Podemos identificar três suposições subjacentes à conclusão de Coase:

1. Primeiro, os agentes envolvidos – o fazendeiro e o criador de gado de Coase – vendem seus produtos em mercados perfeitamente competitivos.

2. Segundo, o sistema de preços funciona sem custo ou, como ele colocou mais tarde no artigo, não há custos envolvidos na realização de transações de mercado.

3. Finalmente, Coase assumiu a existência de uma atribuição inicial de direitos legais sobre os recursos relevantes, com base na ideia de que a presença desses direitos era necessária para induzir negociações.

Uma das características definidoras da literatura sobre o teorema de Coase é a ausência de uma declaração singular e geralmente aceita do teorema. A multiplicidade de "teoremas de Coase" alimentou a controvérsia sobre a correção do teorema como uma proposição na lógica econômica, bem como disputas sobre o domínio de sua aplicabilidade no mundo real. As raízes dessas perspectivas divergentes residem em visões diferentes quanto ao contexto dentro do qual a atividade contemplada pelo teorema se desenrola, as suposições subjacentes ao teorema (incluindo o conteúdo atribuído a elas) e os resultados reivindicados para ele. Como observou Usher, o teorema de Coase é "o único teorema com um nome estabelecido, mas sem conteúdo universalmente reconhecido". Não é pequena a responsabilidade que recai sobre o próprio Coase. Sendo ele nem um economista moderno no que diz respeito aos métodos formais nem ciente de que estava apresentando uma ideia que seria rotulada como um "teorema," sua análise exibe uma flexibilidade que a torna sujeita a múltiplas interpretações e, como veremos, a uma ampla gama de críticas.

Dadas essas suposições, Coase afirmou duas coisas. Primeiro, a alocação de recursos que emerge será eficiente, no sentido de maximizar o valor da produção. Chamaremos isso de "afirmação de eficiência." Em segundo lugar, a decisão sobre a quem os direitos serão inicialmente atribuídos não afetará a alocação final de recursos. Chamaremos isso de "afirmação de invariância."

A interpretação subsequente de Stigler sobre a descoberta de Coase, que ele codificou como o "teorema de Coase," apareceu na terceira edição de seu *Theory of Price* (1966). Foi expressa de forma muito mais concisa do que a formulação original de Coase, lembrando tanto a discussão sobre externalidades na literatura das décadas de 1940 e 1950, que ele havia tratado com certa profundidade em edições anteriores de seu texto sobre teoria dos preços, quanto o primeiro teorema fundamental da economia do bem-estar:

> *O teorema de Coase ... afirma que, sob competição perfeita, os custos privados e sociais serão iguais* (Stigler 1966)

Nas mãos de Stigler, nenhuma suposição explícita além da competição perfeita era necessária. Os motivos por trás da decisão de Stigler de codificar o resultado de Coase como o "teorema de Coase" e de expressá-lo como fez são desconhecidos, e até mesmo o extenso arquivo de Stigler na Universidade de Chicago não oferece pistas. Mas duas possibilidades se destacam. Stigler estava obviamente encantado com o resultado de Coase, como deixou claro em múltiplos comentários subsequentes – chegando a rotular Coase como um Arquimedes moderno. Sua decisão de aplicar o rótulo de "teorema" pode, portanto, ser nada mais do que um floreio retórico provocativo de Stigler. Mas é mais provável que houvesse um método na "loucura" de Stigler – um desejo de elevar o resultado de Coase ao nível de um corolário do Primeiro Teorema Fundamental da Economia do Bem-Estar, que explicitamente assumia a eliminação dos efeitos externos. Lido dessa forma, o teorema de Coase não era tanto uma prescrição para lidar com externalidades quanto uma justificativa para não se preocupar com elas, já que as forças da competição frequentemente eliminariam esse impedimento à eficiência.

As declarações de Coase e Stigler foram tanto o ponto de partida para subsequentes reformulações do teorema quanto sugestivas de um dos contrastes fundamentais encontrados na literatura do teorema de Coase—o quadro mais amplo dentro do qual o teorema está situado. Coase postulou um cenário de negociação com poucos participantes – cada uma de suas ilustrações lida com apenas dois agentes—enquanto a formulação de Stigler, se não a sua ilustração real do livro didático, sugere um teorema que é uma proposição na teoria dos mercados competitivos. Assim, encontramos o escopo prospectivo do teorema de Coase se estendendo do fazendeiro e do criador de gado até um sistema em grande escala de permissões de poluição dentro de um sistema walrasiano. As implicações potencialmente muito diferentes desses dois quadros para as estratégias de modelagem empregadas e para as conclusões alcançadas têm pesado bastante nos debates sobre a validade do teorema.

O teorema de Coase foi enunciado de várias maneiras, algumas delas permutações da declaração de Coase e outras da de Stigler. A conclusão mais importante a ser tirada disso é que não há consenso sobre o que o teorema de Coase realmente afirma. As declarações do teorema apresentadas a seguir podem parecer semelhantes a um exame casual, mas refletem suposições, resultados e ênfases diferentes entre a eficiência e a invariância que estão no cerne dos debates sobre a validade e aplicabilidade do teorema. Embora o foco na lista de teoremas que segue seja nas declarações originais de tipos específicos do teorema de Coase, contrapartes contemporâneas podem ser identificadas facilmente na literatura. Itálico foi adicionado, exceto onde indicado, para enfatizar elementos originais ou únicos nas declarações específicas dos teoremas.

Essa sequência de teoremas de Coase reflete uma variedade de pressupostos subjacentes que supostamente sustentam o teorema e afirmações contrastantes sobre o que o teorema afirma. Devemos, portanto, dedicar alguma atenção para explorar esses diversos pressupostos e afirmações antes de voltarmos nossa atenção para as controvérsias que geraram muitas dessas declarações contrastantes do teorema.

Muita da ambiguidade, confusão e controvérsia em torno do teorema de Coase é um artefato da colisão entre a abordagem mais solta e intuitiva de Coase sobre o assunto e a ênfase crescente da profissão em modelagem formal. Muitas das reinterpretações subsequentes podem, então, ser vistas como tentativas de apertar os pressupostos subjacentes do teorema ou de esclarecer melhor o que Coase deve ter tido em mente (na visão dos comentaristas) ao apresentar seu resultado.

Como nossa sequência demonstra, comentaristas subsequentes adicionaram uma variedade de pressupostos adicionais à formulação de Coase. A reinterpretação de Calabresi de 1968, que rapidamente se tornou uma referência tanto na economia quanto na literatura jurídica, adicionou o pressuposto de racionalidade, e outros seguiram rapidamente, seja explicitamente em suas declarações do teorema ou em suas análises dele. O pressuposto de racionalidade, no entanto, tem implicações importantes para os desafios teóricos e comportamentais ao teorema. Outros pressupostos, como os que se referem aos efeitos de renda (ou riqueza ou bem-estar) e conjuntos de produção/utilidade convexos, foram adicionados como resultado dos debates sobre a validade do teorema—embora, como ficará claro na próxima seção, não haja uma conclusão definitiva quanto à sua necessidade.

A declaração de Calabresi pode ser vista como uma abordagem mais expansiva e similar a um teorema sobre o original de Coase. A versão Stigleriana do teorema de Regan, exposta em sua crítica frequentemente citada ao argumento de Coase, tanto introduz um requisito de informação quanto vincula explicitamente o teorema à teoria do equilíbrio competitivo. Enquanto isso, Hoffman e Spitzer assumem tanto mercados competitivos quanto dois agentes para uma negociação (além de fornecer a declaração mais longa do teorema na literatura). A tensão entre o ambiente competitivo e o exemplo de barganha de pequenos números de Coase formou a base dos ataques de Regan e de muitos outros ao teorema, como veremos.

O conteúdo atribuído ao pressuposto de custos de transação zero esteve no centro de várias controvérsias sobre o teorema. Nem todos, no entanto, sentiram-se compelidos a restringir o teorema de Coase a um mundo assim. A reinterpretação inicial e amplamente citada de Coase por Ralph Turvey informou o leitor de que os agentes negociarão acordos eficientes na presença de externalidades, desde que "sejam capazes e dispostos a negociar para seu mútuo benefício" (Turvey 1963), um sentimento refletido nas declarações acima citadas de Miller, Cooter e Ulen, e Dixit e Olson. Sugeriu-se que o teorema se mantém se os custos de transação forem "negligenciáveis" (Worcester e Jr. 1972), "suficientemente baixos" (Baird 1975), "relativamente sem custo" (Miller 1978), "zero ou próximo de zero" (Beckmann e Wesseler 2007), ou sempre que os ganhos mútuos potenciais "excedam [os] custos de negociação necessários" (Nicholson 1989). Essa permissão para pequenos, mas positivos, custos de transação tem sido particularmente prevalente na literatura de manuais (Medema 2015c).

Nossa sequência de teoremas de Coase não fornece mais acordo sobre os resultados afirmados ou sobre a mensagem real do teorema do que sobre os pressupostos.

Coase, como já notamos, fez reivindicações tanto sobre eficiência (maximização do valor da produção) quanto sobre invariância alocativa. Muitas declarações do teorema replicam as duas reivindicações de Coase – às vezes referidas como a versão “forte” do teorema de Coase – mas outras, como Calabresi e Dixit e Olson, afirmam apenas a proposição de eficiência (o teorema “fraco” de Coase). Para outros, no entanto, o resultado verdadeiramente novo é a reivindicação de invariância, e assim encontramos declarações, como as de Hurwicz e Allen (assim como Lazear 2000), nos informando que o teorema abrange apenas a reivindicação de invariância, com eficiência geralmente assumida. E embora tenha sido quase universalmente admitido – como Coase fez – que a distribuição de renda variará dependendo de a quem os direitos são inicialmente atribuídos, Schwab nos forneceu um teorema que afirma que a alocação e a distribuição não são afetadas.

Essa diferença de ênfase entre eficiência e invariância deve-se em parte à crença amplamente compartilhada de que a reivindicação de invariância é incorreta como questão de lógica econômica. Mas também tende a ser uma função dos interesses daqueles que utilizam o teorema. No debate sobre os méritos relativos das soluções negociadas e pigouvianas, que foi particularmente importante durante as primeiras duas décadas e mais do teorema de Coase, a eficiência estava no centro da discussão. Para muitos estudiosos do direito, em contraste, e particularmente antes que a análise econômica ocupasse um lugar proeminente no pensamento jurídico, os efeitos invariantes das regras legais foram o *insight* revolucionário.

Isso nos leva a uma característica final que emerge de nossa sequência de teoremas de Coase—o domínio do teorema. A reinterpretação de Calabresi o estende além das externalidades para todas as falhas de mercado relacionadas à eficiência, e outros, incluindo Foss e Foss, e Foros e Hansen, veem o teorema como uma proposição geral na exploração de ganhos de troca. A implicação para o lado da invariância é uma mais geral de irrelevância institucional. Isso, combinado com a atenção crescente dada aos efeitos de estruturas institucionais alternativas, nos ajuda a entender a ênfase crescente colocada pelos economistas na reivindicação de invariância nos últimos anos.

Uma interpretação dessa variedade de teoremas de Coase é que muitos economistas simplesmente não entendem ou não entendiam o teorema de Coase. Mas isso perde o ponto histórico, pois nunca houve um "teorema de Coase" singular a ser compreendido – um fato que, por si só, distingue o teorema de Coase de outros teoremas em economia. A falta de qualquer declaração geralmente aceita do teorema—quanto aos pressupostos ou aos resultados – teve um papel importante em estimular as controvérsias sobre ele e na natureza do debate de vai-e-vem tanto sobre a validade teórica do teorema quanto sobre sua relevância.

Embora algumas vozes questionando o teorema de Coase tenham sido ouvidas durante a década de 1960, foi a década de 1970 que trouxe uma explosão de controvérsia sobre o teorema de Coase – uma controvérsia que continua, embora um pouco diminuída, até hoje. Os primeiros anos da controvérsia apresentaram uma série de debates, que se estenderam por cerca de duas décadas nas principais revistas da profissão, sobre a validade do teorema como uma proposição em lógica econômica. A progressão típica aqui foi a de "desprova" por oponentes do teorema, muitas vezes com uma defesa acompanhada de abordagens pigouvianas, seguida por tentativas dos apoiadores do teorema de defender o teorema contra a suposta desprova –geralmente afirmando mostrar o erro da desprova em questão, embora às vezes modificando o próprio teorema. Nos anos mais recentes, no entanto, a natureza da discussão mudou um pouco, com parte do foco mudando para a derivação de condições sob as quais o teorema pode ser mostrado como válido e aquelas sob as quais não pode. Dado que muitos dos argumentos contra o teorema continuam a ser discutidos na literatura após as refutações dessas críticas terem sido oferecidas, é importante lidar com as várias posições. Como veremos, muito disso tem a ver com a definição de custos de transação e a natureza da vida em um mundo onde eles estão ausentes. E como também veremos, essa análise da história do teorema ajuda a esclarecer as condições sob as quais as reivindicações de eficiência e invariância do teorema podem ser sustentadas e assim nos levará ao ponto em que podemos declarar um teorema de Coase válido e útil.

Alguns dos desafios mais complicados que enfrentam o teorema de Coase envolvem situações em que os efeitos do lado do consumidor entram em cena. As complicações potenciais introduzidas aqui são várias. Primeiro, como Buchanan e Stubblebine (1962) apontaram cedo e Hovenkamp (1990) enfatizou mais recentemente, a não comparabilidade das funções de utilidade impede qualquer reivindicação forte sobre invariância. Além disso, Hovenkamp observou, se os consumidores maximizam a utilidade em vez de riqueza, não há garantia de que a reivindicação de Coase para maximização da utilidade seja mantida. O foco nos efeitos distributivos e a identificação de efeitos normativos relevantes tornam as comparações de utilidade entre indivíduos um problema central na análise econômica.

Em termos mais gerais, existe o problema de que os consumidores em uma economia podem ter diferentes preferências e disposições a pagar que se tornam relevantes ao considerar negociações de externalidades. Portanto, enquanto os efeitos alocativos de um mercado competitivo podem ser previsíveis, as trocas privadas de externalidades em um mundo de consumidores com diferentes disposições a pagar e níveis de bem-estar podem introduzir complexidades adicionais e tornar as implicações do teorema de Coase mais difíceis de analisar ou aplicar. Essas são as questões que a seção 4 abordará em mais detalhes, ao examinar a lógica econômica e as controvérsias em torno da validade e aplicabilidade do teorema de Coase em relação aos custos de transação, eficiência e invariância.

Um segundo conjunto de questões surge devido aos efeitos diferenciais de renda que podem acompanhar atribuições alternativas de direitos, tanto sob condições de barganha (Dolbear 1967) quanto de competição (Hurwicz 1995). Coase, por sua vez, mais tarde afastou essas objeções com base no argumento de que os efeitos de renda “normalmente serão tão insignificantes que podem ser desprezados com segurança” mas “normalmente” não é suficiente para resgatar a proposição de invariância de um teorema. Que pressupostos seriam necessários para validar a alegação de invariância aqui e, assim, salvar um (forte) teorema de Coase nesse contexto? Dolbear (1967) sugeriu que preferências paralelas (utilidade quase-linear) excluiria esses problemas, um resultado que foi formalizado posteriormente por Hurwicz (1995) e refinado ainda mais por, por exemplo, Chipman e Tian (2011). A presença de bens públicos na relação – por exemplo, crianças em um contexto de casamento/divórcio – introduz uma complicação adicional (Zelder 1993), e aqui, como Chiappori (2010) e Chiappori, Iyigun e Weiss (2015) mostraram, a utilidade transferível em todos os ambientes institucionais relevantes é necessária para garantir a invariância. Bergstrom (2017) recentemente generalizou vários dos resultados mencionados anteriormente ao demonstrar que a invariância ocorre desde que as fronteiras de possibilidade de utilidade sejam paralelas, o que será o caso se as funções de utilidade forem uniformemente afins em bens privados. A natureza bastante restritiva dessas condições sugere uma limitação significativa no escopo da proposição de invariância, embora isso seja mitigado um pouco pela descoberta de Russell (1995) de que uma suposição de preferências heterogêneas salva a invariância em um ambiente competitivo, independentemente da forma das preferências individuais—pelo menos onde um grande número garante diversidade suficiente (e, portanto, heterogeneidade).

Os defensores do teorema enfatizam que a crítica dos efeitos de renda não se aplica às alterações na estrutura existente de direitos. Os argumentos aqui são dois. Primeiro, com direitos de propriedade totalmente especificados, uma alteração na responsabilidade não pode ocorrer sem uma compensação completa; caso contrário, os direitos não foram totalmente especificados em primeiro lugar, violando o que é geralmente considerado uma das premissas centrais do teorema. Com essa compensação sendo paga, a distribuição de riqueza não é afetada. Isso também elimina a crítica de que o teorema falha em considerar os interesses das gerações futuras. Segundo, em um mundo de custos de transação zero, o impacto potencial de uma redistribuição de direitos será totalmente considerado em contratos e/ou capitalizado nos valores dos recursos, deixando a distribuição de riqueza inalterada e não fornecendo espaço para efeitos de renda.

O terceiro desafio ao teorema de Coase nesse aspecto, inicialmente levantado por Mishan (1965), refere-se à preocupação de que o valor que os indivíduos atribuem aos direitos pode ser uma função da propriedade – como quando o valor que uma vítima de poluição está disposta a aceitar (WTA) em pagamento por permitir que o poluidor contamine o seu ar é maior do que o valor que ela está disposta a pagar (WTP) para induzir uma redução nas emissões. O preço pelo qual um acordo é feito provavelmente varia com a atribuição de direitos, dando origem a diferentes níveis de produção e externalidades (Pareto-ótimos) e negando a invariância. Essas divergências WTA/WTP podem ocorrer por uma variedade de razões, incluindo a utilidade marginal decrescente da renda quando os agentes negociam sobre utilidade em vez de riqueza per se (Hovenkamp 1990), possibilidades mínimas de substituição (Hanemann 1991) e a endogeneidade dos gostos e preferências dos consumidores (Kahneman e Tversky 1979; Samuelson e Zeckhauser 1988; Thaler 1980). De particular preocupação aqui são os efeitos de dotação, que podem gerar menos negociação de direitos do que o postulado pelo teorema de Coase e, no limite, a falha em concluir qualquer acordo. Embora o Paretiano não possa olhar com desdém para tais resultados, a alegação de invariância claramente perde toda a sua força na presença de tais divergências.

Um quarto problema surge em situações nas quais um ou mais agentes têm renda/riqueza insuficiente para pagar um suborno que gere eficiência. Isso não representa um problema quando os agentes negociam sobre riqueza, em vez de utilidade, já que a riqueza aumentará mais do que o valor do suborno e os agentes poderiam pegar emprestado, se necessário, para financiar o suborno em um mundo de custos de transação zero. No entanto, valores subjetivos apresentam uma dificuldade maior, e a afirmação de Shavell (2009) de que a invariância é provável de se manter, ou pelo menos aproximadamente assim se o valor subjetivo do direito não for grande em relação aos ativos das partes, novamente nos afasta do domínio dos teoremas.

Então, o que devemos fazer com as implicações de tudo isso para o teorema de Coase? Uma abordagem seria impor suposições adicionais – por exemplo, racionalidade e restrições apropriadas sobre preferências para excluir esses efeitos. Outra abordagem é inserir uma qualificação dos efeitos renda, uma solução vista em várias declarações do teorema que aparecem em nossa liturgia. Uma terceira resposta tem sido declarar (ou insistir) o teorema sem a tese da invariância. No entanto, essa solução rouba do teorema o que muitos consideram seu *insight* central – que a atribuição inicial de direitos não impacta a alocação de recursos.

O problema do carona é o reverso da questão da extorsão e torna-se relevante quando saímos do quadro de negociação de dois agentes proposto por Coase. Essa possibilidade, também introduzida por Wellisz (1964), foi levantada repetidamente nos anos seguintes e recentemente abordada de forma mais formal por, por exemplo, Dixit e Olson (2000) e Ellingsen e Paltseva (2016). Se aqueles afetados pela poluição precisam pagar ao poluidor para reduzir as emissões, um pagamento de A ao poluidor resulta em emissões reduzidas que beneficiam B, C, D, ..., bem como A. É do interesse de B aproveitar o pagamento de A, C, D, ..., beneficiando-se do ar limpo sem precisar pagar por isso. Mas, como cada agente enfrenta o mesmo incentivo, os pagamentos totais ao poluidor ficarão aquém do nível necessário para gerar a quantidade ótima de poluição e de produção poluidora. Parisi (1995) considera essas situações de carona as mais recidivas ao antídoto coasiano e Baliga e Maskin (2003) nos dizem que mesmo um coasiano convicto deve concordar que o teorema de Coase não se aplica nessas circunstâncias.

Baliga e Maskin (2003) observam, corretamente, que o teorema de Coase requer excludibilidade, para que o problema de carona não se manifeste. A questão, então, é se isso exige uma suposição adicional. A resposta aqui, como nas questões de entrada e extorsão, reside na distinção entre classes abertas e fechadas (Holderness

1989). Se A está comprando de B o direito de estar livre de danos, esse direito não está completo a menos que inclua a capacidade de excluir outros de seus benefícios. Assim, a presença de carona viola a suposição do teorema de direitos de propriedade totalmente especificados. Mas há uma segunda linha de ataque contra o argumento do carona, lançada, ironicamente, por dois dos críticos mais severos do teorema – Mishan (1967b) e Dick (1974) – durante os primeiros estágios do debate. Eles argumentaram que o carona é uma manifestação de custos de transação positivos – é uma informação custosa que possibilita a revelação incompleta de preferências – e, portanto, não é um argumento legítimo contra o teorema de Coase. Isso nos leva ao que talvez seja a questão mais controversa na literatura sobre o teorema de Coase – as implicações da informação privada para a validade do teorema.

Os problemas gerados pela incerteza são bem ilustrados por Cooter (1982), que aponta que a estratégia de maximização da utilidade esperada que é ótima contra a distribuição das estratégias de um oponente pode não ser ótima contra a estratégia realmente jogada – dando origem a acordos negociados ineficientes quando expectativas e realidade divergem. Uma preocupação adicional resulta das implicações do teorema de Myerson-Satterthwaite (1983) para as alegações de eficiência do teorema de Coase. O que Myerson e Satterthwaite demonstraram, de forma resumida, é que, para um bem indivisível, não há um equilíbrio de Bayes-Nash eficiente quando agentes racionais possuem informação privada. As dificuldades que isso apresenta para o teorema de Coase foram notadas por Samuelson (1985) e por Fudenberg e Tirole (1991), e ampliadas por McKelvey e Page (1999; 2002). A generalização de Myerson-Satterthwaite por McKelvey e Page revela que a capacidade dos agentes de empregar estrategicamente informação privada irá enviesar a solução negociada na direção do detentor dos direitos de propriedade, significando que haverá um nível de poluição ineficientemente alto quando os poluidores são designados aos direitos relevantes e um nível de poluição ineficientemente baixo (um nível ineficientemente alto de mitigação) quando as vítimas possuem os direitos. Com base nisso, McKelvey e Page

(1999, 246) oferecem um "teorema de Coase com informação privada":

> *Teorema de Coase (Informação Privada): Para dois jogadores com preferências quase-lineares [e] informação privada, em qualquer jogo não cooperativo onde os direitos de propriedade estão definidos e aplicados, não existe nenhum equilíbrio de Bayes-Nash que seja totalmente eficiente e o equilíbrio de Bayes-Nash mais eficiente exibe um viés nos resultados a favor da parte que é atribuída os direitos de propriedade.*

Seria difícil formular uma versão do teorema mais contrária ao original de Coase. A existência de informação privada não implica inevitavelmente no fracasso de barganhas do tipo teorema de Coase. Maskin (1994) argumenta que o poder monopolista está na raiz dos problemas de informação identificados por Farrell (1987). Se nenhum agente é grande o suficiente para ter poder de mercado significativo e os agentes redigem contratos que maximizam o excedente, eles negociarão um acordo eficiente. Isso, afirma Maskin, dá suporte a uma postura bastante laissez-faire em relação às externalidades (1994). Gomes e Jehiel (2005), por sua vez, demonstram que quando os agentes são capazes de elaborar contratos de longo prazo sobre ações presentes e futuras, o equilíbrio resultante será eficiente e, se os jogadores forem suficientemente pacientes, independente do ponto de partida. Uma abordagem muito diferente para o problema vem de Schmitz (2001), que mostra que o teorema de Coase pode se manter válido com informação privada quando os direitos de propriedade não foram atribuídos. A intuição aqui é que o incentivo para exagerar as avaliações diminui quando os direitos de propriedade são incertos, uma vez que os agentes não sabem se serão compradores ou vendedores desses direitos antes do julgamento. Isso, por sua vez, aumenta a probabilidade de que as partes alcancem uma solução eficiente sem recorrer ao julgamento. O problema com todas essas descobertas, claro, é que elas carecem da generalidade geralmente associada ao teorema de Coase.

As ineficiências resultantes da informação privada devem-se, em última análise, ao comportamento estratégico na disputa pelos excedentes da negociação. Coase (1988b, 161), por sua vez, considerava o comportamento estratégico como não importante. Em seu mundo, os agentes estão dispostos a uma divisão razoável dos ganhos da troca; ninguém vai ameaçar rasgar a nota de $100 que o grupo encontrou na calçada. Cooter, por outro lado, considera essas preocupações estratégicas como um obstáculo quase intransponível para resultados negociados que aumentem a eficiência, defendendo um "teorema de Hobbes" – que os agentes nunca concordarão com uma distribuição do excedente – que é talvez tão forte quanto o do teorema de Coase (Cooter 1982). Hirshleifer, da mesma forma, contrasta o teorema de Coase (as pessoas nunca irão perder uma oportunidade de cooperar por meio de troca mutuamente vantajosa) com o que ele chama de "teorema de Maquiavel" (ninguém jamais perderá uma oportunidade de obter uma vantagem unilateral explorando outra parte). Ambos, diz ele, são verdades parciais e, na realidade, os agentes encontrarão uma posição ótima entre esses dois (Hirshleifer 1994).

5 Incidência de Impostos

The adage 'free as air' has become obsolete by Act of Parliament. Neither air nor light have been free since the imposition of the window-tax. We are obliged to pay for what nature lavishly supplies to all, at so much per window per year; and the poor who cannot afford the expense are stinted in two of the most urgent necessities of life.

Charles Dickens (1850, p. 461)

5.1 Introdução

Idealmente, a análise de política tenta considerar políticas completamente especificadas, ter uma visão abrangente do problema em questão no que diz respeito a instrumentos potencialmente úteis e apresentar e avaliar alternativas de maneira comparável. Esses recursos são particularmente importantes no desenvolvimento uma teoria da tributação e no exame de assuntos pertinentes à economia do setor público.

Completude significa que uma política deve ser totalmente articulada em todos aspectos pertinentes. Um requisito importante no cenário atual é equilíbrio orçamentário[33]. Apesar de sua familiaridade, seus ditames às vezes são esquecidos, o que pode desviar a análise – por exemplo, omitindo receita e efeitos de substituição de despesas que, em muitos casos, estão em oposição aos dos impostos que os financiam, e também por ignorar consequências distributivas das despesas. Além disso, ao examinar qualquer imposto ou política de despesas, existe uma variedade de maneiras de tornar a especificação completa, cada uma das quais pode ter implicações diferentes para a extensão da redistribuição e outras relevantes para o bem-estar considerações. Portanto, orientações adicionais na escolha de como preencher o sistema é necessário.

A abrangência indica a necessidade de considerar todos os instrumentos de política pertinentes. Normalmente, não se usaria uma chave de fenda ou uma faca para martelar pregos se houvesse um martelo. Da mesma forma, ao considerar como alguém pode empregar a tributação de bens e doações para aumentar a receita ou aumentar a redistribuição, a disponibilidade do imposto de renda deve ser mantida em mente. Normalmente, é melhor usar a ferramenta mais diretamente adequada para a tarefa em questão. Além disso, quando outros mecanismos também estão em uso, deve-se atentar para as interações que podem passar despercebidas ao focar em um único instrumento de política. Notavelmente, muitas políticas interagem com o imposto de renda no que diz respeito aos efeitos sobre a oferta de trabalho e a

[33] Veja, por exemplo, a discussão de Musgrave (1959, pp. 212-215) sobre “incidência tributária diferencial” e considere também a noção mais ampla de “incidência de orçamento equilibrado”.

distribuição de renda. Por essas e outras razões, não podemos avaliar a tributação da receita de capital independentemente de outros instrumentos fiscais que aumentem a receita de pessoas físicas na mesma parte da distribuição de renda.

A comparabilidade refere-se à ideia de que é mais fácil avaliar as escolhas entre diferentes tipos de maçãs do que entre maçãs e laranjas (ou, pior, maçãs e elefantes). Ao avaliar uma determinada política, geralmente é mais útil dar foco a uma dimensão de cada vez e, além disso, distinguir características intrínsecas das incidentais. Ao comprar um automóvel, não é muito útil comparar um SUV branco, superdimensionado, a um carro subcompacto vermelho, econômico. Para os compradores com orçamento apertado, escolher o SUV porque odeiam a cor vermelha ou porque ele vem com um forno de micro-ondas grátis que por acaso atende a uma necessidade atual seria tolice. Em vez disso, é mais esclarecedor realizar uma série de comparações de veículos que são semelhantes em todas as dimensões, exceto em uma dimensão, por exemplo, investigar subcompactos em outras cores se alguém odeia vermelho e visitar uma loja de eletrodomésticos se um novo forno de micro-ondas for desejado.

Essa noção de senso comum de comparabilidade recebeu muito pouca atenção na teoria da tributação. A comparabilidade é na verdade uma ideia extremamente poderosa. Em particular, verifica-se que em uma variedade surpreendentemente ampla de casos, a comparabilidade é mais nítida quando se completa o sistema – digamos, um plano de imposto sobre bens de luxo ou uma proposta de despesa em parques – usando uma técnica particular: um ajuste no imposto de renda (e transferência) que atinge um resultado geral neutro em termos de distribuição. Na verdade, a disponibilidade do imposto de renda como um instrumento lança uma luz diferente sobre a análise de políticas como impostos sobre commodities, impostos sobre dividendos e ganhos de capital, impostos sobre propriedades e doações, seguro social, provisão de bens públicos e regulamentação econômica. Acontece que geralmente não é sensato usar várias formas indiretas de tributação, políticas de despesas ou regulamentos para redistribuir a renda se houver um imposto de renda disponível, como geralmente acontece nas economias desenvolvidas. Entre os muitos instrumentos de política geralmente pensados em termos distributivos ou de arrecadação, o imposto de renda ocupa um lugar especial e, por causa desse papel, os demais precisam ser analisados de forma diferente de como costumam ser.

Juntos, completude, abrangência e comparabilidade são aspectos essenciais de uma visão integrada de tributação, gastos do governo e redistribuição. Só podemos compreender cada instrumento de política – cada peça do quebra-cabeça – se as outras também estiverem na mesa e as relações entre elas forem compreendidas. A maioria das análises é muito mais especializada, concentrando-se em uma política específica, na verdade, frequentemente em um único aspecto de uma política específica, e há boas razões para essa divisão de trabalho. No entanto, a pesquisa é mais bem orientada e seus resultados são empregados de forma mais eficaz pelos formuladores de políticas se a estrutura mais ampla e integrada da qual fazem parte for bem compreendida e mantida claramente em vista.

5.2 Sistema Tributário

Os princípios de justiça tributária, simplicidade e eficiência econômica são atributos desejáveis em qualquer sistema tributário. O conceito de justiça tributária está relacionado com equidade entre os agentes econômicos da sociedade. Na teoria de tributação ótima, a justiça social está associada ao bem-estar da sociedade como uma função de utilidades individuais.

Um dos principais problemas de um sistema tributário é a diferença existente entre os indivíduos com relação a uma série de fatores, em particular quanto à dotação de recursos e suas preferências. Essas são características relevantes para determinação de tributos, mas são informações privadas e que não são perfeitamente reveladas na economia. O sistema tributário deveria levar em conta as diferenças entre as preferências dos agentes econômicos. Se a observação dessas últimas em cada indivíduo fosse possível, com custo zero e fosse feita de forma perfeita, o governo poderia utilizar o *lump sum tax*, isto é, um imposto de montante fixo, único imposto que não gera ineficiência na alocação de recursos da economia. O *lump sum tax* é eficiente no sentido de o produto de sua arrecadação independer do comportamento do agente econômico e depender de características do indivíduo que, em princípio, não podem ser alteradas.

O arcabouço teórico usado na literatura da tributação ótima traduz-se pela modelagem dos efeitos ocasionados pela tributação no comportamento dos agentes econômicos de modo que seja consistente com a especificação das utilidades e a análise das consequências desse comportamento.

O conceito de utilidade individual e de bem-estar social é de grande importância para a análise da teoria da tributação ótima. Bem-estar social é um indicador do bem-estar da sociedade e depende das utilidades individuais. O primeiro passo para o cálculo do imposto ótimo é obter a função de utilidade do agente econômico que pode depender dos bens de consumo e da oferta de trabalho ou da renda como um todo e da oferta de trabalho. No primeiro caso, um sistema completo de demanda precisa ser estimado.

A impossibilidade da verificação das características inerentes a cada um dos agentes econômicos torna inevitável a utilização de impostos distorcivos, o que impede de se ter uma economia com eficiência de Pareto – situação em que um agente não pode melhorar sem que o bem-estar de outro piore. A teoria da tributação preocupa-se, portanto, com a escolha de características “facilmente” observáveis como base de tributação que estejam associadas de forma sistemática às características não-observáveis e nas quais há o real interesse em se tributar [Atkinson e Stiglitz (1976, p. 56)].

A eficiência econômica está relacionada com as distorções que um sistema tributário provoca no comportamento dos agentes econômicos. Um sistema tributário é dito eficiente quando a alocação de recursos é feita de modo a minimizar a interferência nas decisões econômicas dos agentes. É importante ter se em mente que as condições necessárias que caracterizam alocações eficientes de recursos no sentido de Pareto raramente são satisfeitas. Portanto, usualmente, a análise da tributação ótima centra-se na teoria do *second best*, que fundamenta a formulação de políticas do governo em situações de impossibilidade da remoção de algumas distorções existentes na economia.

A simplicidade de um sistema tributário é avaliada pelos custos administrativos que podem ser diretos ou indiretos. Os custos administrativos diretos são aqueles necessários para o funcionamento do sistema e são, geralmente, arcados pela Receita Federal. Os custos indiretos são os arcados pelos contribuintes e podem assumir diversas formas: preenchimento de formulários dos impostos, custos de advogados e contadores, entre outros.

Os modelos de tributação ótima utilizam a análise econômica para estudar a combinação dos três critérios citados anteriormente: equidade, simplicidade e eficiência econômica. Geralmente, a questão da simplicidade fica em segundo plano devido à dificuldade na modelagem da relação entre as alíquotas tributárias e os custos administrativos. Essa negligência é uma das maiores limitações nos modelos de tributação ótima. O que se questiona com maior frequência na literatura de tributação ótima é o *trade-off* entre equidade e eficiência na economia.

Qual a melhor estrutura tributária para alcançar, de forma simultânea, os objetivos de redistribuição de renda e eficiência econômica do governo? O que é melhor para a sociedade, tanto em termos de política redistributiva quanto de eficiência? O modelo de Ramsey (1927) e sua extensão para uma economia com muitos agentes, apresentada por Diamond e Mirrlees (1971), são de extrema relevância para a teoria da tributação ótima sobre o consumo, pois tentam responder a tais questões. No primeiro modelo, que trata unicamente de questões de eficiência na economia, a alíquota de imposto sobre um bem qualquer guarda relação inversa com a sua elasticidade-preço da demanda. A análise de Diamond e Mirrlees modifica o esboço do imposto ótimo sobre bens pois aspectos sobre equidade são levados em conta na economia.

Nesse caso, a estrutura das alíquotas é determinada pela seletividade, de acordo com a ponderação dada aos bens consumidos pelos mais pobres.

O sistema tributário influi na distribuição de renda a partir do momento em que se utiliza de diversos instrumentos de arrecadação que vão impactar de modo diferenciado cada contribuinte. Ao dispor de tributos que incidem sobre diferentes fatos econômicos (como a posse ou a transferência de bens, a aferição de renda, o consumo ou a poupança), o modo como o Estado organiza o seu sistema tributário irá impactar cada agente de modo particular, a depender de seu perfil econômico em cada tipo de tributação pertinente.

A incidência tributária é o estudo dos efeitos das políticas tributárias sobre os preços e a distribuição de utilidades/bem-estar. O que acontece com os preços de mercado quando um imposto é introduzido ou alterado? Exemplos:

- O que acontece quando se impõe um imposto de R$ 1 por pacote de cigarros?
- O que acontece quando se introduz um subsídio para produtores?
- Efeito sobre o preço: efeitos distributivos sobre os fumantes, para os produtores, acionistas, agricultores etc.

Esta é uma análise positiva: tipicamente o primeiro passo na avaliação de políticas; é uma entrada para pensar mais tarde sobre qual política maximiza o bem-estar social. A análise empírica é uma grande parte desta literatura porque a própria teoria é amplamente inconclusiva sobre magnitudes, embora informativa sobre sinais e estática comparativa. A incidência de impostos não é um exercício contábil, mas uma caracterização analítica de mudanças nos equilíbrios econômicos quando os impostos são alterados. Ponto-chave: os impostos podem ser transferidos. Os impostos afetam diretamente os preços das mercadorias, que afetam as quantidades por causa de respostas comportamentais, que afetam indiretamente o preço de outros bens. Conhecer a incidência é fundamental para a análise de políticas. Idealmente, queremos conhecer o efeito de uma mudança de impostos nos níveis de utilidade de todos os agentes da economia. Normalmente, costumamos olhar para os impactos nos preços ou na renda, em vez da utilidade. Uma simplificação útil é agregar agentes econômicos em alguns grupos.

- imposto sobre a gasolina: produtores versus consumidores
- BF: destinatários versus não-recebedores
- imposto de renda: rico versus pobre
- imposto predial: região ou país
- seguridade social: através das gerações

Exemplo 5.1. *Considere uma economia em que nosso agente representativo tem preferências que podem ser representadas pela seguinte função de utilidade*

$$U(C_1, C_2) = C_1 C_2 \tag{5.1}$$

cuja dotação inicial é $(\overline{C}_1, \overline{C}_2) = (1, 0)$*. Suponha que o bem 2 pode ser produzido através do bem 1 a partir da seguinte tecnologia:* $f(C_1) = \dfrac{C_1}{a}$*.*

No equilíbrio de mercado o produtor tem lucro zero:

$$p_2 f(C_1) - 1C_1 = 0 \quad \Longrightarrow \quad p_2 = a \tag{5.2}$$

A partir disso, podemos construir a restrição do agente como $C_1 + aC_2 = 1$*, em que a renda do agente é a soma de suas dotações.*

Para encontrar as escolhas ótimas do agente, usamos o fato de que a taxa marginal de substituição é igual à razão de preços. Assim,

$$TMS_{C_1,C_2} \equiv \frac{UMgC_1}{UMgC_2} = \frac{p_1}{p_2}$$

$$\frac{C_2}{C_1} = \frac{1}{c} \quad \Longrightarrow C_1 = aC_2 \tag{5.3}$$

Substituindo esse resultado na restrição orçamentária que derivamos acima, encontramos:

$$\begin{aligned} C_1 + aC_2 &= 1 \\ aC_2 + aC_2 &= 1 \\ C_2 &= \frac{1}{2a} \end{aligned} \tag{5.4}$$

e $C_1 = \frac{1}{2}$. *Logo, a utilidade indireta*[34] *é* $U(C_1, C_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2a}\right) = \frac{1}{4a}$.

Agora considere um imposto t *sobre o bem* 2 *cuja arrecadação é totalmente devolvida para o consumidor na forma de um transferência lump-sum. A nova restrição orçamentária é*

$$C_1 + (a+t)C_2 = 1 + T \tag{5.5}$$

Para encontrar as escolhas ótimas do agente, usamos o fato de que $TMS = \frac{p_1}{p_2}$. *Assim,*

$$\frac{UMgC_1}{UMgC_2} \equiv \frac{C_2}{C_1} = \frac{1}{a+t} \tag{5.6}$$

Substituindo esse resultado na restrição orçamentária que derivamos acima, encontramos:

$$\begin{aligned} C_1 + (a+t)C_2 &= 1 + T \\ (a+t)C_2 + (a+t)C_2 &= 1 + T \\ C_2 &= \frac{1+T}{2(t+a)} \end{aligned} \tag{5.7}$$

[34] A função de utilidade indireta expressa o nível máximo de utilidade que um consumidor pode alcançar, dado o orçamento disponível e os preços dos bens.

e

$$C_1 = \frac{1+T}{2} \tag{5.8}$$

Considerando que $T = tC_2$, tem-se que:

$$\begin{aligned} C_2 &= \frac{1+tC_2}{2(t+a)} \\ 2(t+a)C_2 - tC_2 &= 1 \\ C_2(2a+t) &= 1 \\ C_2 &= \frac{1}{2a+t} < \frac{1}{2a} \end{aligned} \tag{5.9}$$

Observe que o consumo do bem tributado é menor do que na situação sem tributação.

Para o consumo do bem 1 temos:

$$\begin{aligned} C_1 &= \frac{1+T}{2} \\ &= \frac{1+tC_2}{2} \\ &= \frac{1}{2}\left[1 + t\left(\frac{1}{2a+t}\right)\right] \\ &= \frac{1}{2}\left(\frac{2a+t+t}{2a+t}\right) \\ &= \frac{a+t}{2a+t} > \frac{1}{2} \end{aligned} \tag{5.10}$$

Temos, portanto, que o consumo do bem 1 aumenta, e o consumo do bem 2 cai como consequência da imposição de um tributo sobre o bem 2.

Quanto ao bem estar, temos que

$$U(t) = \left(\frac{a+t}{2a+t}\right)\left(\frac{1}{2a+t}\right)$$

$$= \frac{a+t}{(2a+t)^2} \tag{5.11}$$

Portanto, a diferença de bem-estar entre a situação sem imposto e a situação com imposto é

$$U(0) - U(t) = \frac{1}{4a} - \frac{a+t}{(2a+t)^2} = \frac{t^2}{4a(2a+t)^2} \tag{5.12}$$

Há duas coisas a serem ressaltadas. Primeiro está o fato de que $U(0) > U(t)$: existe perda de bem-estar, pois $t^2 > 0$ para todo a. Em segundo está o fato de que esta perda é de segunda ordem[35]*. A derivada de $U - U(t)$ avaliada em $t = 0$ é zero.*

Veja que

$$\begin{aligned} \frac{\partial(U - U(t))}{\partial t} &= \frac{1}{4a}\left[\frac{2t(2a+t)^2 - 2t^2(2a+t)}{(2a+t)^4}\right] \\ &= \frac{1}{4a}\left[\frac{2t(2a+t)[(2a+t) - 2t]}{(2a+t)^4}\right] \\ &= \frac{t}{2a}\left[\frac{(2a-t)}{(2a+t)^3}\right]\Bigg|_{t=0} = 0 \end{aligned} \tag{5.13}$$

[35] Ao analisar a perda de bem-estar devido a uma política fiscal, como a introdução de um imposto, é comum investigar os impactos de primeira e segunda ordem.

- Impactos de primeira ordem: referem-se ao efeito direto e inicial da mudança na política (neste caso, o imposto) sobre o bem-estar do consumidor. Em termos simples, seria o impacto imediato da perda de poder de compra ou de consumo.
- Impactos de segunda ordem: referem-se aos ajustes subsequentes que o consumidor faz em resposta ao imposto, levando em consideração sua otimização de utilidade. Estes ajustes geralmente envolvem mudanças nas escolhas de consumo que minimizam a perda de bem-estar. Portanto, quando se diz que a perda de bem-estar é de segunda ordem, isso geralmente indica que o impacto direto do imposto é pequeno, e a maior parte do impacto se dá através dos ajustes subsequentes que o consumidor realiza. Tecnicamente, a perda de bem-estar de segunda ordem é proporcional ao quadrado da magnitude do imposto, o que sugere que a perda inicial (de primeira ordem) é compensada por esses ajustes otimizadores do consumidor.

e que

$$
\begin{aligned}
\frac{\partial^2(U-U(t))}{\partial t^2} &= \frac{1}{2a}\left[\frac{(2a+t)^3(2a-2t)-3t(2a-t)(2a+t)^2}{(2a+t)^6}\right] \\
&= \frac{1}{2a}\left[\frac{t^2+4a^2-8at}{(2a+t)^4}\right] \\
&= \frac{1}{2a}\left[\frac{t^2+4a^2-8at}{(2a+t)^4}\right]\Bigg|_{t=0} \\
&= \frac{1}{(2a)^3} > 0 \qquad (5.14)
\end{aligned}
$$

Queremos saber o quão geral são estes resultados. O que podemos falar em geral sobre incidência? E sobre perdas de bem-estar?

5.3 Análise de Equilíbrio Parcial da Incidência de Impostos

A incidência tributária é o estudo de quem suporta o ônus econômico de um imposto. Em termos gerais, é a análise positiva do impacto dos impostos na distribuição do bem-estar em uma sociedade. Começa com o *insight* muito básico de que a pessoa que tem a obrigação legal de efetuar um pagamento de imposto pode não ser a pessoa cujo bem-estar é reduzido pela presença do imposto. A incidência estatutária de um imposto refere-se à distribuição de pagamentos de impostos com base na obrigação legal de remeter impostos ao governo. Os economistas, com razão, concentram-se na incidência econômica, que mede as mudanças no bem-estar econômico da sociedade decorrentes de um imposto.

A incidência econômica difere da incidência estatutária por causa de mudanças no comportamento e consequentes mudanças nos preços de equilíbrio. Os consumidores compram menos de um produto tributado, e portanto, as empresas produzem menos e compram menos insumos - o que altera o preço líquido de cada insumo. Assim, o trabalho do analista de incidência é determinar como esses outros preços mudam e como essas mudanças afetam diferentes tipos de indivíduos.

As análises de incidência são abundantes na literatura, mas podem ser classificadas em algumas categorias. Em particular, quando esses estudos analisam os efeitos distributivos dos impostos entre os grupos, Atkinson e Stiglitz (1980) observam que nós economistas usamos cinco maneiras diferentes de dividir os contribuintes em grupos.

1. Primeiro, podemos nos concentrar no impacto dos impostos sobre os consumidores, em oposição aos produtores ou fornecedores de fatores (como mão-de-obra, capital e terra). Um diagrama de equilíbrio parcial pode identificar tanto a perda do excedente do consumidor quanto a perda do excedente do produtor resultante de um imposto.

2. Segundo, podemos restringir o foco para analisar o impacto de um imposto especificamente nas demandas relativas de diferentes fatores e nos retornos desses fatores (como capital, trabalho ou terra). A análise de equilíbrio geral pioneira de Harberger (1962) simplesmente ignora o lado do consumidor, assumindo que todo mundo gasta seu dinheiro da mesma maneira e, em seguida, ele deduz o ônus de um imposto sobre o capital em oposição ao trabalho.

3. Terceiro, podemos agrupar indivíduos por alguma medida de bem-estar econômico. Qualquer classificação desse tipo nos permite analisar a progressividade de um sistema tributário. Normalmente, os contribuintes são agrupados por alguma medida de renda e, em seguida, os dados nos dizem quanto cada grupo ganha de cada fator e quanto cada grupo gasta em cada produto.

4. Quarto, os impostos podem ser avaliados com base na incidência regional. Essa análise pode se concentrar em diferenças regionais dentro de um país ou pode se concentrar em diferenças internacionais.

5. Finalmente, os impostos podem ter efeitos intergeracionais. Por exemplo, a criação de um sistema de impostos e transferências que seja parcial ou totalmente

financiado por dívida trará uma transferência das gerações futuras para alguns ou todos os membros da geração atual.

Muitos dos princípios fundamentais da incidência de impostos podem ser ilustrados no ajuste de equilíbrio parcial mais simples. Portanto, começamos considerando a análise de equilíbrio parcial de um imposto sobre o consumo de um produto.

Na ausência de tributação, o equilíbrio é atingido quando oferta e demanda são iguais e a equação

$$D(p) = S(p), \tag{5.15}$$

é satisfeita.

Agora considere a introdução de um imposto sobre o consumo, τ. Se o imposto é recolhido dos consumidores, o novo equilíbrio deverá satisfazer a condição abaixo:

$$D(p + \tau) = S(p), \tag{5.16}$$

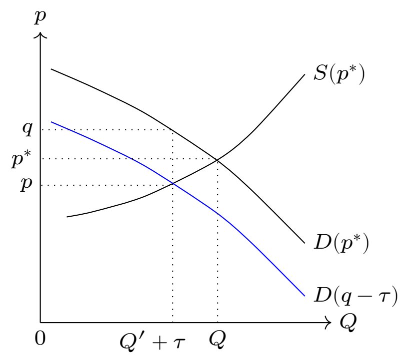
enquanto que se o imposto é coletado dos produtores ($q = p + t$), o novo equilíbrio deverá satisfazer a condição

$$D(q) = S(q - \tau). \tag{5.17}$$

Comparando as equações, é claro que a determinação da quantidade de equilíbrio, o preço pago pelos consumidores e a receita dos produtores não depende de que lado do mercado o imposto é cobrado. Diagramaticamente, a imposição de um imposto pode ser analisada como uma mudança na curva de demanda ou como uma mudança na curva de oferta. O equilíbrio resultante é o mesmo em ambos os casos. Este princípio, de que a incidência de um imposto não depende de que lado do mercado ele é cobrado, transita para contextos muito mais gerais.

Segue-se imediatamente deste princípio que a incidência final de um imposto

Figura 5.1 – *Incidência de um Imposto*



não pode ser avaliada simplesmente observando onde o imposto é cobrado de forma imediata. A fim de examinar a incidência de um imposto de consumo, começamos por caracterizar a mudança no equilíbrio que resulta da imposição do imposto. Por conveniência, pensamos no imposto como sendo coletado dos consumidores. Neste caso o preço ao consumidor é $q = p + t$ e o preço ao produtor é p. Assim, o equilíbrio é

$$D(p+\tau) = S(p) \tag{5.18}$$

Diferenciando ambos os lados em relação ao imposto τ:

$$\frac{dD(p)}{dp} \cdot \frac{d(p+\tau)}{d\tau} = \frac{dS(p)}{dp} \cdot \frac{dp}{d\tau} \tag{5.19}$$

Como $\dfrac{d(p+\tau)}{d\tau} = \dfrac{dp}{d\tau} + 1$, a equação acima torna-se:

$$\frac{dD(p)}{dp} \cdot \left(\frac{dp}{d\tau} + 1\right) = \frac{dS(p)}{dp} \cdot \frac{dp}{d\tau} \tag{5.20}$$

Agora, substituindo as elasticidades-preço, temos:

$$\eta_D \cdot \frac{D(p)}{p} \cdot \left(\frac{dp}{d\tau} + 1\right) = \eta_S \cdot \frac{S(p)}{p} \cdot \frac{dp}{d\tau} \tag{5.21}$$

Rearranjando para resolver $\frac{dp}{d\tau}$, obtemos:

$$\begin{aligned} \eta_D \cdot \left(\frac{dp}{d\tau} + 1\right) &= \eta_S \cdot \frac{dp}{d\tau} \\ \eta_D \cdot \frac{dp}{d\tau} + \eta_D &= \eta_S \cdot \frac{dp}{d\tau} \\ \frac{dp}{d\tau}(\eta_D - \eta_S) &= -\eta_D \end{aligned} \tag{5.22}$$

Assim, o efeito do imposto t sobre o preço p é dado por:

$$\frac{dp}{d\tau} = -\frac{\eta_D}{\eta_D - \eta_S} \tag{5.23}$$

Portanto, $0 \geq \frac{dp}{dt} \geq -1$.

De modo semelhante, se o produtor for responsável pelo pagamento do imposto, temos:

$$D(q) = S(q - t) \tag{5.24}$$

Seguindo o mesmo procedimento,

$$D'(q)dq = S'(q - t)(dq - dt) \tag{5.25}$$

Na proximidade de $t = 0$, temos $q = p$. Podemos reescrever a expressão acima como

$$\frac{dq}{dt} = \frac{\eta_S}{\eta_S - \eta_D} \tag{5.26}$$

Neste ponto, estamos prontos para avaliar a incidência de um imposto sobre consumo.

Portanto, $0 \leq \frac{dq}{dt} \leq 1$.

Mais interessante é notar que,

$$\frac{dp}{dt} + 1 = -\frac{\eta_D}{\eta_D + \eta_S} + 1 = \frac{\eta_S}{\eta_D + \eta_S} = \frac{dq}{dt} \tag{5.27}$$

o que mostra a irrelevância da incidência legal.

Considere as mudanças nos excedentes do consumidor e produtor decorrentes da introdução de um pequeno imposto.

Na ausência de impostos preexistentes, podemos aproximar o excedente do produtor conforme o número de unidades vendidas vezes a variação do preço do produtor ($Q\ dp$), e o excedente dos consumidores como o número de unidades compradas vezes a mudança no preço ao consumidor ($Q\ dq$). Por convenção, para tornar ambos os valores negativos, já que $dp < 0$ e $dq > 0$, definimos a carga do consumidor igual a $-Q\ dq$. Assim, para um imposto pequeno, a variação no excedente do consumidor é dada por:

$$\frac{dCS}{d\tau} = -\left(\frac{\eta_S}{\eta_S - \eta_D}\right) Q\tau \tag{5.28}$$

ou

$$\frac{dCS}{d\tau} = -D(p)\frac{\eta_S}{\eta_S - \eta_D} \tag{5.29}$$

Para um pequeno imposto τ, o preço ao consumidor aumenta em dq, reduzindo o excedente do consumidor. A variação aproximada é dada por:

$$\Delta CS \approx -Q \cdot dq \tag{5.30}$$

O imposto τ é dividido entre consumidores e produtores:

$$\tau = dq + dp \tag{5.31}$$

A razão entre as variações nos preços está relacionada às elasticidades:

$$\eta_D \cdot dq = -\eta_S \cdot dp \Rightarrow dp = -\frac{\eta_D}{\eta_S} \cdot dq \tag{5.32}$$

Substituindo em τ:

$$\begin{aligned} \tau &= dq + dp && (5.33) \\ &= dq - \frac{\eta_D}{\eta_S} \cdot dq && (5.34) \\ &= dq\left(1 - \frac{\eta_D}{\eta_S}\right) && (5.35) \end{aligned}$$

Logo, temos:

$$dq = \tau \cdot \frac{\eta_S}{\eta_S - \eta_D} \tag{5.36}$$

Substituindo o valor de dq na expressão para ΔCS:

$$\Delta CS \approx -Q \cdot \left(\tau \cdot \frac{\eta_S}{\eta_S - \eta_D}\right) \tag{5.37}$$

$$\Rightarrow \frac{dCS}{d\tau} = -\left(\frac{\eta_S}{\eta_S - \eta_D}\right) \cdot Q \cdot \tau \tag{5.38}$$

A variação no excedente do produtor é:

$$\frac{dPS}{d\tau} = \left(\frac{\eta_D}{\eta_S - \eta_D}\right) Q\tau \tag{5.39}$$

ou

$$\frac{dPS}{d\tau} = S(p)\frac{\eta_D}{\eta_S - \eta_D} \tag{5.40}$$

Para um pequeno imposto τ, o preço ao produtor cai em dp, reduzindo seu excedente. A variação aproximada é dada por:

$$\Delta PS \approx -Q \cdot dp \tag{5.41}$$

O imposto τ é dividido entre consumidores e produtores:

$$\tau = dq + dp \tag{5.42}$$

A razão entre as variações nos preços está relacionada às elasticidades:

$$\eta_D \cdot dq = -\eta_S \cdot dp \Rightarrow dq = -\frac{\eta_S}{\eta_D} \cdot dp \tag{5.43}$$

Substituindo em τ:

$$\begin{aligned} \tau &= dq + dp && (5.44) \\ &= -\frac{\eta_S}{\eta_D} \cdot dp + dp && (5.45) \\ &= dp\left(1 - \frac{\eta_S}{\eta_D}\right) && (5.46) \end{aligned}$$

Logo, temos:

$$dp = \tau \cdot \frac{\eta_D}{\eta_S - \eta_D} \tag{5.47}$$

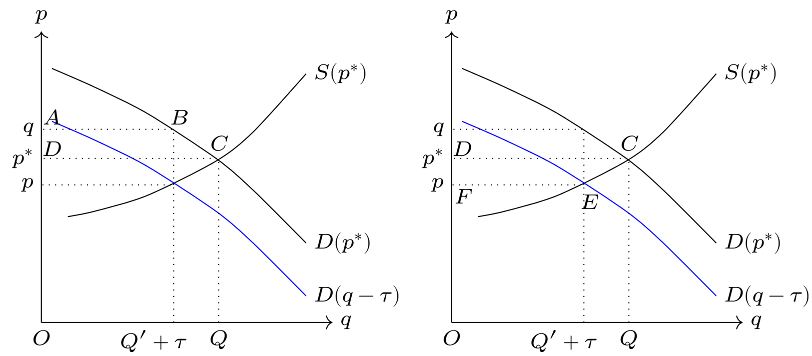
Substituindo o valor de dp na expressão para ΔPS:

$$\begin{aligned} \Delta PS &\approx -Q \cdot \left(\tau \cdot \frac{\eta_D}{\eta_S - \eta_D}\right) && (5.48) \\ \Rightarrow \frac{dPS}{d\tau} &= -(-1) \cdot Q \cdot \left(\frac{\eta_D}{\eta_S - \eta_D}\right) \cdot \tau && (5.49) \\ &= \left(\frac{\eta_D}{\eta_S - \eta_D}\right) \cdot Q \cdot \tau && (5.50) \end{aligned}$$

Observe que, para uma pequena alteração de imposto que começa com nenhum imposto, menos a soma das alterações no excedente do consumidor e do produtor é igual à receita do imposto, uma vez que $D(p) = S(p)$ e a mudança na receita tributária é igual a $d\tau S(p)$. Em termos gráficos, Figura 5.2, a mudança no excedente do consumidor é a área $ABCD$, e a mudança no excedente do produtor é a área

$DCEF$. Para impostos muito pequenos a soma destes impostos é essencialmente $ABEF$, a receita tributária.

Figura 5.2 – *Excedente do Consumidor e do Produtor*



As equações acima ilustram um princípio fundamental que se repetirá com frequência em nossa discussão sobre incidência tributária. Os impostos tendem a ser suportados por fornecedores ou demandantes inelásticos. É instrutivo considerar os casos limitantes.

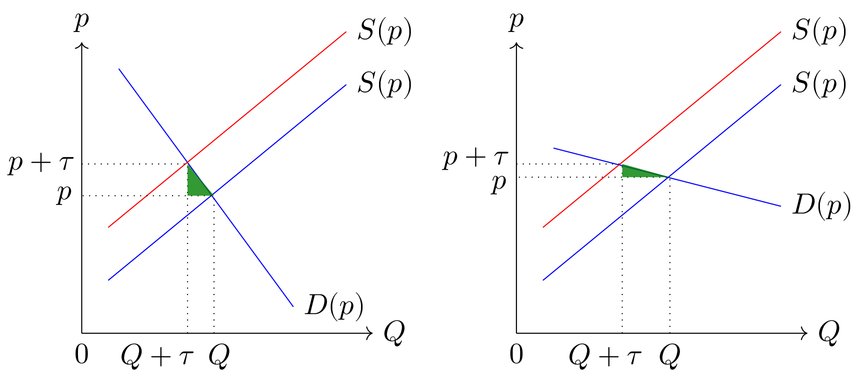
- Se a demanda é completamente inelástica $\eta_D = 0$ ou a oferta é perfeitamente elástica $\eta_S = +\infty$, os consumidores arcarão com todo o ônus de um imposto sobre consumo.

- Por outro lado, se a oferta é perfeitamente inelástica $\eta_S = 0$ ou a demanda é perfeitamente elástica $\eta_D = \infty$, todo o imposto será suportado pelos fornecedores.

Veja a Figura 5.3 para compreender como o repasse do imposto depende das elasticidades.

Mais geralmente, os impostos são suportados por aqueles que não podem se ajustar facilmente. Quanto maior a capacidade dos compradores de substituir a mercadoria tributada por outras commodities, maior será sua capacidade de transferir impostos. Da mesma forma, se os produtores não têm fatores fixos e podem deixar

Figura 5.3 – *Demanda Inelástica versus Demanda Elástica*



uma indústria onde os impostos estão sendo cobrados, sua curva de oferta é perfeitamente elástica e o imposto deve ser suportado pelos consumidores. Pois, se os vendedores fossem forçados a arcar com o imposto, ganhariam uma taxa de retorno abaixo do normal, levando-os a cessar a produção. Assim, no novo equilíbrio, os produtores recebem o mesmo preço para produzir como no antigo equilíbrio, enquanto o preço pago pelos consumidores sobe pelo valor total do imposto.

Embora essa análise descreva com precisão os efeitos da introdução de um imposto sobre consumo em um mercado pequeno, onde não há distorções pré-existentes, é difícil se estender a outros casos. Em geral, mudanças na curva de demanda, tais como seriam causadas por uma mudança de impostos, serão associadas a mudanças na demanda por outros produtos. Isso alterará seus preços, levando a mudanças nos preços dos fatores, o que afetará a posição das curvas de oferta e demanda. Ao considerar os impostos que afetam uma grande parte da economia, é necessário, portanto, adotar uma perspectiva de equilíbrio geral, e não a visão de equilíbrio parcial tomada acima. Dois princípios que emergem desta análise de equilíbrio parcial permanecerão válidos. Primeiro, a incidência de impostos não depende de qual lado do mercado o imposto é avaliado. Segundo, os impostos serão deslocados por esses agentes e fatores que são mais elásticos na oferta ou demanda.

O governo aumenta impostos para aumentar receita para financiar bens ou para redistribuir renda de ricos para pobres. Mas aumentar a receita fiscal geralmente tem um custo de eficiência: para gerar R$ 1 de receita, precisa reduzir o bem-estar dos indivíduos em mais de R$ 1. Os custos de eficiência vêm da distorção do comportamento.

A carga de peso morto (também chamada de sobrecarga) da tributação é definida como a perda de bem-estar criada por um imposto para além da receita fiscal gerada pelo imposto. No diagrama simples de oferta e demanda, o bem-estar é medido pela soma do excedente do consumidor e do excedente do produtor.

A perda de bem-estar da tributação é medida como uma mudança no excedente do consumidor + do produtor menos o imposto cobrado. A ineficiência de qualquer imposto é determinada pela medida em que consumidores e produtores mudam seu comportamento para evitar o imposto; a perda de peso morto é causada por indivíduos e empresas que fazem escolhas ineficientes de consumo e produção, a fim de evitar a tributação. Se não houver alteração nas quantidades consumidas, o imposto não tem custos de eficiência.

5.4 Receita Tributária e Impostos sobre Consumo: Elasticidade Crítica

Na Tabela 5.1, são apresentados o preço de equilíbrio (P_e) e a quantidade de equilíbrio (Q_e), juntamente com o imposto sobre consumo arrecadado (T) e o gasto total dos compradores (TE) para diferentes níveis de imposto sobre consumo (t). A tabela baseia-se nas funções lineares gerais de demanda e de oferta

$$P_d = a + bQ \tag{5.51}$$

e

$$P_s = c + dQ + t \tag{5.52}$$

em que P_d e P_s são, respectivamente, o preço da demanda e o preço da oferta. As suposições são que $b < 0$, $a > 0$, $c > 0$, $d > 0$, e $a > c$. As equações de equilíbrio para a quantidade trocada e o preço, em termos dos parâmetros, são:

$$Q_e = \frac{c-a}{b-d} + \frac{t}{b-d} \tag{5.53}$$

e

$$P_e = a + b\left[\frac{c-a}{b-d} + \frac{t}{b-d}\right] \tag{5.54}$$

Tabela 5.1 – *Alíquotas do Imposto sobre Consumo, Preço e Quantidade de Equilíbrio, Receita Tributária e Gasto Total*

t	P_e	Q_e	T	TE
0	250	150	0	37.500
30	265	135	4.050	35.775
60	280	120	7.200	33.600
90	295	105	9.450	30.975
120	310	90	10.800	27.900
150	325	75	11.250	24.375
180	340	60	10.800	20.400
210	355	45	9.450	15.975
240	370	30	7.200	11.100
270	385	15	4.050	5.775
300	400	0	0	0

A receita do imposto sobre consumo e o gasto total são dados por:

$$T = tQ_e \tag{5.55}$$

e

$$TE = P_e Q_e \tag{5.56}$$

Na Tabela 5.1, a forma específica das equações de demanda e oferta é:

$$P_d = 400 - Q \tag{5.57}$$

e

$$P_s = 100 + Q + t \tag{5.58}$$

Em geral, existem duas alíquotas de imposto que resultarão em receitas tributárias idênticas. Por exemplo, uma receita tributária de \$10.800 pode ser obtida se o imposto específico for de \$120 ou \$180 por unidade. À medida que a alíquota do imposto sobre consumo aumenta no intervalo $0 < t < 150$ por unidade, a receita tributária aumenta. Mas, à medida que a alíquota do imposto aumenta no intervalo $150 < t < 300$ por unidade, a receita tributária diminui.

Por que isso acontece? A razão pela qual o aumento na alíquota do imposto sobre consumo resulta em uma redução da receita tributária é, essencialmente, que a elasticidade-preço da demanda não permanece constante. Eu postulei uma curva de demanda linear. À medida que o preço aumenta, a quantidade demandada diminui

– um movimento para cima e para a esquerda ao longo da curva de demanda. A elasticidade da demanda, portanto, aumenta em valor absoluto – ela se torna mais negativa – à medida que a quantidade demandada se desloca para cima na curva de demanda. Assim, a queda percentual na quantidade demandada logo se torna igual ao aumento percentual na alíquota do imposto. Como

$$T = tQ_e \tag{5.59}$$

disso decorre que:

$$\%\Delta T = \%\Delta t + \%\Delta Q_e \tag{5.60}$$

sendo que o sinal do segundo termo é sempre oposto ao do primeiro termo. À medida que Q_e se aproxima de zero, a variação percentual absoluta de Q_e se torna cada vez maior em comparação à variação percentual de t, mesmo que a taxa de variação de Q_e para uma dada variação em t permaneça constante.

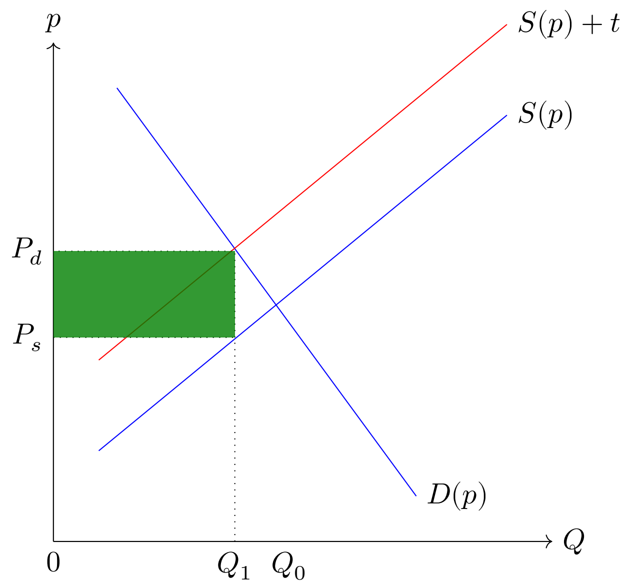
A alíquota de imposto t^* que resulta na receita tributária máxima (T^*) é aquela em que a elasticidade da demanda é tal que a variação percentual de t é igual (e oposta) à variação percentual de Q_e. Com essa alíquota, aumentos adicionais no imposto resultam em queda da receita tributária. Referirei-me a essa alíquota t^* como a alíquota crítica.

Na Figura 5.4, a curva de oferta sem imposto é $S(p)$, com intercepto vertical no ponto c e inclinação ${}^{dP}\!/_{dQ} = d$, ambos positivos. A imposição de um imposto específico t por unidade resulta em um deslocamento vertical paralelo para cima da curva de oferta – a inclinação permanece igual a d, enquanto o intercepto vertical aumenta para $c + t$.

A curva de demanda é $D(p)$, com intercepto a e inclinação ${}^{dP}\!/_{dQ} = b$, onde

$b < 0$. O imposto arrecadado será positivo, a menos que o imposto unitário exceda o valor $a - c$, caso em que a quantidade vendida cai para zero.

Figura 5.4 – *Imposição, Incidência e Arrecadação de um Imposto Específico (Imposto sobre Consumo)*



Ao aumentar o imposto unitário de t_1 para t_2, a curva de oferta se desloca para cima, de $S + t_1$ para $S + t_2$. A quantidade vendida diminui de Q_1 para Q_2.

A priori, não se pode prever o sinal da variação líquida na receita tributária. No entanto, ao impor limites para a elasticidade da demanda, é possível derivar condições necessárias e suficientes para concluir que a receita tributária pode não aumentar com o aumento do imposto unitário.

Para desenvolver as condições relativas ao imposto específico que maximiza a receita tributária, toma-se o diferencial total da equação (5.55), obtendo-se:

$$dT = tdQ_e + Q_e dt \tag{5.61}$$

ou, na forma de derivada:

$$\frac{dT}{dt} = t\frac{dQ_e}{dt} + Q_e \tag{5.62}$$

Pela equação (5.53), a derivada de Q_e em relação a uma mudança no imposto específico é:

$$\frac{dQ_e}{dt} = \frac{1}{b-d} \tag{5.63}$$

Essa derivada é negativa, uma vez que b é negativo e d é positivo. Combinando as equações (5.53), (5.55) e (5.63), obtemos:

$$\frac{dT}{dt} = \frac{c-a}{b-d} + \frac{2t}{b-d} \tag{5.64}$$

Como $a > c > 0$, o primeiro termo na equação (5.64) é positivo, enquanto o segundo termo, $\frac{2t}{b-d}$, é negativo. Um valor de t tal que o valor absoluto do segundo termo exceda o do primeiro fornecerá a condição necessária:

$$\frac{dT}{dt} < 0 \quad \Longrightarrow \quad 2t > |c-a| \tag{5.65}$$

ou, como $a > c$, o valor de t avaliado quando $\frac{dT}{dt} = 0$ (isto é, quando $t = t^*$) resulta em:

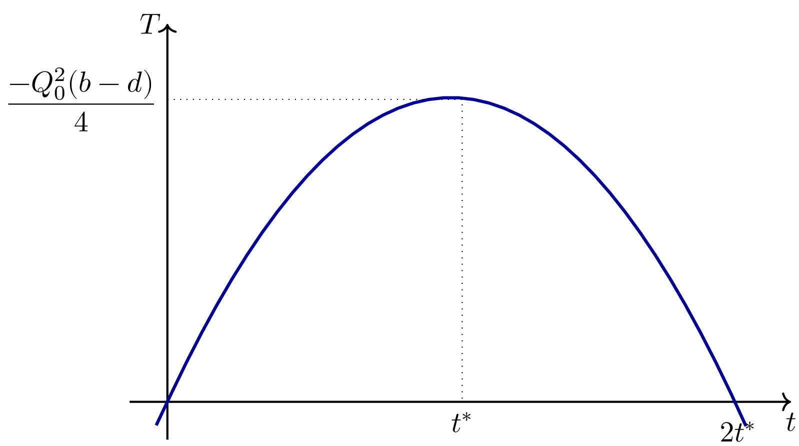
$$t^* = \frac{a-c}{2} \tag{5.66}$$

A quantidade trocada diminui à medida que t aumenta o preço de equilíbrio, até que o preço de equilíbrio finalmente faça a quantidade demandada cair para zero. Note que quando $t = t^*$, a quantidade comprada é igual à metade da quantidade comprada inicialmente, com imposto específico igual a zero.

A receita tributária máxima ocorre quando t= t^*, a taxa crítica. Nesse valor de t, o aumento em T devido ao aumento do imposto específico é justamente igual à diminuição em T devido à queda nas vendas. A inclinação instantânea de T, $\dfrac{dT}{dt}$, é igual a zero quando $t = t^*$.

A Figura (5.5) apresenta a relação entre a quantidade de receita tributária, T, e o imposto específico t. Essa relação é não linear em t, pois Q_e é determinado por t. A receita tributária máxima ocorre quando $t = t^*$, a alíquota crítica. Nesse ponto, o aumento em T devido ao acréscimo do imposto específico é exatamente igual à redução em T causada pela queda nas vendas. A inclinação instantânea de T, dT/dt, é igual a zero quando $t = t^*$.

Figura 5.5 – *Receita Tributária (T) com Preço de Equilíbrio de Mercado e Vendas em Diversos Níveis de Imposto Específico (t)*



Se $t = 0$, pela equação (5.53), $Q_e = \dfrac{c-a}{b-d}$. Essa quantidade é Q_0. Avaliando a equação (5.53) com $t = t^*$, onde $t^* = \dfrac{a-c}{2}$, temos:

$$Q_e = \frac{c-a}{2(b-d)} \tag{5.67}$$

ou seja, $\dfrac{Q_0}{2}$. Essa quantidade é Q^*.

Avaliando a equação (5.55) com $t = t^*$ e com $Q_e = Q^*$, o valor máximo da receita do imposto específico é $T^* = t^*Q^*$ ou:

$$T^* = -\frac{(Q_0)^2}{4(b-d)} \tag{5.68}$$

em que Q_0 é a quantidade inicial trocada no equilíbrio sem imposto e é determinada apenas pelos parâmetros a, b, c, d – os parâmetros de oferta e demanda do modelo.

Para derivar a equação (5.68), primeiro escrevemos a equação da receita tributária:

$$T = tQ_e \tag{5.69}$$

Como

$$Q_e = \frac{c-a}{b-d} + \frac{t}{b-d} \tag{5.70}$$

e

$$Q_0 = \frac{c-a}{b-d} \tag{5.71}$$

então

$$T = tQ_e = t\left[Q_0 + \frac{t}{b-d}\right] \tag{5.72}$$

Agora, tomamos a derivada total:

$$\begin{aligned}
\frac{dT}{dt} &= t\left(\frac{dQ_e}{dt}\right) + Q_e \\
&= t\left(\frac{1}{b-d}\right) + \left[Q_0 + \frac{t}{b-d}\right] \\
&= \frac{t}{b-d} + Q_0 + \frac{t}{b-d} \\
&= \frac{2t}{b-d} + Q_0
\end{aligned} \tag{5.73}$$

em que $\frac{dT}{dt} = 0$, a receita tributária é maximizada, pois a segunda derivada total é negativa. Seja t^* o valor de t que maximiza T. Então,

$$\frac{dT}{dt} = 0 \quad \Longleftrightarrow \quad \frac{2t^*}{b-d} + Q_0 = 0 \tag{5.74}$$

Logo,

$$\begin{aligned}
2t^*(b-d) &= -Q_0 \\
t^* &= -\frac{Q_0(b-d)}{2}
\end{aligned} \tag{5.75}$$

No ponto T^*, temos $Q^* = \frac{Q_0}{2}$. Assim, por substituição:

$$T^* = t^*Q^* = \left(-\frac{Q_0(b-d)}{2}\right)\left(\frac{Q_0}{2}\right) = -\frac{Q_0^2(b-d)}{4} \tag{5.76}$$

Como a condição crítica depende do aumento (absoluto) na elasticidade da demanda à medida que a taxa do imposto específico se move para a esquerda e para cima na curva de demanda, discutirei a elasticidade da demanda na quantidade, preço e taxas do imposto específico que resultam em T^*. A seguinte notação será útil:

E^* é a elasticidade crítica, $\left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right)$ avaliada em $t = t^*$.

E_0 é a elasticidade inicial, $\left(\frac{dQ}{dP}\right)\left(\frac{P}{Q}\right)$ avaliada em $t = 0$.

Q^* e P^* são a quantidade e o preço de equilíbrio crítico, respectivamente, avaliados em $t = t^*$.

Q_0 e P_0 são a quantidade e o preço de equilíbrio iniciais, respectivamente, avaliados em $t = 0$.

Para derivar E^*, primeiro reescreva a equação da posição da demanda (5.51) com a quantidade demandada como uma função explícita do preço, e avaliada em $Q = Q^*$:

$$Q^* = -\frac{a}{b} + \frac{P}{b} \tag{5.77}$$

Como $\frac{dQ}{dP} = \frac{1}{b}$, $Q^* = \frac{Q_0}{2}$, e $P^* = P_0 + \frac{a - P_0}{2}$, a elasticidade da demanda em $Q = Q^*$ pode ser simplificada escrevendo-se:

$$E^* = \frac{2}{Q_0}\left(\frac{a}{b}\right) + 1 \tag{5.78}$$

Para derivar a equação (5.78), primeiro escrevemos a elasticidade-preço da demanda:

$$E^* = \left(\frac{dQ}{dP}\right)\left(\frac{P^*}{Q^*}\right) \tag{5.79}$$

Como $\dfrac{dQ}{dP} = \dfrac{1}{b}$, $Q^* = \dfrac{Q_0}{2}$, e $P^* = P_0 + \dfrac{a - P_0}{2}$, por substituição temos:

$$\begin{aligned}
E^* &= \left(\frac{1}{b}\right)\left(\frac{P_0 + \dfrac{a - P_0}{2}}{Q_0/2}\right) && (5.80)\\
&= \frac{2}{bQ_0}\left(P_0 + \frac{a}{2} - \frac{P_0}{2}\right) && (5.81)\\
&= \frac{2}{bQ_0}\left(\frac{P_0}{2} + \frac{a}{2}\right) && (5.82)\\
&= \frac{1}{bQ_0}(P_0 + a) && (5.83)
\end{aligned}$$

Agora, lembrando que quando $Q = Q_0$, então $P_0 = a + bQ_0$. Substituindo isso na equação acima, temos:

$$\begin{aligned}
E^* &= \frac{1}{bQ_0}(a + bQ_0 + a) && (5.84)\\
&= \frac{1}{bQ_0}(2a + bQ_0) && (5.85)\\
&= \frac{2a}{bQ_0} + 1 && (5.86)\\
E^* &= \left(\frac{2}{Q_0}\right)\left(\frac{a}{b}\right) + 1 && (5.87)
\end{aligned}$$

que é a equação (5.78).

Avaliada em $Q = Q_0$, e simplificada, a elasticidade da demanda pode ser escrita como:

$$E_0 = \frac{1}{Q_0}\left(\frac{a}{b}\right) + 1 \tag{5.88}$$

Para derivar a equação (5.88), siga o mesmo procedimento, mas avalie a elasticidade em $Q = Q_0$ e $P = P_0$:

$$E_0 = \left(\frac{dQ}{dP}\right)\left(\frac{P_0}{Q_0}\right) = \left(\frac{1}{b}\right)\left(\frac{P_0}{Q_0}\right) \tag{5.89}$$

Como $P_0 = a + bQ_0$, temos:

$$E_0 = \left(\frac{1}{b}\right)\left(\frac{a + bQ_0}{Q_0}\right) \tag{5.90}$$

$$= \frac{a}{bQ_0} + \frac{bQ_0}{bQ_0} \tag{5.91}$$

$$= \frac{a}{bQ_0} + 1 \tag{5.92}$$

$$E_0 = \left(\frac{1}{Q_0}\right)\left(\frac{a}{b}\right) + 1 \tag{5.93}$$

que é a equação (5.88).

Finalmente, para ver a relação entre as duas elasticidades que deve ser mantida, defina:

$$X = \frac{1}{Q_0}\left(\frac{a}{b}\right) \tag{5.94}$$

de modo que as duas elasticidades podem ser expressas como:

$$E_0 = X + 1 \tag{5.95}$$

e

$$E^* = 2X + 1 \tag{5.96}$$

Pela equação (5.95), $X = E_0 - 1$, então a elasticidade crítica pode ser escrita como:

$$E^* = 2E_0 - 1 \tag{5.97}$$

Nessa forma, a elasticidade crítica é facilmente determinada. Ela é duas vezes a elasticidade sem imposto, menos um. Por exemplo, suponha que, inicialmente, sem imposto sobre o consumo, a elasticidade da demanda (E_0) seja $-0,5$. A elasticidade crítica é, portanto, -2, pois $2 \times (-0,5) - 1 = -2$. Isso indica que aumentos no imposto específico resultarão em aumentos cada vez menores na receita tributária, até que a elasticidade da demanda se torne igual ou mais elástica que -2. A partir desse ponto, aumentos no imposto específico resultarão em quedas cada vez maiores na receita tributária. Com a especificação em valor absoluto das elasticidades, a elasticidade que maximiza a receita tributária é duas vezes o valor original (sem imposto), mais um. Por exemplo, se $E_0 = 1$, então $E^* = 3$.

5.5 Peso Morto e Bem-Estar

A única justificativa econômica para as perdas sociais incorridas pela sociedade em razão do imposto se dá quando o governo possui um outro projeto para o qual ele corretamente faz a avaliação social, ou seja, calcula o valor presente líquido, à taxa social de desconto (se, para isso, desloca fundos públicos), levando em conta preços-sombras sociais dos recursos deslocados da economia. É para isso que existe a avaliação social de projetos, que difere da avaliação privada. Note que o projeto deve ser empreendido pelo governo, pois não interessa ao setor privado. O governo cria o imposto, gera perda social, mas, com o que foi arrecadado, financia, por exemplo, uma represa ou universidades federais.

Arnold Harberger[36] tem vários estudos em que mostra como avaliar socialmente construções de represas, de auto-estradas e várias outras coisas. Falta-nos a cultura da avaliação social de projetos. Se um banco de desenvolvimento financia projetos com fundos públicos, ele deve, então, descontar os valores futuros pela taxa social de desconto, que é uma média ponderada entre a produtividade marginal do capital dos investidores e a taxa de desconto intertemporal dos poupadores, ponderação essa que depende intimamente das elasticidades-juro desses agentes.

O imposto da janela fornece um exemplo histórico dramático e claro dos possíveis efeitos distorcidos da tributação[37]. Imposto na Inglaterra em 1696, o imposto – uma espécie de antecessor do imposto predial moderno – era cobrado sobre residências com o passivo tributário baseado no número de janelas. O imposto levou a esforços para reduzir as notas fiscais através de medidas como o fechamento de janelas e a construção de casas com muito poucas janelas. Às vezes, pisos inteiros de casas não tinham janelas. Apesar dos efeitos perniciosos à saúde e da estética, e apesar dos protestos generalizados, o imposto persistiu por mais de um século e meio: foi finalmente revogado em 1851. O efeito adverso mais sério da taxa de janela foi sobre a saúde humana. Com base em uma série de estudos de médicos se descobriu que as condições insalubres resultantes da falta de ventilação adequada e ar fresco incenti-

[36] Ver Gary Becker, Armen Alchian, Harold Demsetz, Axel Leijonhuvfud, Joseph Ostroy, William Zame, Gordon Tullock e Ken Binmore como referências extras.

[37] Oates, W. E. & Schwab, R. M. (2015). "The Window Tax: A Case Study in Excess Burden". *Journal of Economic Perspectives*, 29(1): 163–180.

vavam a propagação de numerosas doenças como disenteria, gangrena e tifo. Em um exemplo, em 1781, uma epidemia de tifo matou muitos cidadãos em Carlisle. O Dr. John Heysham traçou as origens do surto em uma casa habitada por seis famílias pobres e descreveu a habitação desta maneira:

> *In order to reduce the window tax, every window that even poverty could dispense with was built up, and all source of ventilation were thus removed. The smell in this house was overpowering, and offensive to an unbearable extent. There is no evidence that the fever was imported into this house, but it was propagated from it to other parts of town, and 52 of the inhabitants were killed.*

O imposto não consistia em uma série de taxas marginais crescentes, mas incluía uma série de "entalhes" - pontos nos quais uma janela adicional trazia um grande aumento no passivo tributário. Considere, por exemplo, as reformas introduzidas em 1747, sob as quais o Parlamento aumentou e reformulou a estrutura de taxas do imposto. Os 2 *shillings* fixos por moradia foram destacados do imposto sobre janelas e impostos, além de um novo cronograma de taxas de janelas. Sob o novo cronograma de tarifas, havia um imposto de 6 *pences* em cada janela de uma casa com 10 a 14 janelas, de 9 *pences* por janela em casas com 15 a 19 janelas e de 1 *shilling* por cada janela em casas com mais de 20 janelas. Como resultado, podemos esperar encontrar, por exemplo, muito mais casas com 9, em vez de 10, janelas.

O imposto da janela, aliás, tinha um antecedente: o imposto da lareira. Imposto em 1662 por Carlos II após a Restauração, o imposto da lareira consistia em uma taxa de 2 *shillings* por cada lareira e fogão em casas na Inglaterra e no País de Gales. O imposto foi muito impopular em parte por causa do caráter intrusivo do processo de avaliação. Os "homens da chaminé" (como eram chamados os assessores e cobradores de impostos) tinham que entrar na casa para contar o número de lareiras e fogões, e havia um grande ressentimento contra essa invasão da santidade da casa. A taxa de janelas, ao contrário, não exigia acesso ao interior da habitação: os "espiadores de janelas" podiam contar janelas de fora, simplificando o procedimento de avaliação e evitando a necessidade de invasão do interior.

Ambos os impostos destinavam-se a ser um indicador visível da capacidade de pagamento. Conforme apontado em uma discussão na Câmara dos Comuns (1850), pouco antes da revogação do imposto sobre janelas,

> *o imposto sobre janelas, quando aplicado pela primeira vez, não era destinado a impostos sobre janelas, mas como um imposto predial, a casa era considerada um critério seguro do valor da propriedade de um homem, e as janelas eram assumidas apenas como o índice do valor das casas.*

Mas como Adam Smith (1776 [1937], p. 798) observou em A Riqueza das Nações, o número de janelas poderia ser uma medida muito pobre do valor de uma habitação:

> *casa de dez libras de aluguel no país pode ter mais janelas do que uma casa de quinhentas libras de aluguel em Londres; e, embora seja provável que o habitante do primeiro seja um homem muito mais pobre do que o último, ainda que sua contribuição seja regulada pelo imposto da janela, ele deve contribuir mais para o apoio do Estado.*

Em 1846, oficiais médicos pediram ao Parlamento a abolição do imposto da janela, declarando-o "mais prejudicial à saúde, bem-estar, propriedade e indústria dos pobres e da comunidade em geral".

Um sistema tributário cria uma descontinuidade se uma pequena mudança de comportamento levar a uma mudança discreta nas taxas de imposto média e marginal. Uma pessoa que possuía uma casa com 9 ou menos janelas não pagava impostos. Mas seu vizinho, cuja casa tinha 10 janelas, pagaria um imposto de 6 *pence* por cada janela: a taxa marginal de imposto para a 10ª janela era de 60 *pence*.

As descontinuidades levam a grandes perdas de peso morto: ou seja, um esquema tributário com descontinuidades fornece fortes incentivos para que os contribuintes distorçam o comportamento e se localizem em torno de uma descontinuidade.

A perda de peso morto correspondia a 62% dos impostos pagos por esses consumidores. Ou seja, para cada dólar cobrado, a versão simulada do imposto da janela impunha um encargo adicional de 62 centavos às famílias. O excesso de carga é particularmente grande para as famílias que escolheram 9 janelas. Esses consumidores pagaram zero no imposto da janela e, portanto, para eles, todo o ônus do imposto é um ônus excessivo. O imposto da janela é, portanto, um exemplo bastante impressionante de um imposto que levou a um comportamento radical de evitar impostos com altos níveis associados de excesso de carga. Se o imposto da janela era um imposto ruim que gerava efeitos adversos e críticas intensas, por que persistiu por um período tão longo? O uso continuado do imposto da janela foi, em parte, pelo menos, uma resposta a um cenário de extrema pressão orçamentária, na qual o governo percebeu pouco espaço para redução de quaisquer taxas tributárias.

Temos em mente um objetivo pedagógico. O conceito de excesso de carga (ou "perda de peso morto") é para economistas aspecto nevrálgico da análise tributária. Mas para os leigos a noção é realmente bastante misteriosa; os economistas das finanças públicas geralmente têm alguma dificuldade, por exemplo, em explicar aos contribuintes os custos com o bem-estar das distorções induzidas pelos impostos na alocação de recursos. O imposto da janela é um exemplo de como um imposto pode ter sérios efeitos colaterais adversos no bem-estar social. Além de suas consequências questionáveis para a equidade tributária, o imposto da janela resultou em óbvias e custosas má alocações de recursos.

Deadweight loss ou custos de eficiência é o dinheiro que o governo queima quando cria impostos, tarifas, subsídios, quotas, tetos de preço, pisos de preço e outras coisas mais que até seriam toleráveis, não fosse pelo mau uso que o governo faz e pela incompetência do governo em não calculá-los. Vamos usar aqui uma abordagem com quantidades discretas e vamos ver consumidores e produtores que o governo expulsa do mercado quando faz incidir sobre este um imposto (*excise tax*).

A troca econômica traz ganhos para todas as partes envolvidas na transação. Para entendermos isso, devemos, primeiro, entender os conceitos de excedente do produtor (o ofertante) e excedente do consumidor (o demandante).

Considere o mercado de um bem qualquer, digamos, frango. É importante especificar o período de tempo em que a transação é relevante, por exemplo, por semana. Assim, quando se diz que um consumidor consome ou demanda um frango, deve-se subentender que ele demanda um frango por semana. O mesmo vale para a oferta e a mesma unidade de tempo deve subjazer a todas as quantidades mencionadas, a todas as utilidades subsumidas e funções de custo consideradas. Sem essa especificação, os conceitos de oferta e demanda perdem totalmente o sentido.

Para simplificar, em vez de considerar uma curva de oferta agregada e uma de demanda agregada genéricas, impessoais, suponha que cada demandante demande um e apenas um frango e que, similarmente, cada ofertante oferte um e apenas um frango. Cada um deles terá um nome, para que você possa entender o que significa, na realidade, a intervenção do governo no mercado. Os demandantes são ordenados de forma descendente a partir da valoração mais alta, ou seja, desde aquele que está disposto a pagar mais até aquele disposto a pagar menos, como na Tabela 5.2.

Os ofertantes também são pessoas, só que eles serão ordenados de forma ascendente, desde o que está disposto a ofertar por menos até aquele disposto a ofertar por mais, como na Tabela 5.3.

Juntando tudo numa tabela só, temos o seguinte quadro (Tabela 5.4) de preços de demanda e de oferta para as diversas quantidades de frango transacionadas:

Assim, o número 3 relativo a Maria significa que Maria possui a terceira maior disposição a pagar pelo frango (R$ 7), não que ela demande três unidades.

A lista de preços de oferta e de demanda pode ser mais bem visualizada mediante sua representação gráfica. No eixo horizontal coloco as quantidades e no eixo

Tabela 5.2 – *Preços de Demanda*

Demandante	Preço de demanda
1. João	9
2. Pedro	8
3. Maria	7
4. Carla	$6,50$
5. Antônio	$5,50$
6. Dora	3
7. Gustavo	1
8. José	0
9. Patrícia	0
10. Amélia	0

Tabela 5.3 – *Preços de Oferta*

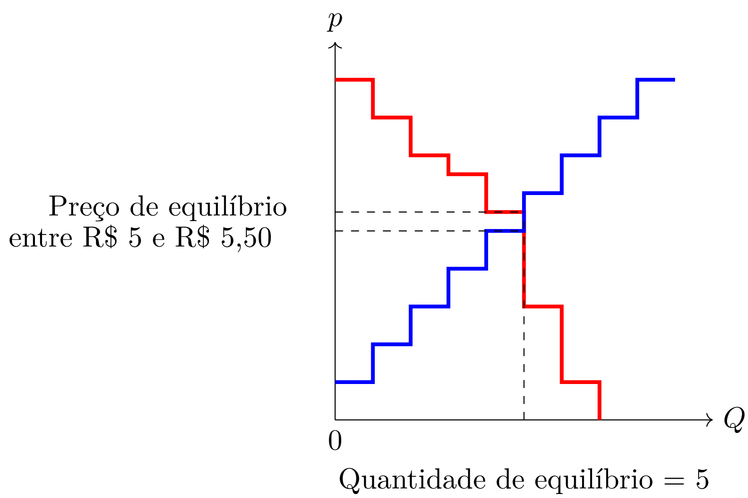
Ofertante	Preço de oferta
1. Catarina	1
2. Augusto	2
3. Roberto	3
4. Aline	4
5. Paula	5
6. Marcelo	6
7. Tiago	7
8. Luíza	8
9. Agnes	9
10. César	10

Tabela 5.4 – *Preços de Demanda e de Oferta por Unidade*

Quantidade	Preço de demanda	Preço de oferta
1	9	1
2	8	2
3	7	3
4	$6,50$	4
5	$5,50$	5
6	3	6
7	1	7
8	0	8
9	0	9
10	0	10

vertical o valor monetário correspondente ao preço, seja de oferta ou de demanda.

Figura 5.6 – *Curvas de Oferta e de Demanda do Mercado de Frangos*



Pela tabela acima - e com a ajuda do gráfico -, podemos ver que ao preço de R$ 5 por frango, a quantidade ofertada por semana é igual à quantidade demandada. Um equilíbrio de mercado é o comércio de 5 frangos no período ao preço de R$ 5 a unidade.

Para mostrarmos que a quantidade de equilíbrio é de 5 unidades, vejamos o que ocorre quando a quantidade comercializada é de 4 unidades ou de 6 unidades. Se a quantidade comercializada fosse de 4 unidades, tanto o quinto ofertante (Paula) quanto o quinto demandante (Antônio) veriam a possibilidade de ganhos no comércio da quinta unidade. Com efeito, o demandante Antônio pagaria R$ 5,50 pelo frango e a ofertante Paula aceitaria R$ 5 por ele, de modo que existe espaço para a troca. Se o sexto ofertante, Marcelo, decidisse ofertar a sexta unidade, ele teria um custo marginal de R$ 6, mas a sexta demandante, Dora, só estaria disposta a pagar R$ 3 pela sexta unidade. Assim, Marcelo não conseguiria cobrir o seu custo marginal.

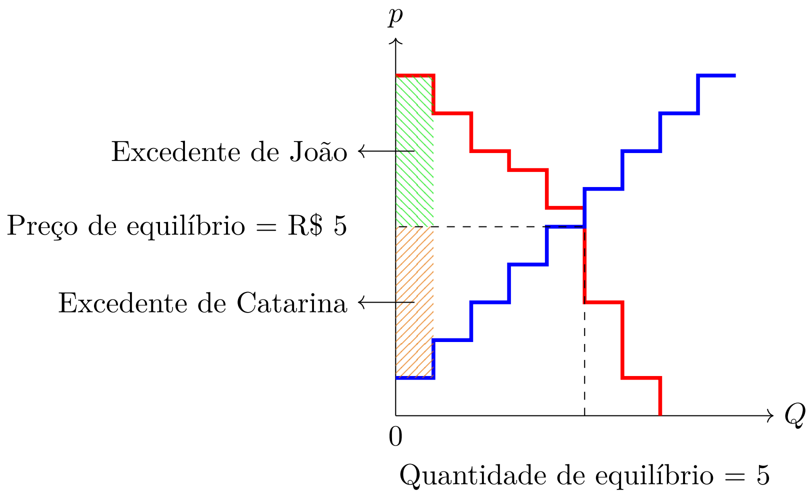
Disse acima que um equilíbrio de mercado é o comércio de 5 frangos no período ao preço de R$ 5 a unidade. Ora, outro equilíbrio seria o comércio de 5 frangos ao preço de R$ 5,50 a unidade. Procedendo do mesmo modo como fiz acima, é fácil ver que esse também é um equilíbrio. Mas qual é a diferença entre os dois? É o preço de equilíbrio! Paula exige no mínimo R$ 5 para ofertar a quinta unidade e Antônio está disposto a pagar no máximo R$ 5,50 por ela. Portanto, qualquer preço entre R$ 5 e R$ 5,50 pode equilibrar a oferta com a demanda. Qual preço prevalecerá? É nesse ponto que a Microeconomia abre espaço para a barganha e para isso há uma literatura abundante e cursos específicos. Independentemente do resultado da barganha, o importante é que o preço final acordado não afetará os ganhos totais de troca, afetará, porém, toda a distribuição dos ganhos de troca entre ofertantes e demandantes, conforme mostrarei logo adiante.

A leitura econômica da tabela 5.4 é, no entanto, mais sutil e revela muito do fenômeno da troca. Suponha, então, que o preço de equilíbrio é R$ 5. João está disposto a pagar R$ 9 pelo primeiro frango e Catarina está disposta a ofertá-lo por R$ 1. Existe vantagem nessa troca, ambos saem ganhando. João pagaria R$ 9 pelo frango, mas pagou apenas R$ 5. Teve um excedente de R$ 4 = R$ 9 - R$ 5. Catarina ofertaria o frango por R$ 1, mas recebeu R$ 5 por ele. Teve um excedente de R$ 4 = R$ 5 - R$ 1. O excedente dessa transação é a soma dos excedentes das partes da transação, ou seja, R$ 8 = R$ 4 + R$ 4. Outra forma de se obter o excedente total dessa primeira transação é pela diferença entre a disposição máxima a pagar e a disposição mínima para ofertar, a saber, R$ 8 = R$ 9 - R$ 1. Isso é ilustrado na Figura 5.7.

Já na transação do segundo frango, Pedro estaria disposto a pagar R$ 8, mas pagou R$ 5, de modo que seu excedente de consumidor nessa segunda transação foi de R$ 3 = R$ 8 - R$ 5. Augusto ofertaria o frango por R$ 2, mas recebeu R$ 5, de modo que seu excedente de produtor (ou ofertante) nessa segunda transação foi de R$ 3 = R$ 5 - R$ 2. O excedente da segunda transação é, portanto, R$ 6.

Continuando dessa maneira, podemos construir a Tabela 5.5 com o excedente do demandante, o excedente do ofertante e o excedente total para cada quantidade

Figura 5.7 – *Distribuição do Excedente de Troca da Primeira Transação*



comercializada ao preço de equilíbrio de R$ 5.

A última coluna é fácil de entender. Tome, por exemplo, a linha referente à quantidade de 3 unidades. O excedente total do comércio de 3 frangos é o excedente total do comércio do primeiro, do segundo e do terceiro frango, ou seja, R$ 8 + R$ 6 + R$ 4 = R$ 18. Note agora que, quando a quantidade comercializada é a de equilíbrio, 5 frangos, o ganho (ou excedente) total de trocas é maximizado, a saber, R$ 21. Com 5 unidades, o excedente total dos demandantes é R$ 11 e o excedente total dos ofertantes é R$ 10. O ganho total de trocas, R$ 21, foi repartido entre os demandantes, que ficaram com R$ 11, e os ofertantes, que ficaram com R$ 10. Assim, temos que:

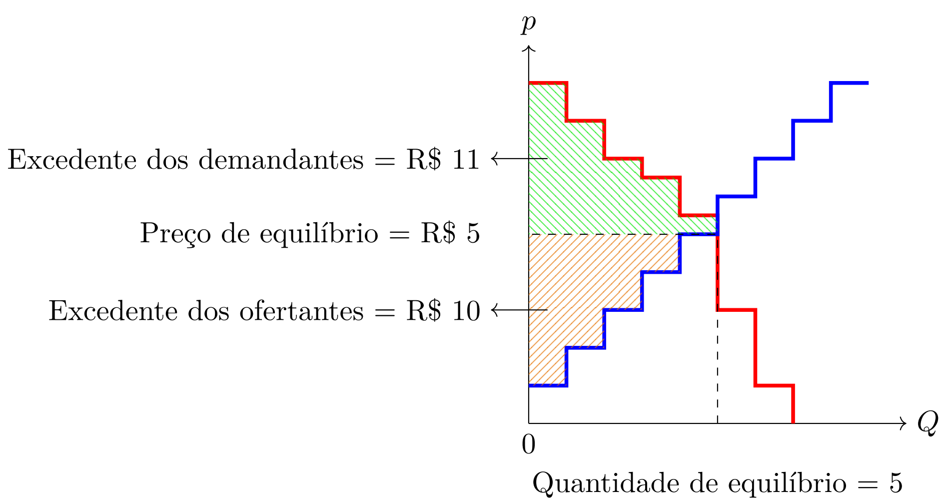
$$\text{ganho total de trocas} = \text{excedente dos demandantes} + \text{excedente dos ofertantes}$$
$$\text{R\$ } 21 = \text{R\$ } 11 + \text{R\$ } 10$$

O Gráfico 5.8 ilustra a distribuição dos ganhos totais de troca entre excedente dos demandantes e excedente dos ofertantes para o caso em que o preço de equilíbrio é de R$ 5 por unidade:

Tabela 5.5 – *Excedentes Privados por cada Transação e Excedente Total das Unidades Comercializadas*

Quantidade	Excedente do demandante	Excedente do ofertante	Excedente de cada transação	Ganho total da troca
1	4	4	8	8
2	3	3	6	14
3	2	2	4	18
4	$1,50$	1	$2,50$	$20,50$
5	$0,50$	0	$0,50$	21
6	-2	-1	-3	18
7	-4	-2	-6	12
8	-5	-3	-8	4
9	-5	-4	-9	-5
10	-5	-5	-10	-15

Figura 5.8 – *Distribuição do Excedente*



Neste ponto é que se revela a importância da hipótese de que cada demandante demanda apenas uma unidade e de que cada ofertante oferta apenas uma unidade, pois com ela podemos perceber que a existência de excedentes significa que pessoas participam do mercado e que podemos saber, neste exemplo, quem precisamente elas são, conforme a Tabela 5.6

Note que quem determina o preço de equilíbrio são Antônio e Paula. Antônio

Tabela 5.6 – *Nomes das Pessoas que Participam do Mercado de Frangos*

Transação	Excedente do demandante		Excedente do ofertante	
1	João	4	4	Catarina
2	Pedro	3	4	Augusto
3	Maria	2	2	Roberto
4	Carla	$1,50$	1	Aline
5	Antônio	$0,50$	0	Paula

quer pagar, no máximo, R$ 5,50 e Paula quer receber, no mínimo, R$ 5. É o poder de barganha de cada um deles que determinará qual será o preço de equilíbrio, que pode ser qualquer coisa entre R$ 5 e R$ 5,50. Esses são os chamados agentes marginais. Eles são os últimos nas respectivas filas, aqueles que estão praticamente indiferentes entre participar ou não do mercado. Antônio é, dentre os demandantes que participam do mercado, aquele com menor disposição a pagar, é aquele para quem o frango vale menos. Paula é, dentre os ofertantes, aquele com maior disposição a ofertar ou maior custo marginal. Se Paula quiser ofertar um frango, seu custo marginal será R$ 5. Isso quer dizer que os recursos que ela teria de deslocar da economia para ofertar um frango valem R$ 5.

Socialmente só há vantagem nessa quinta transação porque a sociedade, ao abrir mão de R$ 5, consegue gerar um excedente de R$ 0,50 que será dividido entre Paula e Antônio. Paula não está "roubando" da sociedade, pois ela também é parte da sociedade. A sociedade não perde com isso. Os recursos que Paula desloca da economia são remunerados também, cada fator recebe um excedente na forma de pagamento. Os trabalhadores empregados na produção recebem seu salário e o capital recebe sua remuneração, ou seja, as pessoas que abriram mão de dinheiro para utilizá-lo como capital no processo produtivo são compensadas pelo consumo de que abriram mão durante o tempo de produção. Os recursos adicionais usados na produção de Paula, devido à substituibilidade entre bens, implicam redução do uso de alguns dos outros bens. Alguns trabalhadores de alguma firma fornecedora serão demitidos, algum capital não será aceito, mas as perdas não superarão os ganhos, desde que o preço do frango posto à venda por Paula não seja inferior a R$ 5. Tudo isso já está embutido no custo marginal de R$ 5 que Paula enfrenta. Logo, o excedente de R$ 0,50 da quinta transação é um ganho social puro, não é um roubo. Quando Paula tenta maximizar seu lucro barganhando pesado com Antônio, quem ganha é a sociedade. Não é só Paula que ganha. Todas as pessoas que direta ou indiretamente estão envolvidas no deslocamento dos recursos da economia para a produção da quinta unidade ganham.

Além disso, Antônio, o consumidor, ganha também e, ao barganhar pesado pela redução do preço, agindo em seu próprio interesse, ele não afeta o fato de que a sociedade ganha. O que ele afeta é a distribuição desse ganho social. Já Paula tanto gera o ganho social como também, pela barganha, contribui para a determinação da distribuição do ganho. Como esse ganho social marginal deve-se unicamente à decisão privada de Paula e Antônio de participarem da economia, nada mais justo que essa fatia adicional do bolo fique com eles. É isso que economistas despreparados não conseguem entender: eles não enxergam que a sociedade não perde com a liberdade de comércio, que, pelo contrário, ganha e que a busca da satisfação dos interesses privados é o melhor mecanismo para o enriquecimento da nação. Eles não entendem que o produtor é quem aumenta a renda nacional ou o excedente social ao decidir produzir e que o papel do consumidor nessa história toda é, por um lado, contribuir para a distribuição do excedente na sociedade, pela barganha de preços, e, por outro, ser aquilo que motiva o produtor a decidir produzir.

Nosso exemplo do mercado de frango é simples, mas os princípios básicos abordados são válidos para mercados competitivos em situações reais. Note também que os ganhos de troca foram maximizados. Os agentes possuem recursos distintos e diversidade de desejos e capacidades. Mediante o sistema de preços de mercado, na ausência de impedimentos às transações, sejam impedimentos de caráter institucional ou informacional, e sob a condição de direitos de propriedades bem definidos, a alocação de equilíbrio exauriu todos os possíveis ganhos de troca e os distribuiu da melhor maneira.

A pergunta que podemos fazer agora é: "Qual o efeito dos impostos e dos subsídios governamentais sobre o mercado?" A lição básica é que impostos sobre o consumo de bens causam distorções que reduzem a eficiência do mercado. Surpreendentemente, os subsídios também distorcem e reduzem a eficiência. Vou me restringir aqui aos efeitos do imposto apenas. Ao montante monetário dessa perda dá-se o nome de custo de eficiência. Esse é o termo que Arnold Harberger prefere usar no lugar do termo *deadweight loss*.

Para entendermos isso, voltemos ao nosso mercado de frango, em que a quantidade e o preço de equilíbrio eram determinados competitivamente. Suponha que o governo cria um **imposto de R$ 3 sobre a transação de frango**.

Tabela 5.7 – *Preços de Demanda e de Oferta e Excedente Total com Imposto de R$ 3*

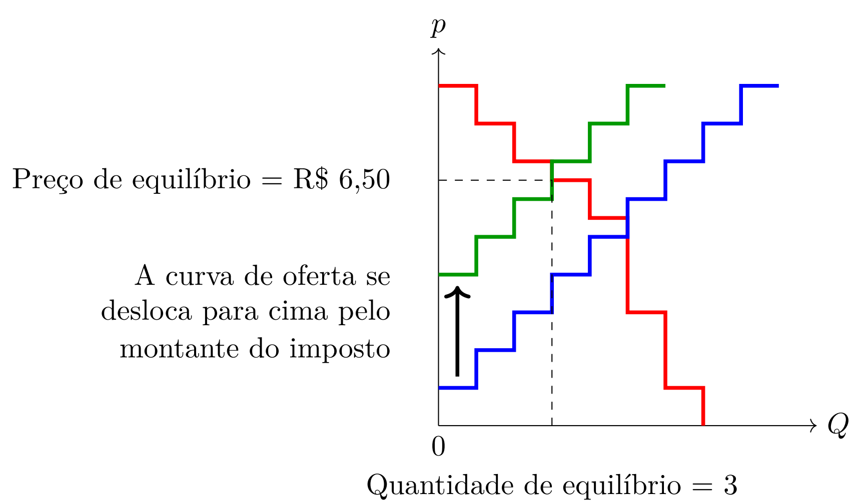
Quantidade	Preço de demanda	Custo marginal	Preço de oferta	Excedente das trocas	Excedente total
1	9	1	4	5	5
2	8	2	5	3	8
3	7	3	6	1	9
4	$6,50$	4	7	$-0,50$	$8,50$
5	$5,50$	5	8	$-2,50$	6
6	3	6	9	-6	0
7	1	7	10	-9	-9
8	0	8	11	-11	-20
9	0	9	12	-12	-32
10	0	10	13	-12	-45

Assim, se um ofertante cobra R$ 5 pelo frango, ele tem que separar R$ 3 e pagar ao governo na forma de imposto, ficando com apenas R$ 2 no bolso. Se a produção do primeiro frango custa a Catarina R$ 1, então ela não mais ofertará essa primeira unidade a R$ 1, mas a R$ 4 = R$ 1 + R$ 3, ou seja, **o preço cobrado será o custo marginal de cada unidade produzida mais o valor do imposto cobrado**.

Ela repassa todo o imposto para o preço do produto. Mas ao contrário do que reza o pensamento popular, não é o consumidor, em geral, que arca inteiramente com esse custo adicional. **Os ofertantes todos também arcarão com parte disso. Afinal, com o preço de venda maior, as vendas serão reduzidas, afetando o excedente total dos ofertantes. Antes, o preço de oferta coincidia com o custo marginal. Agora, o preço de oferta é o custo marginal mais o imposto**.

O excedente das trocas é maximizado com a comercialização de 3 unidades a qualquer preço entre R$ 6,50 e R$ 7. Para simplificar, vamos supor que o preço de equilíbrio é R$ 6,50. O gráfico abaixo ilustra esse caso:

Figura 5.9 – *Curvas de Oferta e de Demanda do Mercado de Frangos com Imposto de R$ 3*



O excedente total dos consumidores é R$ 4,50. De fato, João valoriza a primeira unidade em R$ 9, paga R$ 6,50 por ela e ganha, portanto, R$ 2,50; pela segunda unidade, Pedro paga R$ 6,50, mas como a valoriza em R$ 8, ele ganha R$ 1,50; pela terceira (e última), Maria paga R$ 6,50, mas como ela a valoriza em R$ 7, ela ganha R$ 0,50. Somando os excedentes privados dos demandantes, encontramos R$ 4,50 (= R$ 2,50 + R$ 1,50 + R$ 0,50).

Já para os ofertantes (Catarina, Augusto e Roberto), a receita proveniente da venda de 3 unidades é R$ 19,50 = 3 x R$ 6,50. Só que eles têm que pagar ao governo R$ 3 por cada frango vendido, ou seja, R$ 9 = 3 x R$ 3. Sobram em caixa apenas R$ 10,50. O custo de produzir 3 unidades é R$ 6, de modo que o excedente do produtor é R$ 4,50 (=R$ 19,50 - R$ 9 - R$ 6).

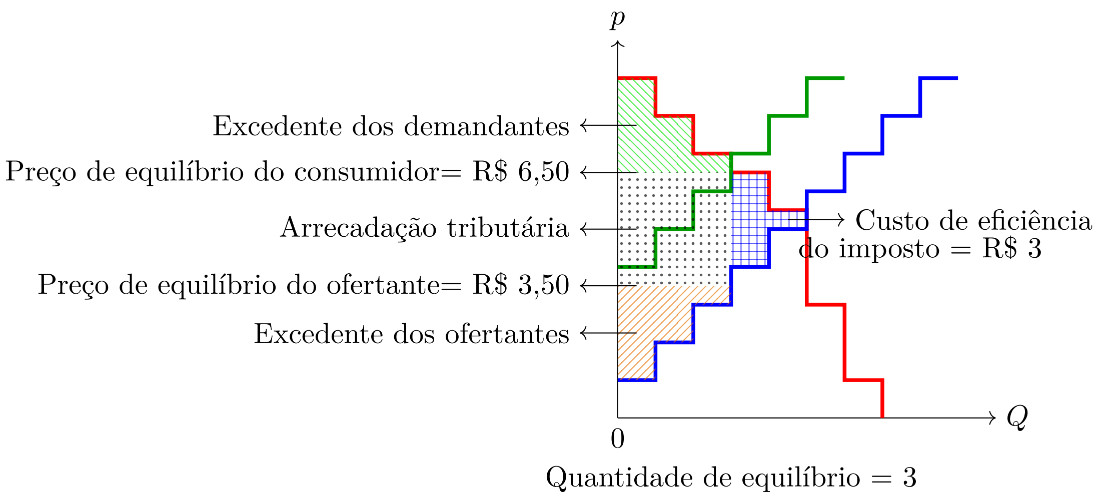
Para nós, os ofertantes são Catarina, Augusto e Roberto. Catarina vende o frango a R$ 6,50, já que esse é o novo preço de equilíbrio. Mas ela tem que pagar R$ 3 de imposto ao governo, retendo, portanto, em seu caixa apenas R$ 3,50 na forma de receita líquida (pós-imposto). Como seu custo marginal é R$ 1, o excedente de Catarina é, por conseguinte, R$ 2,50. Augusto também vende o frango a R$ 6,50 e paga R$ 3 de imposto, retendo, assim, uma receita líquida de R$ 3,50. Como seu custo marginal é R$ 2, seu excedente é R$ 1,50. Finalmente, Roberto vende o frango a R$ 6,50, paga R$ 3 de imposto, retendo uma receita líquida de R$ 3,50 e, como seu custo marginal é R$ 3, seu excedente é R$ 0,50. Note que a soma das receitas líquidas de imposto é R$ 10,50. A soma dos custos marginais de cada um dos ofertantes é R$ 6 (= R$ 1 + R$ 2 + R$ 3), de modo que o excedente total dos ofertantes é R$ 4,50 (=R$ 10,50 - R$ 6). De fato, a soma dos excedentes individuais de Catarina, Augusto e Roberto é R$ 2,50 + R$ 1,5 + R$ 0,50 = R$ 4,50, como já sabemos.

Resumindo, os ofertantes ficaram com um excedente total de R$ 4,50, os demandantes ficaram com um excedente total de R$ 4,50 e o governo ficou com R$ 9. O ganho total é R$ 18 (= R$ 4,50+R$ 4,50+R$ 9). Mas recorde que, no mercado de frango sem impostos, o excedente total era de R$ 21. Portanto, a incidência de imposto sobre o consumo de frango reduziu em R$ 3 os ganhos de troca. A isso chamamos custo de eficiência do imposto.

Na situação sem impostos, a quantidade comercializada em equilíbrio era de 5 unidades. Considere o caso em que o preço de equilíbrio era R$ 5 (recorde que poderia ser qualquer preço entre R$ 5 e R$ 5,50). Com o imposto, a quantidade comercializada caiu para 3 unidades. Logo, o imposto exclui a comercialização das duas últimas unidades. Como essas duas últimas transações deixaram de existir, também desapareceram os ganhos de troca que elas proporcionavam aos consumidores e aos produtores. Quem são essas pessoas expulsas do mercado pelo governo? São elas: Carla, Antônio, Aline e Paula.

Essas duas unidades que deixaram de ser comercializadas proporcionavam aos consumidores expulsos do mercado pelo governo um excedente de R$ 2,50 e aos produtores expulsos um excedente de R$ 0,50. Logo, o custo de eficiência do imposto é o valor dos ganhos de troca provenientes das quantidades que deixaram de ser comercializadas. O Gráfico 5.10 finalmente ilustra o que acontece.

Figura 5.10 – *Curvas de Oferta e de Demanda do Mercado de Frangos com Imposto de R$ 3*



Essa área hachurada denota os ganhos de troca das duas últimas unidades que foram excluídas do comércio pelo imposto.

Até agora nos concentramos na incidência de políticas governamentais: como as intervenções afetam os preços de equilíbrio e os retornos dos fatores. Ou seja, determinamos como as políticas afetam a distribuição dos preços em um mercado. Um segundo conjunto geral de questões é como impostos afetam o todo.

O governo pode aumentar impostos para:

1. aumentar a receita para o fornecimento de bens públicos (estradas, defesa etc.)

2. redistribuir renda

Mas aumentar a receita fiscal geralmente tem um custo de eficiência: para gerar R$ 1 de receita precisamos reduzir o bem-estar dos indivíduos tributados em mais de R$ 1. Os custos de eficiência surgem da distorção do comportamento individual.

Há um grande conjunto de estudos sobre como implementar políticas que minimizam custos de eficiência (tributação ótima). Esta é a teoria central das finanças públicas, que é então adaptada ao estudo dos programas de transferência, seguridade social etc. Começamos com uma análise positiva de como medir o custo de eficiência de um determinado sistema fiscal. Computar o peso morto nos dá o custo de tributação. Vamos ver que esse número não é definido de forma única.

Concluiremos que

- O peso morto aumenta com com o tamanho absoluto das elasticidades ($\eta_S > 0, -\eta_D > 0$) $\Rightarrow$ Mais eficiente para tributar bens relativamente inelásticos
- O peso morto aumenta com o quadrado da alíquota τ: impostos pequenos têm custos de eficiência relativamente baixos, impostos grandes têm custos de eficiência relativamente altos $\Rightarrow$ Mais eficiente para distribuir impostos em todos os bens para manter as taxas de imposto baixas
- Distorções pré-existentes (como uma externalidade positiva que não é corrigida) tornam o custo da tributação mais alto: passar do triângulo para o trapézio

5.6 Excedente Marshalliano e Triângulo de Harberger

Para entendermos como os impostos se traduzem em uma perda de bem-estar vamos formalizar esta situação.

O governo aumenta impostos para aumentar receita para financiar bens ou para redistribuir renda de ricos para pobres. Mas aumentar a receita fiscal geralmente tem um custo de eficiência: para gerar R$ 1 de receita, precisa reduzir o bem-estar dos indivíduos em mais de R$ 1. Os custos de eficiência vêm da distorção do comportamento.

A carga de peso morto (também chamada de sobrecarga) da tributação é definida como a perda de bem-estar criada por um imposto para além da receita fiscal gerada pelo imposto. No diagrama simples de oferta e demanda, o bem-estar é medido pela soma do excedente do consumidor e do excedente do produtor.

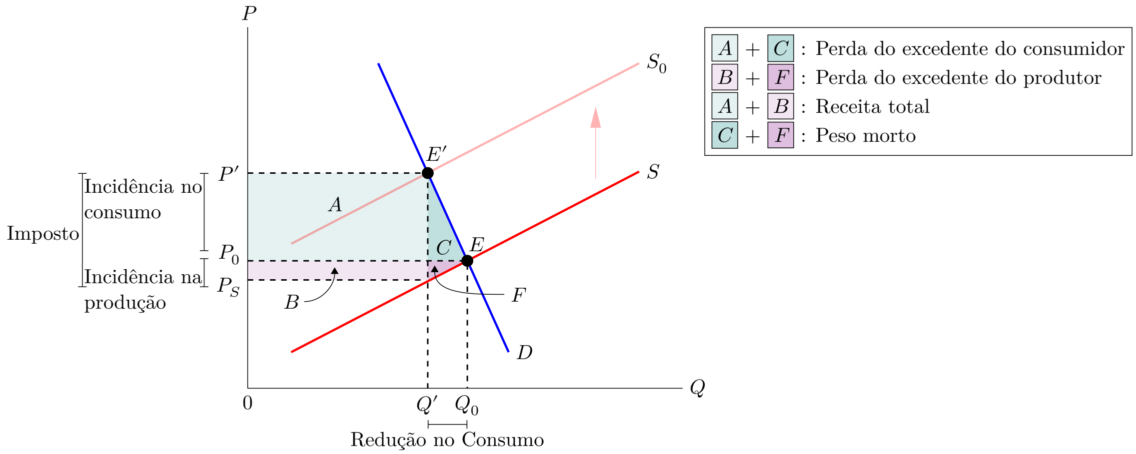
A perda de bem-estar da tributação é medida como uma mudança no excedente do consumidor e no excedente do produtor menos o imposto cobrado. A ineficiência de qualquer imposto é determinada pela medida em que consumidores e produtores mudam seu comportamento para evitar o imposto; a perda de peso morto é causada por indivíduos e empresas que fazem escolhas ineficientes de consumo e produção, a fim de evitar a tributação. Se não houver alteração nas quantidades consumidas, o imposto não tem custos de eficiência.

Dado um mercado, considere a introdução de um imposto específico (*excise tax*) t. Sem perda de generalidade, suponha que o imposto incide sobre a oferta. Assim, o ofertante cobra o custo marginal mais o imposto específico t, deslocando a curva de oferta S paralelamente para cima e para a esquerda até a curva S_0, que dista de S uma distância vertical exatamente igual ao valor do imposto. Graficamente, o peso morto pode ser representado na Figura 5.11.

O custo de eficiência, também conhecido como triângulo de Harberger[38], pode

[38] No seu trabalho empírico, Harberger calcula triângulos de perda de peso morto com base em curvas de demanda construídas ao devolver a receita fiscal aos consumidores de forma global, formando assim uma economia de equilíbrio geral simples, onde todos os mercados se ajustam e nenhum fundo fica sem explicação. Nesse contexto, os impostos afetam os preços e distorcem as decisões individuais, apesar do fato de que as receitas fiscais são eventualmente devolvidas aos consumidores. Como a devolução da receita aos consumidores compensa a quantia dos impostos pagos, mas não compensa a distorção nas decisões individuais, os consumidores são prejudicados pela imposição dos impostos. As curvas de demanda que Harberger utiliza não são demandas marshallianas, uma vez que os consumidores recebem reembolsos fiscais, nem são demandas compensadas ou equivalentes de Hicks, pois as utilidades mudam. Elas são algo diferente: 'demandas Harbergerianas'. Harberger observa que essas curvas de demanda podem ser usadas para gerar medidas de bem-estar que são aproximações de segunda ordem daquelas baseadas em demandas hicksianas; essa afirmação é tornada mais precisa por Diewert (1976) e McKenzie e Pearce (1976), que identificam várias propriedades de medidas de bem-estar baseadas em demandas Harbergerianas.

Figura 5.11 – *Peso Morto*



ser computado como

$$\begin{aligned}
\mathcal{H} &= -\frac{1}{2}(\Delta Q)\,\Delta P \\
&= -\frac{1}{2}(\Delta Q)\,t \\
&= \frac{1}{2}|\Delta Q|t
\end{aligned} \tag{5.98}$$

O custo de eficiência é uma medida da perda social decorrente da distorção de preços causada pela tributação. Vamos deduzir essa medida de forma analítica.

Suponha que as curvas de oferta e demanda são dadas pelas expressões lineares

$$Q^d = a - bP \tag{5.99}$$

$$Q^s = c + dP \tag{5.100}$$

O equilíbrio de mercado é o par (Q_0, P_0). Ao preço de equilíbrio P_0, a quantidade Q_0 é a que iguala quantidade demandada com quantidade ofertada. Para determinar o equilíbrio, fazemos $Q^d = Q^s$, ou seja, a igualdade da oferta com a demanda, de onde temos $a - bP = c + dP$. Portanto, isolando P, encontramos o preço de equilíbrio:

$$P_0 = \frac{a-c}{b+d} \tag{5.101}$$

A esse preço, a quantidade de equilíbrio é:

$$\begin{aligned} Q_0 &= a - bP_0 \\ &= a - b\left(\frac{a-c}{b+d}\right) \\ &= \frac{ab + ad - ab + bc}{b+d} \\ &= \frac{ad+bc}{b+d} \end{aligned} \tag{5.102}$$

Note que encontramos Q_0 substituindo o preço de equilíbrio P_0 na função demanda. Se o tivéssemos substituído na função oferta, o resultado seria o mesmo. De fato:

$$\begin{aligned} Q_0 &= c + dP_0 \\ &= c + d\left(\frac{a-c}{b+d}\right) \\ &= \frac{bc + dc + ad - dc}{b+d} \\ &= \frac{ad+bc}{b+d} \end{aligned} \tag{5.103}$$

Quando o governo introduz um imposto sobre o ofertante, a curva de oferta se desloca para cima e para a esquerda. No novo equilíbrio, a quantidade transacionada é menor: ela cai de Q_0 para Q_1. O ponto crucial é que o preço pago não é mais o preço recebido. À quantidade Q_1, o demandante paga $P^d > P_0$, mas o ofertante recebe $P_s < P_0$. A diferença, como já sabemos, é precisamente o valor do imposto: $t = P^d - P^s$. Queremos encontrar o valor de Q_1, que é a quantidade de equilíbrio após o tributo. O consumidor paga P^d pela quantidade Q_1, isto é, $Q_1 = a - bP^d$, de modo que $P_d = \dfrac{a - Q_1}{b}$. Já o ofertante recebe P^s, ou seja, $Q_1 = c + dP^s$, de modo que $P^s = \dfrac{Q_1 - c}{d}$.

Portanto,

$$\begin{aligned} t &= P^d - P^s \\ &= \frac{a - Q_1}{b} - \frac{Q_1 - c}{d} \\ &= \frac{ad + bc - (b + d)Q_1}{bd} \end{aligned} \tag{5.104}$$

donde $bdt = ad + bc - (b + d)Q_1$.

Logo,

$$\left(\frac{bd}{b + d}\right) t = \frac{ad + bc}{b + d} - Q_1 \tag{5.105}$$

Como $Q_0 = \dfrac{ad + bc}{b + d}$, então

$$Q_1 - Q_0 = -\frac{bd}{b + d} t \quad \iff \quad \Delta Q = \left(-\frac{bd}{b + d}\right) t \tag{5.106}$$

que é a variação da quantidade de equilíbrio em decorrência do imposto.

Temos, assim, uma expressão alternativa para o custo de eficiência:

$$\begin{aligned}
\mathcal{H} &= -\frac{1}{2}\left(\Delta Q\right)t \\
&= -\frac{1}{2}\left(-\frac{bd}{b+d}t\right)t \\
&= \frac{1}{2}\left(\frac{bd}{b+d}\right)t^2
\end{aligned} \tag{5.107}$$

Essa expressão enfatiza o papel das inclinações das curvas de demanda e oferta agregadas no mercado e têm uma função instrumental na teoria, pois as inclinações sugerem a consideração de variações locais.

Quando essas curvas são estimadas econometricamente, surge outra expressão alternativa, baseada nas elasticidades-preço de demanda e oferta. A elasticidade é um conceito local, ou seja, referente a pequenas variáveis de preços.É claro que toda variação observável é discreta, mas, se suficientemente pequena, podemos considerar a elasticidade em termos das derivadas da função. Seja:

$$\eta_0 = \left.\frac{dQ^d}{dP}\frac{P}{Q^d}\right|_{P_0} \tag{5.108}$$

a elasticidade-preço da demanda agregada avaliada ao preço de equilíbrio P_0 pré-imposto e seja, analogamente:

$$\varepsilon_0 = \left.\frac{dQ^s}{dP}\frac{P}{Q^s}\right|_{P_0} \tag{5.109}$$

a elasticidade-preço da oferta.

Como $Q^d = a - bP$ e $Q^s = c + dP$, temos que $\dfrac{dQ^d}{dP} = -b$ e $\dfrac{dQ^s}{dP} = d$. Portanto, substituindo essas derivadas nas expressões para as elasticidades, temos $\eta_0 = -b\dfrac{P_0}{Q_0}$

e $\varepsilon_0 = d\dfrac{P_0}{Q_0}$. Isolando os coeficientes b e d, encontramos $b = -\eta_0\dfrac{Q_0}{P_0}$ e $d = \varepsilon_0\dfrac{Q_0}{P_0}$. Defina $V_0 = P_0Q_0$, que é o valor monetário das transações de equilíbrio. Esse valor equivale ao faturamento que o setor tributado gera em equilíbrio pré-tributo. Seja $\theta = \dfrac{t}{P_0}$ a taxa que expressa a magnitude do tributo como proporção do preço inicial de mercado: $t = \theta P_0$. Substituindo os coeficientes b e d, em termos das elasticidades, na expressão $\mathcal{H} = \dfrac{1}{2}\left(\dfrac{bd}{b+d}\right)t^2$, encontramos:

$$\begin{aligned}
\mathcal{H} &= \frac{1}{2}\left(\frac{bd}{b+d}\right)t^2 \\
&= \frac{1}{2}\left(\frac{-\eta_0\dfrac{Q_0}{P_0}\varepsilon_0\dfrac{Q_0}{P_0}}{-\eta_0\dfrac{Q_0}{P_0}+\varepsilon_0\dfrac{Q_0}{P_0}}\right)t^2 \\
&= \frac{1}{2}\left(\frac{-\eta_0\varepsilon_0}{-\eta_0+\varepsilon_0}\right)\frac{\left(\dfrac{Q_0}{P_0}\right)^2}{\left(\dfrac{Q_0}{P_0}\right)}t^2 \\
&= -\frac{1}{2}\left(\frac{\eta_0\varepsilon_0}{\varepsilon_0-\eta_0}\right)\frac{Q_0}{P_0}t^2
\end{aligned}$$

Ora, $\dfrac{Q_0}{P_0} = \dfrac{P_0Q_0}{P_0^2}$, isto é, $\dfrac{Q_0}{P_0} = \dfrac{V_0}{P_0^2}$. Logo,

$$\begin{aligned}
\mathcal{H} &= -\frac{1}{2}\left(\frac{\eta_0\varepsilon_0}{\varepsilon_0-\eta_0}\right)V_0\left(\frac{t}{P_0}\right)^2 \\
&= -\frac{1}{2}\left(\frac{\eta_0\varepsilon_0}{\varepsilon_0-\eta_0}\right)V_0\theta^2 \\
&= \frac{1}{2}\left(\frac{|\eta_0|\varepsilon_0}{\varepsilon_0+|\eta_0|}\right)V_0\theta^2 \qquad (5.110)
\end{aligned}$$

Esta expressão é útil quando a variação da quantidade de equilíbrio decorrente do imposto é negligenciável relativamente à quantidade comercializada de equilíbrio antes do imposto. A unidade de tempo subjacente que caracteriza a quantidade como taxa de fluxo pode ser o mês, o trimestre ou o ano, sempre o que for mais conveniente para a análise. Dada essa condição, observa-se o faturamento do setor no período, V_0, e considera-se o imposto proposto como percentagem θ do preço de equilíbrio, P_0: esses elementos são facilmente observáveis. O trabalho mais técnico repousa sobre as estimativas das elasticidades-preço de demanda e de oferta do setor tributado, o que requer técnicas econométricas apropriada.

Existem, assim, três expressões equivalentes para o custo de eficiência do imposto:

$$\mathcal{H} = \begin{cases} -\frac{1}{2}\left(\Delta Q\right)t \quad \text{ou} & \frac{1}{2}|\Delta Q|t \\ \frac{1}{2}\left(\frac{bd}{b+d}\right)t^2 & \\ -\frac{1}{2}\left(\frac{\eta_0\varepsilon_0}{\varepsilon_0-\eta_0}\right)V_0\theta^2 \quad \text{ou} & \frac{1}{2}\left(\frac{|\eta_0|\varepsilon_0}{\varepsilon_0+|\eta_0|}\right)V_0\theta^2 \end{cases} \tag{5.111}$$

Segundo o balanço da ABIT (Associação Brasileira da Indústria Têxtil e de Confecção), o faturamento do setor têxtil em 2021 foi estimado em 194 bilhões de reais. Vamos supor, apenas para efeito de exemplificação, que o setor não seja distorcido por tributações pré-existentes e que estudos econométricos estimaram uma elasticidade-preço da demanda em $\eta_0 = -1,2154$ e uma elasticidade-preço da oferta em $\varepsilon_0 = 0,8262$. O governo cogita de fazer incidir sobre os ofertantes uma alíquota de imposto de 10% sobre o preço médio observado no ano. Logo,

$$\begin{aligned} \mathcal{H} &= \frac{1}{2}\left(\frac{|\eta_0|\varepsilon_0}{\varepsilon_0+|\eta_0|}\right)V_0\theta^2 \\ &= \frac{1}{2}\left(\frac{1,2154\times 0,8262}{0,8262+1,2154}\right)\times 194.000.000.000\times(0,1)^2 \\ &= 477.100.000 \end{aligned} \tag{5.112}$$

Ou seja, um custo de eficiência de aproximadamente 477 milhões de reais.

Recorde que essa perda já é líquida da arrecadação tributária do governo. Esse valor consiste de todos os excedentes perdidos dos consumidores expulsos do mercado por que o preço pago subiu e de todos os oferentes expulsos porque o preço recebido caiu.

Um caso extremo é quando a oferta é infinitamente preço-elástica, isto é, variações do preço alteram infinitamente a quantidade ofertada. Em termos gráficos, isso corresponde a uma oferta (inversa) agregada fixa horizontal no plano do preço em função da quantidade. Essa é uma condição natural em mercados nos quais há um número muito grande de ofertantes de um produto homogêneo com a mesma tecnologia de produção. Dizemos que esse é um mercado espesso (*thick market*) no lado da oferta. Analisar esse caso, portanto, é útil quando queremos ter uma estimativa imediata dos custos de eficiência do imposto em mercados espessos. No caso em pauta, $d = \infty$ e $\varepsilon_0 = \infty$. Ora:

$$\begin{aligned} \lim_{\varepsilon_0 \to \infty} \mathcal{H} &= \lim_{\varepsilon_0 \to \infty} \left(\frac{1}{2} \left(\frac{|\eta_0| \varepsilon_0}{\varepsilon_0 + |\eta_0|} \right) V_0 \theta^2 \right) \\ &= \frac{1}{2} |\eta_0| V_0 \theta^2 \end{aligned} \tag{5.113}$$

Alternativamente,

$$\begin{aligned} \lim_{d \to \infty} \mathcal{H} &= \lim_{d \to \infty} \frac{1}{2} \left(\frac{bd}{b + d} \right) t^2 \\ &= \frac{1}{2} b t^2 \end{aligned} \tag{5.114}$$

Em outras palavras a magnitude do custo de eficiência, nesse caso, dependerá diretamente das elasticidades-preço da demanda e da oferta.

5.7 Excedente Marshalliano, Variação Equivalente e Variação Compensatória

Considere que um indivíduo tem a seguinte função de utilidade: $u(c_1, \dots, c_N) = u(c)$. O objetivo do indivíduo é

$$\max_c u(c) \tag{5.115}$$

$$\text{sujeito a} \quad p'c \leq Z \tag{5.116}$$

em que $p' = p + \tau$ denota o vetor de preços com o imposto e Z é a riqueza (pode ser zero).

O lagrangeano deste problema é

$$L(c) = u(c) - \lambda(p'c - Z) \tag{5.117}$$

cujas condições de primeira ordem são:

$$\frac{\partial L}{\partial c} = 0 \quad \Longleftrightarrow \quad u_{c_i} - \lambda p' = 0 \tag{5.118}$$

$$\frac{\partial L}{\partial \lambda} = 0 \quad \Longleftrightarrow \quad p'c - Z = 0 \tag{5.119}$$

O multiplicador da restrição orçamentária é λ (perceba a ideia de preço-sombra). A condição de primeira ordem implica: $u_{c_i} = \lambda$. As condições de primeira ordem em conjunto com a restrição orçamentária determinam as funções de demanda marshallianas (ou não compensadas) $c_i(p', Z)$ e a função de utilidade indireta $v(p', Z)$.

Uma propriedade útil é a identidade de Roy: $v_{p'_i} = \lambda c_i$: o efeito em termos de bem-estar de uma mudança de preço dp'_i é o mesmo que tomar $dZ = c_i dp_i$ do consumidor.

Se representarmos a inclinação da curva de indiferença indireta como função da função de utilidade indireta ou como função da função de dispêndio, obteremos duas relações interessantes entre essas funções e as função de demanda pelo bem 1 ou pelo bem 2.

Uma curva de indiferença indireta pode ser pensada como uma curva ao longo da qual a função de utilidade indireta é mantida constante, do mesmo modo que uma curva de indiferença pode ser pensada como uma curva ao longo da qual a função de utilidade é mantida constante. A inclinação de uma curva de indiferença em um ponto qualquer é dada, sabemos, pelo negativo da razão entre as utilidades marginais do bem 1 e do bem 2. E a inclinação de uma curva de indiferença indireta? De modo similar ela é dada pelo negativo da razão entre a utilidade marginal do preço do bem 1, entendida como a taxa de impacto de uma pequena variação do preço desse bem sobre o valor da função de utilidade indireta $\left(\frac{\partial V}{\partial p'}\right)$, e a utilidade marginal da renda, entendida como a taxa de impacto de uma pequena variação na renda sobre o valor da função de utilidade indireta $\left(\frac{\partial V}{\partial Z}\right)$.

Vamos discutir agora o problema do efeito renda e o problema de *path dependence*. Esta exposição será baseada em Auerbach (1985). Começamos com um vetor de preço p^0 e mudamos para um vetor de preço p^1 (em decorrência da imposição de impostos). O excedente do consumidor (marshalliano) é definido como:

$$CS = \int_{p^0}^{p^1} c(p', Z)dp' \tag{5.120}$$

Qual o problema com esta definição? O excedente do consumidor é dependente do caminho quando o preço muda: de p^0 para p^a e de p^a para p^1. Ou seja,

$$CS(p^0 \to p^a) + CS(p^a \to p^1) \neq CS(p^0 \to p^1) \tag{5.121}$$

Na literatura econômica, há consenso de que o problema da dependência do caminho lida com um fenômeno em que a medição de uma mudança de bem-estar causada por mudanças de preço usando o excedente do consumidor da marshalliana produz resultados que diferem dependendo do caminho de medição escolhido. A composição das cestas ótimas resultam de diferentes caminhos entre o nível de utilidade inicial e o final, ao longo dos quais diferentes preços estão mudando.

Agora considere o caso de dois bens com a incidência de dois impostos (um em cada bem). Então, temos:

$$CS = \int_{p_1^0}^{p_1^1} c_1(p_1, p_2^0, Z)dp_1 + \int_{p_2^0}^{p_2^1} c_2(p_1^1, p_2, Z)dp_2 \tag{5.122}$$

ou

$$CS = \int_{p_2^0}^{p_2^1} c_2(p_1^0, p_2, Z)dp_2 + \int_{p_1^0}^{p_1^1} c_1(p_1, p_2^1, Z)dp_1 \tag{5.123}$$

Problema matemático: para estas duas definições serem equivalentes (*path independence*), as derivadas parciais cruzadas devem ser iguais, isto é, $\dfrac{dc_2}{dp_1} = \dfrac{dc_1}{dp_2}$. Isso não será satisfeito para as funções de demanda marshalliana, a menos que não haja efeitos renda, porque os efeitos renda e os níveis de consumo inicial diferem entre bens. E isso não ocorre em uma função de utilidade geral, apenas no caso de funções quase-lineares. Mas eles são iguais para a demanda hicksiana (compensada) [a matriz de Slutsky é simétrica].

O cálculo do peso morto por meio de demandas marshalliana é atraente, já que é fácil. Mas não é atraente por causa da dependência do caminho. O cálculo do peso morto por meio de demandas hicksianas é atraente porque não há dependência de trajetória. Mas não é atraente porque não é observável e depende da medida de utilidade $h(p, u)$. Se adotarmos o cálculo por demandas compensadas, há dois candidatos naturais (utilidade pré-imposto, utilidade pós-imposto). Isso nos leva a variação compensatória e variação equivalente.

Para traduzir a perda de utilidade em unidades monetárias, introduza a função dispêndio. Fixe o nível de utilidade e os preços e procure a cesta que minimiza o custo para alcançar essa utilidade para esses preços:

$$e(p, U) = \min_{c} pc \tag{5.124}$$

$$\text{sujeito a} \quad u(c) \geq U \tag{5.125}$$

Seja μ o multiplicador na restrição de utilidade, então as condições de primeira ordem dadas por $p_i = \mu u_{c_i}$ e a restrição geram funções de demanda hicksianas (ou compensadas), h, que mapeiam preços e utilidade na demanda

$$c_i = h_i(p, u) \tag{5.126}$$

Defina a perda para o consumidor de aumentar as taxas de imposto como

$$e(p^1, u) - e(p^0, u) \tag{5.127}$$

Essa medida, $e(p^1, u) - e(p^0, u)$, é uma função de valor único e, portanto, é uma medida coerente do custo de bem-estar de uma alteração de imposto para os consumidores. Portanto, não há problema de dependência de caminho. Mas agora, qual u você deve usar? Considere a mudança de preços p^0 para p^1 e suponha que o indivíduo tenha renda Z. Temos que:

- $u^0 = v(p^0, Z)$, utilidade inicial
- $u^1 = v(p^1, Z)$, utilidade ao novo preço p^1

Com estas possibilidades definimos a variação equivalente e a variação compensatória.

1. **Variação compensatória**:

$$CV = e(p^1, u^0) - e(p^0, u^0) = Z - e(p^1, u^0) \tag{5.128}$$

Quanto você precisa para compensar o consumidor para que ele seja indiferente entre ter o imposto e não ter o imposto (para atingir o nível de utilidade original a novos preços). Isto é, $e(p^0, u^0) = e(p^1, u^0) - CV$, em que CV é o valor de suas despesas *ex-post* que eu tenho que cobrir para deixá-lo com a mesma utilidade *ex-ante*.

2. **Variação equivalente**:

$$EV = e(p^1, u^1) - e(p^0, u^1) = e(p^0, u^1) - Z \tag{5.129}$$

Quanto dinheiro o consumidor estaria disposto a pagar como um montante fixo para evitar de ter que pagar o imposto (e atingir o novo nível de utilidade pós-imposto). Isto é, $e(p^0, u^1) + EV = e(p^1, u^1)$, em que EV é o valor extra que posso retirar do indivíduo e deixá-lo com a mesma utilidade *ex-post*.

Exemplo 5.2. *Imagine um consumidor cujas preferências sejam representadas pela função de utilidade* $U(x_1, x_2) = \sqrt{x_1 x_2}$. *Nas condições iniciais, os preços dos dois bens e sua renda são* $p_1^0 = 9$, $p_2^0 = 9$ *e* $Z^0 = 90$. *Nas condições iniciais eles passam a* $p_1^1 = 4$, $p_2^1 = 9$ *e* $Z^1 = 90$. *Vamos calcular as variações compensatória e equivalente associadas a essa mudança. Para tal, observe que a função de utilidade de nosso consumidor é uma função do tipo Cobb-Douglas com os coeficientes somando a unidade. Desse modo, podemos fazer* $a = 1/2$. *A função utilidade indireta e a função dispêndio são, respectivamente:*

$$v(p_1, p_2, Z) = \frac{Z}{2\sqrt{p_1 p_2}} \tag{5.130}$$

$$e(p_1, p_2, u) = 2u\sqrt{p_1 p_2} \tag{5.131}$$

A variação compensatória pode ser encontrada como

$$v(p_1^1, p_2^1, Z^1 - CV) = v(p_1^0, p_2^0, Z^0) \quad \Rightarrow \quad \frac{90 - CV}{2\sqrt{4 \times 9}} = \frac{90}{2\sqrt{9 \times 9}} \quad \Rightarrow \quad CV = 30 \tag{5.132}$$

A variação equivalente pode ser encontrada como

$$v(p_1^0, p_2^0, Z^0 + EV) = v(p_1^1, p_2^1, Z^1) \quad \Rightarrow \quad \frac{90 + EV}{2\sqrt{9 \times 9}} = \frac{90}{2\sqrt{4 \times 9}} \quad \Rightarrow \quad EV = 45 \tag{5.133}$$

Usamos estas definições para mensurar o peso morto. O peso morto será o excesso de EV (CV) sobre as receitas obtidas.

Vamos ver o peso morto graficamente usando variação compensatória e variação equivalente.

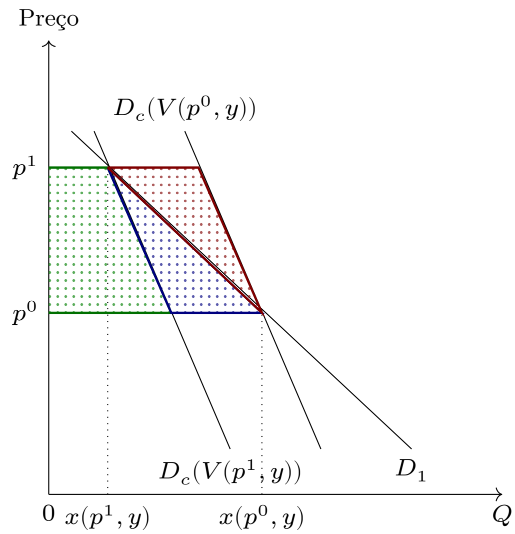
Sejam as funções de demanda hicksianas. Primeiro note que o teorema do envelope implica que $h_i = e_{p_i}(p, u)$. Por isso, pode definir CV ou EV como:

$$e(p^1, u) - e(p^0, u) = \int_{p^0}^{p^1} h(p, u)dp \tag{5.134}$$

Se apenas um preço está mudando, essa é a área sob a curva de demanda de hicksiana para esse bem.

Conforme Auerbach (1985), o excedente do consumidor (o) corresponde à área verde e azul do gráfico, a variação compensatória é a soma das áreas verde, azul e vermelho, e a variação equivalente é a área verde.

Figura 5.12 – *Excedente do Consumidor Marshalliano e Tributação*



Note que $h(p, v(p, Z)) = c(p, Z)$ por causa da dualidade (a solução para o problema maximização deve coincidir com a solução para o problema minimização de gasto no mesmo nível de utilidade indireta). Portanto, os correspondentes hicksianos de CV e EV devem cruzar os marshallianos nos dois preços. Observe que $h(p, u)$ tem uma inclinação mais acentuada que $c(p, Z)$. Isso ocorre porque não há efeito renda. Lembre-se que com uma mudança de preço, $EV < CS < CV$, mas não é verdade com várias mudanças de preço, porque o excedente do consumidor (marshalliano) não está bem definido.

O custo da tributação é, portanto, a mudança no excedente do consumidor menos os impostos pagos, isto é, o que é perdido em excesso de impostos pagos. Além da medida marshalliana, temos duas medidas desse custo, correspondendo a EV e CV:

$$CS(u^1) = EV - (p^1 - p^0)h(p^1, u^1) \quad \text{[Mohring 1971]} \tag{5.135}$$

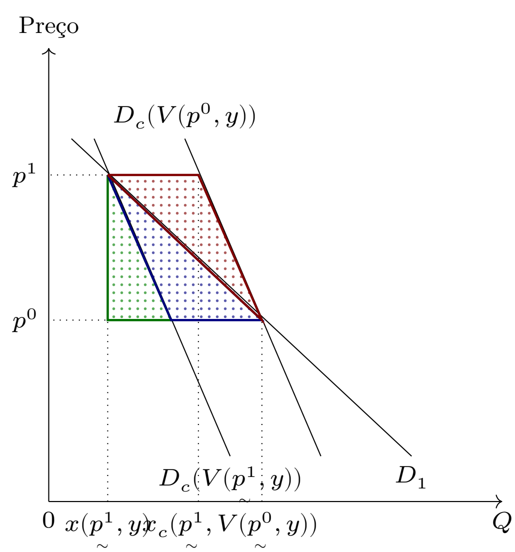
$$CS(u^0) = CV - (p^1 - p^0)h(p^1, u^0) \quad \text{[Diamond e McFadden 1974]} \tag{5.136}$$

A área verde corresponde à variação equivalente, a área vermelha é a variação compensatória e a soma das áreas verde e azul representam o equivalente ao excedente do consumidor marshalliano. As três medidas diferem e correspondem ao peso morto.

Observações desta figura:

1. Em geral, as três medidas do peso morto serão diferentes.
2. EV e CV não mais suportam o excedente do consumidor marshalliano.
3. O ponto chave é que a medida marshalliana superestima o peso morto.
4. No caso especial, sem efeitos renda (utilidade quase-linear), então $CV = EV$ e há uma definição única de excedente do consumidor e peso morto.

Figura 5.13 – *Excedente do Consumidor Hicksiano e Tributação*



A possibilidade de obter as medidas de variação compensatória e de variação equivalente como variações da área abaixo da curva de demanda compensada e acima da linha de preço, parece, à primeira vista promissora: ela sugere que não precisamos conhecer as preferências do consumidor para mensurar variações em seu nível de bem-estar. Precisamos apenas conhecer suas curvas de demanda compensada. O único problema é que usualmente não somos capazes de observar diretamente a demanda compensada do consumidor, mas apenas sua demanda marshalliana, isto é, podemos estimar como a demanda por um bem responde a variações nos preços e na renda do consumidor. Embora teoricamente seja possível, empregando a equação de Slutsky, deduzir o formato das funções de demanda compensada a partir das funções de demanda marshallianas, isso pode implicar procedimentos computacionais complicados. Além disso, limitações frequentes de dados, podem fazer com que as demandas marshallianas estimadas sejam incompletas no número de variáveis consideradas, ou mesmo, como ocorre com grande frequência, estimadas apenas para um agregado de consumidores. Isso dificulta enormemente, quando não impossibilita, a tarefa de deduzir as funções de demanda compensada a partir de funções estimadas de demanda marshalliana.

Considere inicialmente, o caso de um bem cuja demanda não seja afetada por variações na renda. Isso ocorre, por exemplo, caso as preferências do consumidor sejam quase-lineares. Nesse caso, como compensações na renda não alteram a quantidade demandada, as curvas de demanda marshalliana e compensada coincidirão, para quaisquer níveis de renda e utilidade. Isso significa que a medida da variação no excedente do consumidor associada à mudança no preço de um bem será a representação exata tanto da variação compensatória quanto da variação equivalente associadas a essa mudança de preço. Isso sugere também que, desde que a demanda pelo bem em questão seja pouco elástica em relação à renda, a variação no excedente do consumidor será uma boa aproximação das medidas de variação compensatória e equivalente. O mesmo deve ocorrer quando uma variação no preço de nosso bem implicar uma necessidade de variação muito pequena na renda do consumidor para compensar o impacto sobre seu bem-estar.

Além disso, desde que possamos identificar se o bem analisado é um bem normal ou inferior, também podemos dizer em que direção estamos errando quando empregamos a variação no excedente do consumidor como uma medida aproximada da variação compensatória ou da variação equivalente. Podemos, portanto, resumir nossas conclusões com relação à comparação entre as medidas de variação compensatória, variação equivalente e variação no excedente do consumidor, associadas à mudança no preço de um bem, da seguinte maneira:

1. Caso a demanda do bem em questão não seja afetada por alterações na renda do consumidor, como ocorre no caso de preferências quase-lineares, teremos

$$CV = CS = EV \tag{5.137}$$

2. Caso se trate de um bem normal, isto é, caso a demanda desse bem seja crescente em relação à renda do consumidor, então deverão valer as desigualdades

$$CV < CS < EV \tag{5.138}$$

3. Caso, o bem seja um bem inferior, ou seja, caso sua demanda seja decrescente em relação à renda do consumidor, então teremos

$$CV > CS > EV \tag{5.139}$$

Nos dois últimos casos, a variação no excedente do consumidor não pode ser considerada uma medida precisa nem da variação compensatória nem da variação equivalente. Ainda assim, ela indica um limite inferior para uma dessas variações e um limite superior para a outra. Nesse sentido, ela, pode ser justificada como um “compromisso intermediário” entre as medidas de variação compensatória e de variação equivalente.

A diferença entre a demanda marshalliana e a demanda hicksiana (compensada) está na existência do efeito renda na demanda marshalliana (além do efeito substituição). Se a elasticidade da renda na demanda pelo bem (efeito renda) for próxima de zero, então a demanda marshalliana será próxima da demanda hicksiana e $CS \approx CV \approx EV$.

Exemplo 5.3. *Considere um consumidor cujas preferências sejam representadas pela função de utilidade $U(x_1, x_2) = x_1^{1/10} x_2^{9/10}$. A renda desse consumidor e o preço do bem 2 são mantidos constantes em, respectivamente, $Z = 1000$ e $p_2 = 1$, mas o preço do bem 1 sofre uma variação de $p_1^0 = 10$ para $p_1^1 = 8$. Vamos calcular a variação compensatória, a variação equivalente e a variação no excedente do consumidor associadas a essa variação no preço do bem 1.*

A demanda marshalliana pelo bem 1 é dada por

$$x_1(p_1, 1, 100) = \frac{100}{p} \tag{5.140}$$

A variação no excedente do consumidor é encontrada integrando-se essa função entre $p_1 = p_1^0 = 8$ e $p_1 = q_1^1 = 10$:

$$CS = \int_8^{10} \frac{100}{p} dp = 100(\ln 10 - \ln 8) \approx 22,31 \tag{5.141}$$

Para calcularmos a variação equivalente e a variação compensatória, primeiramente determinamos os níveis de utilidade inicial u^0 e final u^1:

$$u^0 = a^a (1-a)^{1-a} \frac{Z}{(p_1^0)^a (p_2^0)^{1-a}} = \left(\frac{1}{10}\right)^{1/10} \left(\frac{9}{10}\right)^{9/10} \frac{1000}{10^{1/10} 1^{9/10}} \approx 573,88 \tag{5.142}$$

$$u^1 = a^a (1-a)^{1-a} \frac{Z}{(p_1^1)^a (p_2^1)^{1-a}} = \left(\frac{1}{10}\right)^{1/10} \left(\frac{9}{10}\right)^{9/10} \frac{1000}{8^{1/10} 1^{9/10}} \approx 586,83 \tag{5.143}$$

A variação compensatória e a variação equivalente são, respectivamente:

$$CV = Z - e(p_1^1, p_2, u^0) = 1000 - 573,88 \frac{8^{1/10} 1^{9/10}}{\left(\frac{1}{10}\right)^{1/10} \left(\frac{9}{10}\right)^{9/10}} \approx 22,07 \tag{5.144}$$

$$EV = e(p_1^0, p_2, u^1) - Z = 586,83 \frac{10^{1/10} 1^{9/10}}{\left(\frac{1}{10}\right)^{1/10} \left(\frac{9}{10}\right)^{9/10}} - 1000 \approx 22,57 \tag{5.145}$$

Nesse exemplo, a variação no excedente do consumidor difere pouco - aproximadamente 1% - tanto da variação compensatória quanto da variação equivalente.

Exemplo 5.4. *Qual é o fardo excedente (também conhecido como perda de peso morto) de um imposto de $1 em roupas, se o preço da comida fosse de $1, o preço original (sem imposto) da roupa era de $1, a renda do consumidor era de $1000 e as preferências do consumidor pode ser representado pela função de utilitário*

$$U = F + 40\sqrt{C} \tag{5.146}$$

em que C é a quantidade consumida de roupas e F a quantidade consumida de alimentos, e a curva de demanda compensada por roupas tem a equação

$$C = \frac{400}{(P_C)^2} \tag{5.147}$$

em que P_C é o preço que ela paga pelas roupas.

Quanto o governo teria que compensar o consumidor pelos danos causados pelo imposto?

Sem impostos a demanda dos consumidores por roupa é

$$C = \frac{400}{1} = 400 \tag{5.148}$$

e sobra 600 *reais para gastar em comida. A sua utilidade é*

$$U = 600 + 40\sqrt{400} = 1400 \tag{5.149}$$

Com o imposto a demanda por roupas é

$$C_{imp} = \frac{400}{2^2} = 100 \tag{5.150}$$

para que, se ela tiver M reais para gastar, gaste $2 \times 100 = 200$ *reais em roupas e* $M - 200$ *para gastar em comida, para que sua utilidade seja*

$$U_{imp} = M - 200 + 40\sqrt{100} = M + 200 \tag{5.151}$$

Para que sua utilidade permaneça inalterada após o imposto, ela precisa de uma renda total M, de modo que $U_{imp} = U$ *ou* $M = 1200$. *Como sua renda original era de* 1000 *reais, a compensação exigida é de* 200 *dólares.*

A compensação exigida é a área sob a curva de demanda compensada da pessoa por roupas, entre o preço sem impostos de $P_C = 1$ *e o preço com impostos de* $P_C = 2$. *A área sob uma curva que representa graficamente uma função é a integral dessa função. Então*

$$CV = \int_1^2 D(P_C)dP_C = \int_1^2 \frac{400}{(P_C)^2}dP_C \tag{5.152}$$

Logo,

$$CV = -\frac{400}{P_C}\bigg|_1^2 = -200 - (-400) = 200 \tag{5.153}$$

Exemplo 5.5. *Suponha que a função dispêndio de um consumidor seja dada por* $e(p_x, p_y, u) = p_x u - 16\frac{(p_x)^2}{p_y}$. *Suponha que o preço do bem x seja 1 e o preço do bem y seja 1. O governo decide tributar o bem y por meio de um imposto específico de R\$ 1 sobre o bem y. A utilidade inicial é* $u = 36$.

1. *Encontre as funções de demanda hicksianas pelos bens* x *e* y.

A demanda hicksiana pelo bem x é

$$x^H(p_x, p_y, u) \equiv \frac{\partial e(p_x, p_y, u)}{\partial p_x} = u - 32\frac{p_x}{p_y} \tag{5.154}$$

A demanda hicksiana pelo bem y é

$$y^H(p_x, p_y, u) \equiv \frac{\partial e(p_x, p_y, u)}{\partial p_y} = 16\left(\frac{p_x}{p_y}\right)^2 \tag{5.155}$$

2. *Qual o custo do imposto para o consumidor?*

   Por definição, o custo para o consumidor do imposto é a mudança no gasto necessário para atingir o nível de utilidade especificado. Ou seja, o custo do imposto sobre o bem y é considerado

$$e(p_x, p_y + t, u) - e(p_x, p_y, u) \tag{5.156}$$

   em que t é o imposto sobre o bem y e p_y é o preço líquido do bem y. Aqui $p_x = p_y = 1$, $t = 1$ e $u = 36$. Então o custo do imposto é

$$e(1, 2, 36) - e(1, 1, 36) \tag{5.157}$$

   Dada a definição da função de despesa na questão

$$\begin{aligned} e(1, 2, 36) &= 36 - 16\frac{1}{2} = 28 \\ e(1, 1, 36) &= 36 - 16\frac{1}{1} = 20 \end{aligned} \tag{5.158}$$

   de modo que o custo do imposto, para o consumidor, seja $28 - 20 = 8$.

3. *Qual a receita dos impostos para o governo?*

   A receita tributária é o imposto por unidade de y, multiplicado pelo número de *unidades de y que o consumidor escolhe comprar. A partir da declaração da pergunta, sua quantidade demandada é*

$$y^H(1,2,36) = 16\frac{1^2}{2^2} = 4 \tag{5.159}$$

para que a receita tributária ty seja igual a 4*.*

5.8 Triângulo de Harberger: Uma Visão Histórica

A sabedoria econômica predominante é altamente crítica a políticas fiscais ou regulamentos governamentais injustificados, externalidades não corrigidas, práticas monopolísticas não controladas e várias outras falhas de mercado. Quando os economistas são desafiados a quantificar os custos econômicos das distorções de preços associadas, é prática padrão – e desde a década de 1960 – usar um pequeno número de suposições para estimar as área relevantes dos “triângulos de Harberger”. Um exercício simples e direto tem inúmeras aplicações e a virtude de produzir respostas em vez de conjecturas.

Em dois artigos influentes publicados em 1964, Arnold C. Harberger (1964a) ofereceu uma derivação clara e persuasiva do método do triângulo para analisar a perda de peso morto e (1964b) aplicou o método para estimar as perdas de peso morto devido ao imposto de renda nos Estados Unidos. Tendo usado anteriormente o método do triângulo para calcular o tamanho da distorção induzida por monopólio na economia dos EUA (1954), imposto de renda corporativo (1959a) e várias e diversas distorções na economia chilena (1959b), Harberger produziu logo em seguida (1966) estimativas do custo do bem-estar dos impostos sobre o capital dos EUA. Em uma pesquisa subsequente, Harberger (1971) esclareceu vários aspectos desse método e abordou várias de suas deficiências percebidas.

Embora a teoria da mensuração da perda de peso morto estivesse bem estabelecida na década de 1950, os economistas muito raramente estimavam as perdas de peso morto antes do aparecimento do trabalho de Harberger. Os trabalhos de Harberger ilustraram as técnicas, a utilidade e a possibilidade realista de realizar tais cálculos e, ao fazê-lo, deram início a uma nova geração de trabalhos. Os triângulos de perda de peso morto tornaram-se conhecidos como “triângulos de Harberger”, devido à ampla influência dos trabalhos de Harberger em pesquisas subsequentes. As novas estimativas de perda de peso morto, por sua vez, levaram a uma importante reavaliação dos efeitos no bem-estar de múltiplas distorções em uma economia e a consideração da possibilidade de que várias falhas de mercado possam transformar os triângulos de Harberger em trapézios ainda maiores.

Compreendeu-se muito antes da década de 1960 que o custo do bem-estar de um imposto sobre mercadorias poderia ser aproximado pelo tamanho do que mais tarde viria a ser chamado de triângulo de Harberger. Jules Dupuit (1844) é geralmente creditado como o primeiro a observar que as curvas de demanda poderiam ser usadas para inferir os efeitos no bem-estar das mudanças de preço. Dupuit era um engenheiro que estava interessado em aplicar princípios econômicos para avaliar obras públicas e não ficou impressionado com a ortodoxia econômica existente (em grande parte devido a Jean-Baptiste Say) que igualava o valor de um bem ao seu preço de mercado. Dupuit observa que os gastos do consumidor constituem um limite inferior à avaliação total de um item. Ele produziu um diagrama muito parecido com o de oferta e demanda, em que as variações de preços induzidas por impostos ou pedágios reduzem a satisfação do consumidor em mais do que a receita que eles geram. Dupuit (1844 [1969], p. 281), tendo o cuidado de distinguir as transferências de recursos (dos consumidores para o estado) através de impostos e pedágios e perdas de recursos devido à ineficiência, descreve o triângulo da perda de bem-estar como “*the utility lost both to the taxpayers and the fisc [the public sector]*”. Ele continuou fazendo outras observações perspicazes, como a de que a área desse triângulo de perda de bem-estar é geralmente uma função do quadrado da alíquota do imposto, e que existe uma taxa de imposto além da qual a receita tributária cai.

A visão de Dupuit foi compartilhada por Fleeming Jenkin, que, como Dupuit, era um engenheiro talentoso, mas não conhecia o trabalho de Dupuit na década de 1840. Jenkin foi estimulado pelo trabalho de Stanley Jevons, professor da Universidade de Manchester, na Inglaterra, que, em seu clássico *The Theory of Political Economy* (1871), discute o efeito das mudanças de preço na utilidade. Jevons, que também aparentemente não tinha conhecimento do trabalho de Dupuit, observa corretamente (1871 [1911], p. 147) que, para um pequeno aumento no preço de uma mercadoria, a utilidade de um consumidor cai em uma quantidade igual ao produto da mercadoria. Jevons expressa reservas, no entanto, sobre a expansão desse método para considerar mudanças não triviais de preço, devido à não constância da utilidade marginal da renda.

Essa abordagem foi insatisfatória para Jenkin, que redescobriu o uso de curvas de oferta e demanda (Jenkin, 1870); e depois aplica a análise para calcular a incidência de um imposto (a distribuição de sua carga entre compradores e vendedores) e a perda de eficiência que ele cria. Jenkin argumenta (1871/72, pp. 109-110) que seu método de determinar o custo para os consumidores (e o de Dupuit, por acaso) é superior à abordagem de Jevons porque:

> *[...] utility, as he [Jevons] defines it, admits of no practical measurement, and he bases his curve, not on the varying estimates of value set by different individuals each on what he has or what he wants, but on the varying utility to each individual of each increment of goods. The above estimate of the gain due to trade deduced from the demand and supply curves as originally drawn in my Recess Studies article is, I believe, novel, and gives a numerical estimate in money of the value of any given trade, which might be approximately determined by observing the effect of a change of prices on the trade; the curves throughout their whole lengths could certainly not, in most cases, be determined by experiment, but statistics gathered through a few years would show approximately the steepness*

of each curve near the market price, and this is the most important information.

A intuição que Dupuit e Jenkin oferecem na obtenção de seus triângulos de bem-estar é essencialmente a mesma hoje. No entanto, vários detalhes sobre os quais eles não se pronunciam se tornaram o foco de mais de um século de pesquisas subsequentes. Uma das primeiras perguntas foi levantada por Léon Walras, professor de Lausanne e fundador da moderna teoria do equilíbrio geral. Ele estava bem ciente do trabalho de Dupuit e, embora admitisse (1874 [1954], p. 443) que fosse "muito completo e engenhoso", sentiu-se compelido a "chamar a atenção para um erro flagrante que Dupuit cometeu em um caso". Walras explica que a disposição do consumidor de pagar por um item é uma função não apenas do seu preço, mas também da renda do consumidor e da utilidade potencialmente disponível ao consumir todas as outras mercadorias. Walras sentiu evidentemente que essa dependência torna o conceito de excedente do consumidor de Dupuit muito específico da situação para servir como uma medida objetiva de satisfação derivada do consumo e, portanto, inapropriada para a medição da perda de peso morto.

Uma defesa prática do método de Dupuit é fornecida por Harold Hotelling, que argumenta que ele depende de propriedades das curvas de demanda que provavelmente serão satisfeitas. Hotelling era inicialmente um estatístico matemático, apesar de ter ocupado um cargo no Departamento de Economia da Columbia de 1931 a 1946, e suas ocasionais incursões na teoria econômica representavam grandes contribuições para a teoria da demanda, a teoria do oligopólio, as finanças públicas e a teoria dos recursos naturais. Em um artigo baseado em seu discurso presidencial à Sociedade Econométrica, Hotelling (1938, p. 242) diz: "neste artigo, apresentaremos de forma revisada um argumento devido essencialmente ao engenheiro Jules Dupuit [...] ". Hotelling analisa a dificuldade de aplicar o método de Dupuit nos casos em que vários preços mudam simultaneamente. Parece que essas situações podem ser tratadas calculando-se os triângulos de perda de peso morto separadamente para cada mercadoria tributada, somando as áreas para obter um total. Infelizmente, a ordem na qual cada preço é alterado afeta o total calculado da perda de peso morto! Como, para mudanças simultâneas de preços, a ordem é perfeitamente arbitrária, uma multiplicidade de respostas reflete que algo no cálculo está certamente errado. Hotelling observa que esse problema desaparece se as chamadas " condições de integrabilidade" forem atendidas. Isso significa essencialmente que as derivadas da curva de demanda entre preços são simétricos; isto é, para quaisquer dois bens i e j, a mudança na quantidade do bem i consumida como resultado de uma mudança unitária no preço de j é igual à mudança na quantidade de bem j consumida como resultado de uma unidade mudança no preço do bem i. Nesse caso, a alteração calculada no bem-estar do consumidor devido a vários preços que variam ao mesmo tempo não é afetada pela ordem em que ocorrem as alterações de preço. No entanto, as condições de integrabilidade são satisfeitas pelas curvas de demanda comuns apenas se os efeitos renda forem inexistentes (o que é possível apenas para um subconjunto de mercadorias) ou se tiverem características muito especiais (como as geradas pelas preferências homotéticas). Hotelling invoca seu trabalho anterior (1932) para argumentar que é improvável que os efeitos renda sejam grandes o suficiente para tornar as curvas de demanda comuns inadequadas para a construção do triângulo no estilo de Dupuit.

Sir John Hicks, professor de Oxford e autor de *Value and Capital* (1946), reavaliou a concepção de excedente do consumidor de Marshall e a imprecisão restante na noção de Marshall da curva de demanda. Hicks considerou o experimento conceitual de compensar totalmente os consumidores pelos efeitos das mudanças de preço em sua renda real, no processo de rastreamento de curvas de demanda correspondentes a diferentes preços. Ele batizou como "variação compensatória" a área entre essa curva de demanda e a linha de preço inicial. Hicks também considerou o efeito sobre as demandas do mercado de extrair dos consumidores o dinheiro que eles pagariam pelo bem para evitar mudanças de preço, usando a "variação equivalente" para se referir à área entre a linha de preço inicial e as curvas de demanda assim geradas.

Portanto, Hicks descreve dois métodos de construção de curvas de demanda que podem ser usados para medir triângulos de perda de peso morto; ambos mantêm a utilidade constante, mas diferem porque são baseados em diferentes níveis de utilidade. A variação compensatória efetivamente mantém a utilidade no nível obtido antes da mudança de preço, enquanto a variação equivalente mantém a utilidade no nível obtido após a alteração do preço.

Medidas de bem-estar do consumidor baseadas em variações compensatórias ou equivalentes têm propriedades desejáveis que intrigaram economistas que trabalham nesta área desde então. Especificamente, as curvas de demanda compensada nas quais elas se baseiam atendem às condições de integrabilidade da Hotelling, tornando os cálculos de bem-estar resultantes definidos de maneira única, mesmo que vários preços mudem simultaneamente. Mudanças endógenas na utilidade marginal da renda não afetam tais medidas, uma vez que os níveis de utilidade são (por construção) mantidos constantes. No entanto, a variação compensatória difere da variação equivalente, e cada uma delas difere do excedente o do consumidor, no qual a renda é mantida constante, na medida em que os efeitos renda são importantes.

O próprio Hicks não se impressionou com a provável importância da distinção entre medidas de bem-estar construídas usando curvas de demanda compensadas e marshallianas. É fácil perceber por que, uma vez que a elasticidade da demanda compensada difere da elasticidade da demanda não compensada correspondente apenas pela propensão marginal do consumidor a gastar no bem em questão. A menos que uma mercadoria represente uma fração extremamente grande do orçamento de um consumidor, as elasticidades de demanda compensadas e não compensadas não diferem muito e quaisquer diferenças entre elas provavelmente serão muito menores do que a incerteza estatística associada às estimativas de elasticidade da demanda. Hicks (1943, p. 40) observa na conclusão de um artigo em que avalia medidas compensadas: “quando, em um artigo anterior [Hicks, 1941], considerei pela primeira vez a possibilidade do tipo de análise que estive aqui trabalhando, rejeitei-o como ‘um negócio complicado, que provavelmente não é de muita importância'. E isso ainda é válido. Não obstante, fico feliz por ter me esforçado para realizá-lo". Hicks (1945–46, p. 68) acrescenta mais tarde: "para a maioria dos propósitos, para os quais a análise do excedente do consumidor é útil, a medida de Marshall é aproximação suficiente; mas, para propósitos claros, é necessário distinguir as medidas básicas e esclarecer suas relações."

Em meados da década de 1950, estava claro que o conhecimento das condições de demanda e oferta era suficiente para calcular as perdas de bem-estar devido a preços distorcidos – mas também que certos ajustes podem ser necessários na aplicação de funções de demanda comuns para esse fim. Nesse momento, no entanto, surgiram dois desenvolvimentos importantes na teoria econômica que não estavam conectados à análise tradicional de bem-estar, mas complicaram qualquer tentativa de medir triângulos de perda de peso morto. O primeiro foi o surgimento de uma rigorosa teoria de equilíbrio geral, que parecia implicar a inadequação de analisar isoladamente um mercado único, uma vez que todos os mercados de uma economia se influenciam. Em particular, Corlett e Hague (1953-54) e Lipsey e Lancaster (1956-57) chamaram a atenção para a possibilidade de que a introdução de distorções em um mercado possa aumentar a eficiência da economia, mitigando os efeitos das distorções em outros lugares. O segundo desenvolvimento foi a análise das dificuldades de obter regras de decisão social bem comportadas a partir das preferências individuais, o que sugeria a impossibilidade de produzir uma medida social geral de políticas que afetam consumidores heterogêneos – de fato, o próprio bem-estar social se tornou um conceito problemático. Como os modelos de equilíbrio geral em escala real eram praticamente inexistentes e como quase qualquer distorção concebível afeta o bem-estar de vários consumidores, a análise aplicada do excedente do consumidor foi frustrada.

Harberger estava bem ciente dessa história intelectual e dos problemas complicados com que se relaciona. Ao motivar sua apresentação dos métodos de estimativa de perda de peso morto, Harberger (1964a, pp. 58–59) escreve:

The measurement of deadweight losses is not new to economics by any means. It goes back at least as far as Dupuit; and more recently Hotelling, Hicks, Debreu, Meade, and H. Johnson have made important contributions. Nonetheless I feel that the profession as a whole has not given to the area the attention that I think it deserves. We do not live on the Pareto frontier, and we are not going to do so in the future. Yet policy decisions are constantly being made which can move us either toward or away from that frontier. What could be more relevant to a choice between policy A and policy B than a statement that policy A will move us toward the Pareto frontier in such a way as to gain for the economy as a whole, say, approximately $ 200 million per year, while policy B will produce a gain of, say, about $30 million per year? What could be more useful to us as a guide to priorities in tax reform than the knowledge that the deadweight losses stemming from the tax loopholes (percentage depletion and capital gains) open to explorers for oil and gas are probably greater in total magnitude than the deadweight losses associated with all the other inefficiencies induced by the corporation income tax? What could be more tantalizing than the possibility (which I believe to be a real one) that the U.S. tariff, whose indirect effect is to restrict the equilibrium value of U.S. exports, produces by this route a gain for the U.S. from a partial exploitation of U.S. monopoly power in world markets which nearly offsets (or perhaps fully or more than fully offsets) the efficiency-losses produced by tariff-induced substitution of more expensive domestic products for cheaper imports? These and similar questions seem to me so interesting, so relevant, so central to our understanding of the economy we live in, that I find it hard to explain why the measurement of deadweight losses should be the province of only a handful of economists rather than at least the occasional hobby of a much larger group.

Os resultados empíricos do trabalho de Harberger são úteis e interessantes. Harberger (1954) constata que a alocação incorreta de recursos devido ao comportamento monopolista na indústria dos EUA gera uma ineficiência igual a aproximadamente 0,1% do PNB dos EUA; que o imposto de renda corporativo dos EUA gera distorções avaliadas em US$ 1 bilhão (0,5% do PNB) anualmente (1959a); que a má alocação de recursos de vários tipos reduz o bem-estar chileno em 15% (1959b); que distorções nas escolhas de trabalho e lazer induzidas pelo imposto de renda pessoal dos EUA reduzem o bem-estar em US$ 1 bilhão (0,4% do PNB) anualmente (1964b); e que todos os impostos sobre a renda dos EUA juntos são responsáveis por perdas econômicas de US$ 2 bilhões (0,8% do PIB) anualmente (1966). Esses resultados se mostraram robustos para a subsequente escolha cuidadosa e o retrabalho, uma vez que os efeitos de cálculos alternativos e especificações metodológicas tendem a se anular.

Por último, pode-se perguntar por que os triângulos familiares de bem-estar levam o nome de Harberger e não, digamos, de Dupuit ou Jenkin ou, de Hotelling. De fato, o termo "triângulo de Harberger" nunca foi usada por Harberger. A primeira referência publicada ao "triângulo de Harberger" é em Rosenberg (1969, p. 173). Em 1980, o termo "triângulo de Harberger" havia penetrado suficientemente no léxico econômico, de tal modo que a pesquisa sobre excesso de carga por Auerbach e Rosen (1980, p. 306) se refere ao "familiar triângulo de Harberger".

Os triângulos Harberger influenciaram o curso subsequente de pelo menos duas correntes de pesquisa econômica: a medição empírica da eficiência econômica e o desenvolvimento da microeconomia normativa aplicada, particularmente nas áreas de finanças públicas e escolha pública.

Harberger (1964a) repreende gentilmente os economistas por sua relutância em medir as perdas de bem-estar devido às distorções econômicas. No início da década de 1960, tais medições podem ter sido um marco na análise profissional, uma vez que a economia estava crivada de fontes potenciais de distorções econômicas significativas. Por exemplo, de 1954 a 1963, a maior taxa marginal de imposto de renda federal sobre indivíduos nos Estados Unidos foi de 91%; em 1964, caiu para 77 por cento e foi de 70 por cento de 1965 a 1980. A taxa de imposto federal sobre as empresas foi de cerca de 50 por cento de 1954 até o final dos anos 1970. Se essas altas taxas de impostos geraram distorções econômicas significativas, é claro, uma questão empírica - mas é uma questão que os economistas podem ter visto como óbvia para tentar responder. Além disso, o início da década de 1960 foi o início de uma era em que a econometria aplicada estava se consolidando, e o trabalho empírico era de maior qualidade e perfil do que antes. Havia um crescente corpo de dados econômicos, técnicas econométricas e poder de computação, e havia um desenvolvimento considerável da teoria da medição do bem-estar. No entanto, no início da década de 1960, os economistas tinham muito pouco a dizer sobre as magnitudes reais da perda de peso morto.

O próprio Harberger estava bastante disposto a aplicar métodos de triângulo para estimar as magnitudes das distorções econômicas, como seu trabalho indica. A publicação da pesquisa de Harberger de 1971 no *Journal of Economic Literature* coincide com um uso acelerado de métodos de triângulo por estudiosos além de Harberger para avaliar os efeitos de bem-estar de várias distorções. É claro que os *papers* de Harberger influenciaram pelo menos uma parte deste trabalho, em alguns casos fornecendo modelos analíticos, em outros simplesmente encorajando outros. Por exemplo, Browning (1975, p. 247) abre sua análise das distorções do mercado de trabalho induzidas pelo sistema de seguridade social observando: “o trabalho seminal de Arnold Harberger sobre a medição do custo do bem-estar da tributação fornece uma técnica que pode ser usada para estimar o custo do bem-estar das distorções nas escolhas de renda e lazer” - e o artigo de Browning então prossegue fazendo exatamente isso.

O impacto do trabalho de Harberger é difícil de medir com precisão, mas um indicador revelador são os números de artigos que contêm estimativas empíricas de triângulos de Harberger publicados nas principais revistas de economia. Para tanto, seja uma amostra de doze periódicos de economia de interesse geral entre 1964, quando Harberger (1964a) e (1964b) apareceu, e 1994, 30 anos depois e no 150º aniversário da publicação do clássico de Dupuit. Em 1964, exatamente um artigo nesses doze periódicos relata estimativas de triângulos Harberger. O número de artigos estimando o tamanho dos triângulos Harberger sobe para sete em 1974 e depois para doze em 1984, mas cai para quatro em 1994. Com certeza, esses números não refletem publicações em diários de campo, monografias e volumes editados e vários outros estabelecimentos profissionais. É também digno de nota que artigos teóricos sobre medição de perda de peso morto (dos quais havia muitos) não são contados nesta tabulação. Nem é uma simples contagem de artigos publicados um indicador convincente de influência intelectual. Mas os números revelam que a medição empírica da perda de peso morto foi muito mais amplamente praticada depois de 1964. Embora os números decrescentes entre 1984 e 1994 possam refletir a novidade cada vez menor da medição da perda de peso morto – já que os principais periódicos preferem publicar artigos apresentando novos métodos de análise – é no entanto verdade que quatro artigos publicados em 1994 constituem um crescimento considerável em relação ao esforço único de 1964.

5.9 Competição Imperfeita

Nesta seção, consideramos os efeitos da tributação em mercados imperfeitamente competitivos. A análise, na maior parte, é de equilíbrio parcial por natureza, e consideramos impostos *ad valorem* e impostos específicos sobre a produção[39]. Mercados imperfeitamente competitivos podem aparecer em uma ampla variedade de formas, e o analista tributário enfrenta a difícil tarefa de determinar qual modelo é apropriado em cada aplicação (consulte Tirole (1988) para uma excelente discussão de diferentes modelos). Em termos gerais, podemos primeiro classificar os modelos com base em se eles consideram produtos homogêneos ou heterogêneos[40]. Modelos com

[39] Diferentemente da concorrência perfeita, o impacto da incidência *ad valorem* e impostos específicos difere em mercados imperfeitamente competitivos.

[40] Dependendo do grau de diferenciação dos produtos, os oligopólios podem ser classificados em mercados de produtos homogêneos ou heterogêneos. Em oligopólio com produtos homogêneos as empresas competem com produtos que são virtualmente idênticos, com pouca ou nenhuma diferenciação em termos de qualidade, funcionalidade ou características. A competição se dá

diferentes empresas produzindo produtos idênticos incluem o oligopólio de Bertrand e o modelo de oligopólio de Cournot-Nash. Aqueles com bens heterogêneos incluem os modelos de concorrência monopolística (por exemplo, Dixit e Stiglitz (1977) e Spence (1976)), modelos de localização (por exemplo, Hotelling (1929) e Salop (1979)) e modelos de diferenciação vertical (por exemplo, Gabszewicz e Thisse (1979) e Shaked e Sutton (1982)). Quer os produtos sejam homogêneos ou heterogêneos, descobriremos que o impacto dos impostos sobre os preços funciona por meio de canais diretos e indiretos (com os canais indiretos diferentes entre os modelos).

5.9.1 Oligopólios

A concorrência de Bertrand é um conceito de equilíbrio de Nash no qual as empresas competem em preços. O equilíbrio de preços é bastante simples: as empresas competem baixando os preços até que todas as empresas definam o preço igual ao seu custo marginal comum. Nenhuma empresa obtém lucros econômicos, não deixando incentivo para entrada ou saída. Os efeitos de um imposto unitário na produção nesse modelo são diretos. Como o preço ao produtor não pode cair abaixo do custo marginal, todo o imposto é repassado aos consumidores. Um mercado baseado no modelo de Bertrand produz o mesmo resultado de equilíbrio que um mercado perfeitamente competitivo com oferta agregada perfeitamente elástica.

Agora, considere o modelo de oligopólio de Cournot-Nash, no qual firmas idênticas competem escolhendo níveis de produção condicionais às expectativas dos níveis

principalmente pelo preço. Exemplos incluem a indústria de aço, o mercado de petróleo (embora haja variações nas fontes de petróleo e nas operações de refino, o petróleo bruto em si é praticamente o mesmo, independentemente do produtor), a indústria de cimento. Em mercados de oligopólio com produtos heterogêneos, as empresas competem não só em termos de preço, mas também por meio de diferenciação de produto, como marca, design, qualidade e funcionalidades. Exemplos incluem a indústria automobilística (embora todas ofereçam automóveis, cada fabricante busca diferenciar seus produtos por meio de design, tecnologia, eficiência de combustível, e percepção da marca), telecomunicações (embora todos ofereçam essencialmente o mesmo serviço básico, diferenciam-se através de pacotes de dados, qualidade de serviço, cobertura, e ofertas promocionais) e a indústria de bebidas (todas ofereçam refrigerantes, cada marca diferencia seus produtos através de sabores, *branding*, e campanhas de marketing).

de produção de seus concorrentes. Prosseguimos em duas etapas: primeiro fixando o número de empresas no mercado em N e depois permitindo a entrada livre sem custo. Para simplificar, assumiremos que as empresas são idênticas e que o equilíbrio é simétrico.

Considere a empresa i no mercado. Sua função de lucro é dada por

$$\max_{q_i} \pi_i(q_i) = (1-\tau_v)p(q_i + Q_{-i})q_i - c(q_i) - \tau_s q_i \tag{5.160}$$

em que q_i é a produção da i-ésima empresa, Q_{-i} é a produção de todas as outras empresas do mercado e $p(Q)$ é a função de demanda inversa para a demanda do mercado Q. A função de custo é $c(q_i)$ e τ_v e τ_s são os impostos *ad valorem* e específicos sobre a empresa.

A condição de primeira ordem para a i-ésima empresa é dada por

$$\frac{\partial \pi_i(q_i)}{\partial q_i} = 0 \quad \iff \quad (1-\tau_v)\left[p'(q_i + Q_{-i})q_i + p(q_i + Q_{-i})\right] - c(q_i)' - \tau_s = 0 \tag{5.161}$$

As condições de segunda ordem são

$$\frac{\partial^2 \pi_i(q_i)}{\partial q_i^2} = (1-\tau_v)p''(q_i)q_i + 2(1-\tau_v)p'(q_i + Q_{-i}) - c''(q_i) < 0 \tag{5.162}$$

Focamos na condição de segunda ordem porque queremos saber quais são as restrições que garantem que o resultado da condição de primeira ordem seja efetivamente a maximização do lucro.

Podemos reescrever a condição de segunda ordem como

$$(1-\tau_v)\left[p''(q_i)q_i + 2p'(q_i + Q_{-i})\right] - c''(q_i) < 0$$
$$(1-\tau_v)\left[p''(q_i)q_i \frac{N}{N}\frac{p'(q_i + Q_{-i})}{p'(q_i + Q_{-i})} + 2p'(q_i + Q_{-i})\right] - c''(q_i) < 0$$

$$
\begin{aligned}
(1-\tau_v)\left[p''(q_i)\frac{Q}{N}\frac{p'(q_i+Q_{-i})}{p'(q_i+Q_{-i})}+2p'(q_i+Q_{-i})\right]-c''(q_i)&<0\\
(1-\tau_v)\left[\frac{\eta}{N}p'(q_i+Q_{-i})+2p'(q_i+Q_{-i})\right]-c''(q_i)&<0\\
\frac{\eta}{N}\tilde{p}'(q_i+Q_{-i})+2\tilde{p}'(q_i+Q_{-i})-c''(q_i)&<0\\
\frac{\eta}{N}\tilde{p}'(q_i+Q_{-i})+\tilde{p}'(q_i+Q_{-i})+\tilde{p}'(q_i+Q_{-i})-c''(q_i)&<0\\
\tilde{p}'(q_i+Q_{-i})\left(1+\frac{\eta}{N}\right)+(\tilde{p}'(q_i+Q_{-i})-c''(q_i))\frac{\tilde{p}'(q_i+Q_{-i})}{\tilde{p}'(q_i+Q_{-i})}&<0\\
\tilde{p}'(q_i+Q_{-i})\left(1+\frac{\eta}{N}\right)+\left(1-\frac{c''(q_i)}{\tilde{p}'(q_i+Q_{-i})}\right)\tilde{p}'(q_i+Q_{-i})&<0\\
\tilde{p}'(q_i+Q_{-i})\left(1+\frac{\eta}{N}\right)+k\tilde{p}'(q_i+Q_{-i})&<0\\
\tilde{p}'(q_i+Q_{-i})\left(1+\frac{\eta}{N}+k\right)&<0\\
\frac{\tilde{p}'(q_i+Q_{-i})}{N}(N+\eta+Nk)&<0 \qquad (5.163)
\end{aligned}
$$

em que $\tilde{p}=(1-\tau_v)p$ é o preço do produtor, $\eta=Q\dfrac{p''}{p'}$ é a elasticidade da inclinação da função de demanda inversa e $k=1-\dfrac{c''}{\tilde{p}'}$ mensura as inclinações relativas das curvas de demanda e de custo marginal. Dado que $p'<0$, as condições de segunda ordem requerem que $N+\eta+Nk>0$.

Em um equilíbrio simétrico precisamos resolver apenas p e q usando as duas equações abaixo, a partir da condição de primeira ordem:

$$p=p(Nq) \qquad (5.164)$$

$$(1-\tau_v)p'(Nq)q+(1-\tau_v)p(Nq)-c'(q)=\tau_s \qquad (5.165)$$

Diferenciando (5.165) com relação a τ_s, por meio do teorema da função implícita, obtemos:

$$\frac{dq}{d\tau_s}=-\frac{-1}{(1-\tau_v)p'+(1-\tau_v)qNp''+(1-\tau_v)Np'-c''}$$

$$\begin{aligned}
&= \frac{1}{(1-\tau_v)p'(1+N)+(1-\tau_v)Qp''-c''} \\
&= \frac{1}{\tilde{p}'(1+N)+(1-\tau_v)Qp''\dfrac{p'}{p'}-c''} \\
&= \frac{1}{\tilde{p}'(1+N)+\tilde{p}'\eta-c''} \\
&= \frac{1}{\tilde{p}'(\eta+N)+\tilde{p}'-c''} \\
&= \frac{1}{\tilde{p}'(\eta+N)+\tilde{p}'\left(1-\dfrac{c''}{\tilde{p}'}\right)} \\
&= \frac{1}{\tilde{p}'(\eta+N)+\tilde{p}'k} \\
&= \frac{1}{\tilde{p}'(\eta+N+k)} \qquad (5.166)
\end{aligned}$$

Segue imediatamente que

$$\frac{dQ}{d\tau_s} = \frac{N}{\tilde{p}'(\eta+N+k)} \qquad (5.167)$$

Usando o fato de que $\tilde{p} = (1-\tau_v)p$ podemos computar

$$\frac{d\tilde{p}}{d\tau_s} = (1-\tau_v)p'\left(\frac{dQ}{d\tau_s}\right) = \frac{N}{\eta+N+k} \qquad (5.168)$$

Se as condições de segunda ordem se mantiverem, então $\eta + N + k > 0$. E, se $\eta + N + k > 0$, a produção cai e o imposto é (até certo ponto) repassado aos consumidores. O grau de deslocamento para frente da taxa unitária sobre o produto depende da inclinação da elasticidade da função de demanda inversa (η), do número de empresas (N) e das inclinações relativas das funções de custo marginal e demanda inversa (k).

O *overshifting* ocorre quando o preço do produtor sobe mais do que o imposto. Como mostramos na seção anterior, esse resultado é impossível em mercados perfeitamente competitivos. Uma vez que mercados imperfeitamente competitivos são permitidos, o *overshifting* se torna uma possibilidade e pode ser garantido em algumas especificações de modelo. O *overshifting* pode ocorrer devido à existência de poder de mercado e comportamento estratégico entre as empresas. As empresas reconhecem que o deslocamento futuro do imposto diminuirá a demanda por seus produtos. Assim, em algumas circunstâncias, eles desejam aumentar o preço mais do que o aumento do imposto para compensar a perda de receita decorrente da diminuição da demanda.

Por definição, a transferência dos impostos para o consumidor ocorre de forma excessiva se a derivada em (5.168) for maior que 1, o que significa que $\eta + k < 0$. Se os custos são lineares na produção, então $c'' = 0$ e $k = 1$. Portanto, uma condição necessária e suficiente para a transferência dos impostos para o consumidor ocorrer de forma excessiva é que $\eta < -1$. Considere uma função de demanda de elasticidade constante com elasticidade de demanda $\varepsilon < 0$. Nesse caso, $\eta = \dfrac{1-\varepsilon}{\varepsilon} < -1$ para todos $\varepsilon < 0$. A transferência dos impostos para o consumidor ocorrerá de forma excessiva sempre e aumentará à medida que a demanda se tornar menos elástica (quando η aumenta em valor absoluto).

Por sua vez, os preços ao produtor aumentam com o aumento de um imposto *ad valorem* da seguinte forma:

$$\frac{d\tilde{p}}{d\tau_v} = \frac{Np\left(1 + \dfrac{1}{\varepsilon}\right)}{\eta + N + k} \tag{5.169}$$

em que $\varepsilon < 0$ é a elasticidade-preço da demanda. A transferência de um imposto *ad valorem* para o consumidor ocorre de forma excessiva quando a variação percentual no preço do produtor excede 100% e ocorre nesse modelo quando $-N < \eta + k < \dfrac{N}{\varepsilon}$. Observe que o tamanho do mercado (N) não afeta a condição de deslocamento para impostos especiais de consumo, mas afeta o grau do deslocamento. Suponha que $-N < \eta + k < 0$ (de modo que $\dfrac{d\tilde{p}}{d\tau_s} > 1$). Então $\dfrac{d\tilde{p}}{d\tau_s dN} = \dfrac{\eta}{(\eta + N + k)^2} < 0$. Em outras palavras, para determinados valores de η e k, o deslocamento excessivo é maximizado para um monopolista e desaparece quando N se aproxima do infinito.

5.9.2 Produtos Diferenciados

Os modelos de oligopólio discutidos acima sofrem com as premissas restritivas de que os bens são idênticos e que nenhuma distinção pode ser feita entre diferentes marcas. Em alguns mercados (por exemplo, mercados de commodities agrícolas), isso pode ser uma suposição razoável. Na maioria dos outros mercados, no entanto, os produtores se esforçam bastante para diferenciar seus produtos. A diferenciação do produto cria algum poder de monopólio, e os resultados no modelo de oligopólio de N fixo indicam que a capacidade de repassar impostos depende muito do número de concorrentes no mercado. Nesta seção, consideramos alguns modelos de produtos diferenciados e examinamos a relação entre concorrência de produtos e incidência tributária.

Começamos com o modelo Dixit e Stiglitz (1977) e seu modelo de competição monopolística. Esse é um modelo um tanto especial, pois cada produto concorre com todos os outros produtos, e o principal objetivo do modelo é ilustrar os benefícios da variedade de produtos. É útil começar com este modelo, pois destaca a importância da diferenciação do produto – um recurso deixado de fora do modelo de oligopólio de bens homogêneos. Considere o seguinte modelo simplificado de variedade de produtos Dixit-Stiglitz, baseado em Krugman (1980).

Os consumidores são idênticos e maximizam uma função de utilidade

$$\max_{x_i} U(x_1, \dots, x_N) = \frac{1}{\theta}\sum_{i=1}^{N} x_i^{\theta}, \quad 0 < \theta < 1$$
$$\text{sujeito a } \sum_{i=1}^{N} p_i x_i = M \tag{5.170}$$

Os bens de consumo entram na função de utilidade simetricamente, mas não são substitutos perfeitos (a menos que θ seja igual a 1).

Montando o lagrangeano, obtemos:

$$L(x_i, \lambda) = \frac{1}{\theta} \sum_{i=1}^{N} x_i^{\theta} - \lambda \left(\sum_{i=1}^{N} p_i x_i - M \right) \tag{5.171}$$

As condições de primeira ordem são:

$$\frac{\partial L}{\partial x_i} = 0 \quad \Longleftrightarrow \quad x_i^{\theta-1} - \lambda p_i = 0, \quad i = 1, \dots, N \tag{5.172}$$

$$\frac{\partial L}{\partial \lambda} = 0 \quad \Longleftrightarrow \quad \sum_{i=1}^{N} p_i x_i = M \tag{5.173}$$

Das condições de primeira ordem podemos derivar as funções de demanda:

$$x_i = (\lambda p_i)^{-\varepsilon}, \quad \varepsilon = \frac{1}{1-\theta} > 1 \tag{5.174}$$

em que λ é a utilidade marginal privada da renda. Se N for grande, podemos assumir que as decisões de preço de uma empresa individual terão um efeito desprezível em λ e a demanda poderá ser escrita como:

$$x_i = A p_i^{-\varepsilon} \tag{5.175}$$

As empresas maximizam os lucros e assumimos que os custos são lineares na forma $cx_i + F$, em que c é o custo marginal e F é o custo fixo. Permitindo que $\tilde{p} \equiv (1 - \tau_v)p$ seja o preço do produtor com um imposto *ad valorem* (τ_v), a regra de preços da empresa é dada pela regra de preços padrão do monopolista:

$$\tilde{p} = \left(\frac{\varepsilon}{\varepsilon - 1} \right) (c + \tau_s) \tag{5.176}$$

Para um imposto especial de consumo ou *ad valorem* aplicado apenas a um determinado setor podemos diferenciar a equação acima. Portanto,

$$\frac{d\tilde{p}}{d\tau_s} = \frac{\varepsilon}{\varepsilon - 1} > 1 \tag{5.177}$$

$$\frac{d\tilde{p}}{d\tau_v} = 0 \tag{5.178}$$

Um imposto especial de consumo é mais do que 100% desviado para a frente (elasticidade constante e resultado de custo linear), enquanto um imposto *ad valorem* não tem impacto no preço do produtor, mas é totalmente repassado aos consumidores.

Uma desvantagem do modelo Dixit-Stiglitz é que todos os produtos são tratados como concorrentes iguais aos de outros produtos. Uma rápida olhada em qualquer número de mercados indica que essa suposição é insustentável. Em seguida, voltamos para um modelo de competição espacial em que as empresas se localizam no espaço do produto para capturar o máximo de clientes em um jogo de entrada simultânea. Utilizamos o modelo de círculos de Salop (1979), desenvolvido para analisar impostos *ad valorem* e impostos especiais de consumo por Kay e Keen (1983). A virtude do modelo de círculo é que ele permite explicitamente modelar o número de firmas em equilíbrio (diferentemente do modelo linear de Hotelling (1929)). Seguindo Salop, assumimos que N empresas idênticas decidem simultaneamente se entram em um mercado em que os consumidores estão localizados uniformemente ao redor do círculo e em que cada consumidor deseja comprar 1 unidade do produto. As empresas que entram localizam-se equidistantemente em torno de um círculo unitário. Assim, em equilíbrio, cada empresa enfrentará uma demanda de $1/N$ (assumindo que o mercado esteja coberto). Cada indivíduo comprará no máximo uma unidade do bem e cada um prefere comprar o bem de qualidade ou local x o mais próximo possível da sua qualidade preferida (x^*). Especificamente, o custo do bem para o consumidor é o preço de compra (p_i) mais um "custo de transporte" que se supõe ser uma constante t vezes a distância de sua localização $|x - x^*|$. A utilidade para um consumidor que compra uma unidade de x é igual a

$$U = s - p - t|x - x^*| \tag{5.179}$$

em que s é uma constante arbitrária suficientemente grande para garantir que $U > 0$.

Considere um consumidor localizado em $\hat{x}$, entre 0 e $\frac{1}{N}$. Esse consumidor será indiferente entre a compra da firma i e da firma $i+1$ se

$$p_i + t\hat{x} = p + t\left(\frac{1}{N} - \hat{x}\right) \tag{5.180}$$

em que p_i é o preço cobrado pela i-ésima empresa e p é o preço cobrado por outras empresas. Por esse preço p_i (tornando o consumidor em $\hat{x}$ indiferente), a demanda pelo i-ésimo bem da empresa, $D(p_i, p)$, será igual a $2\hat{x}$. Resolvendo a equação acima para $\hat{x}$, obtemos:

$$D(p_i, p) = \frac{p - p_i}{t} + \frac{1}{N} \tag{5.181}$$

A empresa maximiza os lucros escolhendo o preço. Ela enfrenta uma taxa de imposto *ad valorem* τ_v e uma taxa de imposto unitária τ_s. Os lucros são dados por

$$\max_{p_i} \pi_i = [(1 - \tau_v)p_i - c - \tau_s]\left(\frac{p - p_i}{t} + \frac{1}{N}\right) - F \tag{5.182}$$

em que c é o custo marginal e F é o custo fixo.

Derivando em relação ao preço e assumindo que em equilíbrio $p = p_i$, obtemos:

$$\frac{\partial \pi_i}{\partial p_i} = 0 \quad \Longleftrightarrow \quad (1 - \tau_v)\left(\frac{p - p_i}{t} + \frac{1}{N}\right) + [(1 - \tau_v)p_i - c - \tau_s]\left(-\frac{1}{t}\right) = 0 \tag{5.183}$$

Manipulando a expressão acima chegamos a:

$$\begin{aligned}(1-\tau_v)\left(\frac{1}{N}\right)+[(1-\tau_v)p_i-c-\tau_s]\left(-\frac{1}{t}\right)&=0\\(1-\tau_v)\left(\frac{t}{N}\right)&=(1-\tau_v)p_i-c-\tau_s\\p_i&=\frac{t}{N}+\frac{c+\tau_s}{1-\tau_v}\end{aligned}\tag{5.184}$$

Podemos reescrever o problema acima em termos do preço do produtor $\tilde{p}$:

$$\tilde{p}=\frac{(1-\tau_v)t}{N}+c+\tau_s\tag{5.185}$$

Assim, os impostos unitários são totalmente repassados no sentido em que o preço do produtor sobe pelo valor total do imposto. A rigor, essa afirmação só é verdadeira se o número de firmas de equilíbrio não for afetado por alterações no imposto especial de consumo.

Precisamos de uma segunda equação para determinar o número de firmas de equilíbrio. Uma condição de lucro zero para a empresa marginal faz isso. Em equilíbrio, cada empresa cobre $1/N$ do mercado. Assim,

$$((1-\tau_v)p_i-c-\tau_s)\left(\frac{1}{N}\right)=F\tag{5.186}$$

Dessa expressão decorre que:

$$\left((1-\tau_v)\left(\frac{t}{N}+\frac{c+\tau_s}{1-\tau_v}\right)-c-\tau_s\right)\left(\frac{1}{N}\right)=F\tag{5.187}$$

E isso implica que

$$N=\sqrt{\frac{(1-\tau_v)t}{F}}\tag{5.188}$$

Embora uma alteração no imposto especial de consumo não afete o número de equilíbrio das empresas, uma alteração no imposto *ad valorem* afeta. Note que $\frac{dN}{d\tau_v} = -\frac{1}{2}\sqrt{\frac{t}{(1-\tau_v)F}} < 0$.

A incidência tributária *ad valorem* pode ser decomposta em dois componentes: um efeito direto e um efeito indireto através da mudança no número de equilíbrio das firmas. Fixando N,

$$\left.\frac{d\tilde{p}}{d\tau_v}\right|_N \equiv \left.\frac{\partial(1-\tau_v)p}{\partial\tau_v}\right|_N = -\frac{t}{N} \tag{5.189}$$

A incidência completa do imposto é

$$\begin{aligned}
\tilde{p} &= \frac{(1-\tau_v)t}{N} + c + \tau_s \\
\tilde{p} &= \frac{(1-\tau_v)t}{\sqrt{\frac{(1-\tau_v)t}{F}}} + c + \tau_s \\
\tilde{p} &= \sqrt{(1-\tau_v)tF} + c + \tau_s \quad \Longrightarrow \quad \frac{d\tilde{p}}{d\tau_v} = -\frac{1}{2}\sqrt{\frac{tF}{(1-\tau_v)}} = -\frac{1}{2}\frac{t}{N}
\end{aligned} \tag{5.190}$$

isto é, exatamente metade da incidência em (5.189) quando N é fixo. Em outras palavras, a saída de firmas reduz pela metade o ônus para os produtores (e aumenta o ônus para os consumidores). As empresas saem porque um aumento na tributação *ad valorem* é equivalente (do ponto de vista da empresa) a um aumento no custo fixo em relação à receita.

6 Tributação Ótima

In duscission of economic science, 'Chicago' stands for an approach that takes seriously the use of economic theory as a tool for analyzing a startlingly wide range of concrete problems, rather than as an abstract mathematical structure of great beauty but little power; for an approach that insists on the empirical testing of theoretical generalizations and that rejects alike fact without theory and theory without facts.

Milton Friedman, 1974

6.1 Introdução

O que torna o estudo do problema de tributação ótima tão interessante e, ao mesmo tempo, complexo é o fato de que se trata modelo de principal e agente. O governo (o principal) quer escolher uma alocação ótima para maximizar o bem-estar social. No entanto, na busca de seus objetivos está sujeito não somente a uma restrição de recursos mas também está limitado pelas ações do contribuinte (o agente), que tem liberdade para escolher aquilo que achar melhor para si.

O objetivo é maximizar o bem-estar social (minimizar o peso morto) sujeito a restrição de receita tributária.

1. Ótimo (*first-best*)

    - Suponha que tenhamos informação perfeita, mercados completos, concorrência perfeita, impostos *lump-sum* sem custo.
    - Resultado: o segundo teorema de bem-estar implica que qualquer alocação eficiente de Pareto pode ser alcançada como um equilíbrio competitivo com transferências *lump-sum* apropriadas.
    - O problema da política econômica se reduz ao cálculo dos impostos de montante fixo (*lump-sum*) necessários para alcançar o equilíbrio desejado.
    - O *trade-off* entre equidade e eficiência desaparece.

2. Problemas:

    - Não há como fazer as pessoas revelarem suas características sem nenhum custo: para evitar pagar uma quantia alta, uma pessoa qualificada fingiria ser inexperiente.
    - Assim, o governo precisa estabelecer impostos como função dos resultados econômicos: renda, propriedade, consumo de bens $\Rightarrow$ Distorção e peso morto

- Então, acabamos com o segundo melhor mundo com tributação ineficiente
- A taxação ineficiente gera custos de eficiência.
- Aqui nós discutimos o melhor imposto sobre mercadorias

3. Quatro principais resultados qualitativos na teoria tributária ótima:

   (a) Regra da elasticidade inversa de Ramsey

   (b) Diamond e Mirrlees: eficiência da produção

   (c) Atkinson e Stiglitz: nenhuma tributação do consumo com tributação de renda não linear

   (d) Chamley/Judd: nenhuma taxação de capital em modelos de horizonte infinito

6.2 O Problema de Taxação em Ramsey

A primeira vez que o problema de tributação ótima foi proposto e formalmente resolvido foi em 1927 em um trabalho visionário de Frank Ramsey A versão que apresentaremos aqui segue a abordagem dual, que se deve principalmente às contribuições de Diamond e Mirrlees, quase meio século depois da contribuição original de Ramsey. Em seu trabalho original, Ramsey usou uma abordagem primal, que torna o problema bem menos tratável, ainda que se tenha provado bastante útil na discussão da tributação intertemporal. No problema de Ramsey de acordo com a abordagem dual de Diamond e Mirrlees o governo fixa impostos sobre a renda de modo a aumentar uma determinada quantia da receita E e minimizar a perda da utilidade.

Um indivíduo (agente representativo), sem preocupações redistributivas, tem as seguintes características:

- Função de utilidade: $u(x_1, \dots, x_N)$
- Z é a renda não-trabalho

- $p_i' = p_i + \tau_i$ é o preço do bem, incluso o imposto
- Restrição orçamentária: $p_1'x_1 + ... + p_N'x_N \leq Z$

O lagrangeano desse problema é dado por:

$$L(x_1, ..., x_N, \alpha) = u(x_1, ..., x_N) + \alpha \left[Z - (p_1'x_1 + ... + p_N'x_N)\right] \tag{6.1}$$

As condições de primeira ordem implicam que $u_{x_i} = \alpha p_i'$. Sejam as demandas marshallianas, $x_i(p', Z)$, e a função de utilidade indireta $V(p', Z)$, em que $\alpha = \dfrac{\partial V}{\partial Z}$ é uma utilidade marginal da renda para o indivíduo. A identidade de Roy implica que $\dfrac{\partial V}{\partial p_i'} = -x_i \left(\dfrac{\partial V}{\partial Z}\right)$.

Duas maneiras equivalentes de configurar o problema de Ramsey pelo lado do governo:

1. O objetivo é maximizar a utilidade indireta do agente representativo sujeita a restrição de receita.

$$\begin{aligned} &\max V(p', Z) \\ \text{sujeito a} \quad &\tau \cdot x = \sum_{i=1}^{N} \tau_i x_i(p', Z) \geq E \end{aligned} \tag{6.2}$$

2. O objetivo é minimizar o custo de eficiência sujeito a restrição de receita.

$$\begin{aligned} &\min CE(p') = e(p', V(p', Z)) - e(p, V(p', Z)) - E \\ \text{sujeito a} \quad &\tau \cdot x = \sum_{i=1}^{N} \tau_i x_i(p', Z) \geq E \end{aligned} \tag{6.3}$$

Observe a equivalência com EV, não com CV.

Vamos analisar o problema pelo lado da função de utilidade indireta. Nós a maximizamos sujeito à restrição da receita tributária. Suponha que o governo aumenta τ_i por $d\tau_i$. Há mudanças na função objetivo do governo uma vez que há aumentos na receita tributária. Há mudanças na função objetivo do governo uma vez que há redução de bem-estar. O ótimo é caracterizado por efeitos de balanceamento de mudanças na receita tributária e mudanças no bem-estar privado.

O lagrangeano desse problema é portanto:

$$
\begin{aligned}
L(p_i', \lambda) &= V(p', Z) + \lambda \left[\sum_{i=1}^{N} \tau_i x_i(p', Z) - E \right] \\
&= V(p', Z) + \lambda \left[\sum_{i=1}^{N} (p_i' - p_i)\, x_i(p', Z) - E \right]
\end{aligned} \tag{6.4}
$$

A condição de primeira ordem é tal que

$$
\begin{aligned}
\frac{\partial L}{\partial p_i'} = 0 \quad \Longleftrightarrow \quad & \frac{\partial V(p', Z)}{\partial p_i'} + \lambda \left[x_i(p', Z) + \sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial p_i'} \right] = 0 \\
& -x_i(p', Z) \left(\frac{\partial V}{\partial Z} \right) + \lambda \left[x_i(p', Z) + \sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial p_i'} \right] = 0 \\
& (\lambda - \alpha)\, x_i(p', Z) + \lambda \sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial p_i'} = 0
\end{aligned} \tag{6.5}
$$

em que $\dfrac{\partial V(p', Z)}{\partial p_i'}$ é a perda de bem-estar individual, $x_i(p', Z)$ é o efeito mecânico e $\sum_j \tau_j \dfrac{\partial x_j(p', Z)}{\partial p_i'}$ são as respostas comportamentais.

As taxas de imposto ótimas satisfazem a fórmula de Ramsey:

$$
\sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial p_i'} = -\frac{x_i(p', Z)}{\lambda} (\lambda - \alpha) \tag{6.6}
$$

para $i = 1, \ldots, N$ define um sistema de N equações e N incógnitas. Note que um bem não deve ser taxado ($j < i$).

A Regra de Ramsey é frequentemente escrita em termos de elasticidades (compensadas) de Hicks para obter uma intuição adicional. Para fazer isso, comece definindo

$$\theta = \lambda - \alpha - \lambda \frac{\partial}{\partial Z}\left(\sum_j \tau_j x_j(p', Z)\right) \tag{6.7}$$

Observe que θ é independente de i, ou seja, é constante entre os bens. Mas o que significa esse termo?

- θ mede o valor para o governo de introduzir um imposto *lump-sum* no montante de R\$ 1.
- Digamos que o governo introduza um imposto *lump-sum* no montante de R\$ 1
  - O governo ganha diretamente λ.
  - Os indivíduos perdem bem-estar no montante de α.
  - O governo perde receita tributária devido ao efeito-renda dx_j, $\frac{\partial}{\partial Z}\left(\sum_j \tau_j x_j(p', Z)\right)$.

Usando a equação de Slutsky sabemos que:

$$\frac{\partial x_j(p', Z)}{\partial p'_i} = \frac{\partial h_j(p', u)}{\partial p'_i} - x_i(p', Z)\frac{\partial x_j(p', Z)}{\partial Z} \tag{6.8}$$

Substituindo este resultado na regra ótima de Ramsey e usando θ, temos:

$$\begin{aligned}
\sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial p'_i} &= -\frac{x_i(p', Z)}{\lambda}(\lambda - \alpha) \\
\sum_j \tau_j \left[\frac{\partial h_j(p', u)}{\partial p'_i} - x_i(p', Z)\frac{\partial x_j(p', Z)}{\partial Z}\right] &= -\frac{x_i(p', Z)}{\lambda}(\lambda - \alpha) \\
\sum_j \tau_j \frac{\partial h_j(p', u)}{\partial p'_i} - x_i(p', Z)\sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial Z} &= -\frac{x_i(p', Z)}{\lambda}(\lambda - \alpha) \\
-\frac{1}{x_i(p', Z)}\sum_j \tau_j \frac{\partial h_j(p', u)}{\partial p'_i} + \sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial Z} &= \frac{1}{\lambda}(\lambda - \alpha) \\
-\frac{1}{x_i(p', Z)}\sum_j \tau_j \frac{\partial h_j(p', u)}{\partial p'_i} &= \frac{1}{\lambda}(\lambda - \alpha) - \sum_j \tau_j \frac{\partial x_j(p', Z)}{\partial Z}
\end{aligned}$$

$$
\begin{aligned}
-\frac{1}{x_i(p',Z)}\sum_j \tau_j \frac{\partial h_j(p',u)}{\partial p_i'} &= \frac{\lambda - \alpha - \lambda \sum_j \tau_j \frac{\partial x_j(p',Z)}{\partial Z}}{\lambda} \\
-\frac{1}{x_i(p',Z)}\sum_j \tau_j \frac{\partial h_j(p',u)}{\partial p_i'} &= \frac{\theta}{\lambda} \\
\frac{1}{x_i(p',Z)}\sum_j \tau_j \frac{\partial h_j(p',u)}{\partial p_i'} &= -\frac{\theta}{\lambda} \qquad (6.9)
\end{aligned}
$$

No ótimo $\theta > 0$, o que implica que a soma das elasticidades-preço ponderadas pelas taxas de impostos são constantes entre os bens. Ou seja, o imposto τ_j sobre o bem j reduz o consumo do bem i (mantendo a utilidade constante) por aproximadamente $dh_i = \tau_j \frac{\partial h_i(p',u)}{\partial p_j'}$. Portanto, a redução total no consumo do bem i devido a todos os impostos é $\sum_j \tau_j \frac{\partial h_j(p',u)}{\partial q_i}$. Ao dividir por x_i, obtém-se a redução percentual no consumo do bem i devido ao sistema tributário (a receita foi normalizada). A quantia $\frac{1}{x_i}\sum_j \tau_j \frac{\partial h_j(p',u)}{\partial p_i'}$ é chamada índice de desencorajamento. A fórmula tributária de Ramsey diz que os índices de desencorajamento devem ser iguais entre todos os bens no nível ótimo.

O problema acima pode ser reescrito a partir das elasticidades compensadas (das demandas hicksianas), como segue:

$$
\sum_j \frac{\tau_j}{1+\tau_j}\varepsilon_{ij}^c = \frac{\theta}{\lambda} \qquad (6.10)
$$

Daqui depreendem-se que se $\varepsilon_{ij} = 0$ para $i \neq j$, ou seja, se as elasticidades-preço cruzadas forem zero, obtemos a famosa regra do inverso da elasticidade, isto é,

$$
\frac{\tau_i}{1+\tau_i} = \frac{\theta}{\lambda}\frac{1}{\varepsilon_{ii}} \qquad (6.11)
$$

Isso implica que quanto maior for a elasticidade-preço da demanda, menor deve ser o imposto ótimo. É o mesmo fato observado no modelo de equilíbrio parcial, em que o peso morto era uma função crescente dos impostos.

A conclusão da proposta de Ramsey é tributar produtos inelásticos para minimizar os custos de eficiência. Mas não leva em conta existem motivos redistributivos. Presumivelmente, as necessidades são mais inelásticas do que os bens de luxo. Portanto, o sistema tributário ideal de Ramsey provavelmente será regressivo. Diamond (1975) estende o modelo de Ramsey para levar em consideração os motivos redistributivos.

As principais limitações da regra de Ramsey é

1. De acordo essa fórmula, o governo deveria tributar bens inelásticos para minimizar os custos de eficiência.

2. Não leva em conta questões redistributivas.

3. O principal problema é que as necessidades básicas tendem a ser mais inelásticas do que os bens de luxo.

4. O sistema tributário ideal de Ramsey provavelmente será regressivo.

5. Diamond (1975) estende o modelo de Ramsey para incorporar aspectos distribuições.

6.3 Produção Eficiente: O Modelo Diamond & Mirrless

O modelo de tributação Diamond-Mirrlees é uma estrutura fundamental na economia do bem-estar, desenvolvida por Peter A. Diamond e James A. Mirrlees nos anos 1970. Esse modelo explora a tributação ótima em uma economia, com foco na maximização do bem-estar social enquanto se minimiza o custo econômico da tributação. A abordagem principal do modelo é que, sob certas condições, é possível implementar uma estrutura de impostos que não distorça as decisões de produção, levando a uma alocação eficiente dos recursos.

Os elementos do modelo Diamond-Mirrlees são:

- Uma economia competitiva com múltiplos consumidores e produtores.
- A produção é realizada por firmas que utilizam insumos (como trabalho e capital) para produzir bens.
- Os consumidores têm preferências sobre bens de consumo e oferecem trabalho às firmas.
- O governo tem a capacidade de taxar os consumidores e as firmas para financiar suas atividades.

O governo busca maximizar uma função de bem-estar social, que depende do bem-estar dos consumidores. O governo utiliza instrumentos de política fiscal (impostos) para redistribuir renda e financiar bens públicos.

O Teorema da Não Distorção na Produção diz que a alocação eficiente dos recursos na produção requer que as decisões de produção não sejam distorcidas pelos impostos. Isto significa que os impostos devem ser projetados de tal forma que as firmas escolham os níveis de produção de maneira eficiente, ou seja, como fariam em um mercado competitivo sem impostos.

O modelo Diamond-Mirrlees sugere que é possível separar a eficiência na produção das questões de redistribuição. Os impostos devem ser aplicados de forma a não distorcer a produção, enquanto o objetivo de redistribuição pode ser alcançado por meio de impostos sobre os consumidores (por exemplo, impostos sobre o consumo ou a renda), sem afetar a eficiência produtiva.

O modelo considera tanto impostos lineares (onde a taxa de imposto é constante) quanto impostos não lineares (onde a taxa de imposto varia com o nível de renda ou consumo). A estrutura ótima de tributação muitas vezes depende de uma combinação desses tipos de impostos. O governo deve escolher impostos que maximizem o bem-estar social, levando em conta tanto a eficiência econômica quanto a equidade na distribuição de renda.

Entretanto, o governo pode não ter informações completas sobre as habilidades dos consumidores ou as tecnologias das firmas. Esse problema de informação assimétrica impõe restrições à forma como os impostos podem ser projetados. Os impostos devem ser projetados de modo a respeitar as restrições de incentivo, ou seja, devem garantir que os indivíduos e as firmas ajam de acordo com seus interesses, mesmo na presença de impostos.

Os principais resultados são:

1. Teorema da Produção Eficiente: a eficiência na produção pode ser mantida mesmo com a tributação, desde que os impostos não distorçam as decisões de produção. Isso significa, por exemplo, que impostos sobre bens intermediários devem ser evitados, enquanto impostos sobre bens finais podem ser usados para redistribuição.

2. Equidade e Eficiência: o modelo mostra que é possível atingir uma distribuição mais equitativa de renda sem sacrificar a eficiência produtiva, separando as decisões de tributação em dois níveis: um para garantir a eficiência e outro para redistribuição.

3. Tributação Ótima: a estrutura de tributação ótima depende de uma análise cuidadosa das preferências dos consumidores, das tecnologias de produção e das restrições informacionais e de incentivo enfrentadas pelo governo.

O modelo Diamond-Mirrlees tem implicações importantes para a formulação de políticas tributárias. Ele sugere que o governo deve focar em tributar os resultados finais (como consumo ou renda) em vez de interferir nas decisões de produção. Além disso, o modelo indica que uma política tributária eficiente pode ser implementada mesmo em um contexto de redistribuição, desde que se respeitem as restrições de eficiência.

Essa estrutura tem sido amplamente influente no debate sobre a tributação ótima e continua a ser uma referência teórica importante na economia pública.

Consideremos uma economia com N consumidores e M firmas. O governo pretende maximizar o bem-estar social, que é uma função das utilidades dos consumidores, enquanto arrecada impostos para financiar despesas públicas.

A função de bem-estar social é dada por:

$$W = W(U_1(x_1, l_1), U_2(x_2, l_2), \ldots, U_N(x_N, l_N)) \tag{6.12}$$

em que $U_i(x_i, l_i)$ é a função utilidade do consumidor i, x_i é o vetor de bens consumidos e l_i é o trabalho oferecido.

Cada firma j produz y_j unidades de um bem usando insumos z_j de acordo com uma função de produção $y_j = f_j(z_j)$.

O consumidor i enfrenta a seguinte restrição orçamentária:

$$p \cdot x_i \leq w_i l_i + \pi_i \tag{6.13}$$

em que p é o vetor de preços dos bens, w_i é o salário e π_i são os lucros recebidos da firma.

O governo arrecada impostos T e gasta G. A restrição orçamentária do governo é

$$T = G \tag{6.14}$$

O problema de maximização do bem-estar social é formulado como:

$$\max_{x_i, l_i, z_j} W(U_1(x_1, l_1), U_2(x_2, l_2), \ldots, U_N(x_N, l_N)) \tag{6.15}$$

Sujeito às seguintes restrições:

1. Equilíbrio de Mercado de Bens

$$\sum_{j=1}^{M} y_j = \sum_{i=1}^{N} x_i \tag{6.16}$$

Isto significa que a produção total de cada bem deve igualar a demanda total.

2. Equilíbrio de Mercado de Trabalho

$$\sum_{j=1}^{M} z_j = \sum_{i=1}^{N} l_i \tag{6.17}$$

A oferta total de trabalho deve igualar a demanda total.

3. Restrição Orçamentária do Governo

$$T = \sum_{i=1}^{N} t_i(x_i, l_i) \tag{6.18}$$

em que $t_i(x_i, l_i)$ representa a função de impostos do consumidor i.

O lagrangeano para o problema de maximização do bem-estar social, incorporando as restrições, é dado por:

$$\mathcal{L} = W(U_1(x_1, l_1), \ldots, U_N(x_N, l_N)) + \sum_{i=1}^{N} \lambda_i \left(w_i l_i + \pi_i - p \cdot x_i\right) + \mu \left(T - G\right) \quad (6.19)$$

Aqui, λ_i são os multiplicadores de Lagrange associados às restrições orçamentárias dos consumidores, e μ é o multiplicador associado à restrição orçamentária do governo.

Para encontrar as condições de primeira ordem (FOCs), derivamos o lagrangeano em relação a x_i, l_i, e z_j.

A derivada do lagrangeano em relação a x_i é:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial W}{\partial U_i} \cdot \frac{\partial U_i}{\partial x_i} - \lambda_i p \quad (6.20)$$

Para maximizar $\mathcal{L}$, a condição de primeira ordem é:

$$\frac{\partial W}{\partial U_i} \cdot \frac{\partial U_i}{\partial x_i} = \lambda_i p \quad (6.21)$$

Essa equação significa que a utilidade marginal do consumo, ponderada pelo peso de bem-estar social $\frac{\partial W}{\partial U_i}$, deve ser igual ao custo marginal do consumo, que é o preço p.

A derivada do lagrangeano em relação a l_i é:

$$\frac{\partial \mathcal{L}}{\partial l_i} = \frac{\partial W}{\partial U_i} \cdot \frac{\partial U_i}{\partial l_i} + \lambda_i w_i \quad (6.22)$$

A condição de primeira ordem é:

$$\frac{\partial W}{\partial U_i} \cdot \frac{\partial U_i}{\partial l_i} = -\lambda_i w_i \tag{6.23}$$

Isso implica que a desutilidade marginal do trabalho, ponderada pelo peso de bem-estar social, deve ser igual ao salário.

Para a produção, a condição de eficiência é obtida ao derivar o lagrangeano em relação a z_j:

$$\frac{\partial \mathcal{L}}{\partial z_j} = \sum_{k=1}^{M} \frac{\partial f_j}{\partial z_{jk}} - \mu \cdot \frac{\partial T}{\partial z_j} \tag{6.24}$$

Para eficiência de produção, o valor marginal do produto dos insumos deve igualar o custo marginal desses insumos, considerando o imposto T.

O Teorema da Produção Eficiente afirma que, sob certas condições, é possível organizar a tributação de maneira que as decisões de produção não sejam distorcidas.

Formalmente, isso é expresso pela condição:

$$\frac{\partial f_j}{\partial z_{jk}} = \frac{w_{jk}}{p_k} \tag{6.25}$$

Ou seja, os preços dos insumos devem refletir seus custos marginais. Isso implica que impostos sobre bens intermediários devem ser evitados.

Agora, vamos considerar a estrutura de impostos ótima. A ideia central é minimizar as distorções econômicas enquanto maximiza o bem-estar social. A Regra de Ramsey para tributação ótima em bens de consumo estabelece que o imposto t_i sobre o bem i deve ser inversamente proporcional à elasticidade-preço da demanda:

$$\frac{t_i}{1+t_i} = -\frac{1}{\eta_i} \tag{6.26}$$

em que η_i é a elasticidade-preço da demanda pelo bem i. Bens com demanda inelástica devem ser mais fortemente tributados para minimizar a distorção no comportamento dos consumidores.

Finalmente, o problema de tributação envolve resolver as equações diferenciais obtidas a partir das condições de primeira ordem e integrar as funções de utilidade e produção para determinar a solução exata da estrutura de impostos.

O resultado final depende de vários parâmetros, incluindo as elasticidades de oferta e demanda, a função de bem-estar social, e as restrições orçamentárias.

Em uma economia aberta, o conjunto de produção é estendido porque é possível negociar a preços lineares (para um país pequeno) com outros países. O resultado Diamond-Mirrlees afirma que uma pequena economia aberta deve estar na fronteira do conjunto de produção estendida. Implica que nenhuma tarifa deve ser imposta sobre bens e insumos importados ou exportados pelo setor produtivo.

Este modelo implica que o governo precisa ser capaz de definir um conjunto completo de alíquotas diferenciadas para cada insumo e cada produto. O governo precisa ser capaz de taxar os lucros totalmente puros (ou a produção está em constante retorno à escala). Caso contrário, pode melhorar o bem-estar tributando indústrias que geram muitos lucros para melhorar as receitas em detrimento da eficiência da produção. O governo pode variar os preços dos bens de consumo sem alterar os preços de produção. Mesmo que o governo esteja limitado à segunda melhor situação no problema de consumo, não há razão para adotar a segunda melhor solução no problema de produção.

A relevância prática do resultado é um pouco menos clara:

- A hipótese 1 (alíquotas diferenciadas) não é realista. Exemplo: insumos de trabalho qualificados e não qualificados devem ser diferenciados. Quando eles não podem (como no atual sistema de imposto de renda), então seria ótimo subsidiar indústrias intensivas de baixa qualificação ou estabelecer tarifas sobre produtos importados intensivos de baixa qualificação (para proteger a indústria doméstica).

- Fórmulas fiscais ótimas, mesmo quando os preços ao produtor não são constantes, tomam a mesma forma que o problema de Ramsey.

7 Tributação da Renda do Trabalho

Most economists would probably agree that the importance of mathematics to economics stems from its usefulness in developing economic intuitions. A mathematical formulation provides a logical test of an economic intuition. Also, the rigorous development of economic ideas can itself suggest new ones. In these ways, the influence of mathematics on economics is analogous to that of empirical work.

Bewley and Shafer, 1988

7.1 Introdução

Este capítulo considera a tributação ótima da renda do trabalho, ou seja, a distribuição justa e eficiente da carga tributária em indivíduos com rendimentos diferentes. Uma grande literatura acadêmica desenvolveu modelos de teoria tributária ótima para esclarecer essa questão. Esses modelos tipicamente postulam que o sistema tributário deve maximizar uma função de bem-estar social sujeita a uma restrição orçamentária do governo, levando em conta como os indivíduos respondem a impostos e transferências. O bem-estar social é maior quando os recursos são distribuídos de forma mais igualitária, mas os impostos e transferências redistributivas podem afetar negativamente os incentivos para trabalhar e ganhar renda em primeiro lugar. Isso cria o *trade-off* clássico entre equidade e eficiência, que está no cerne do problema do imposto sobre o rendimento do trabalho.

Metodologicamente, um objetivo central da análise tributária ótima deve ser lançar luz sobre as questões reais da política tributária e ajudar a projetar melhores sistemas tributários. Teoria e derivações técnicas são muito valiosas para modelar rigorosamente o problema em questão. Um dos principais objetivos deste capítulo é mostrar como tornar tais descobertas teóricas aplicáveis. Como argumentado em Diamond e Saez (2011), os resultados teóricos em análise tributária ótima são mais úteis para recomendações de políticas quando três condições são atendidas:

1. os resultados devem basear-se em mecanismos econômicos que sejam empiricamente relevantes e de primeira ordem para o problema em questão;

2. os resultados devem ser razoavelmente robustos aos pressupostos de modelagem e, em particular, à presença de heterogeneidade nas preferências individuais;

3. a prescrição da política tributária precisa ser implementável – isto é, a política tributária precisa ser relativamente fácil de explicar e discutir publicamente e não muito complexa de administrar em relação à prática real.

Essas condições nos levam a adotar duas escolhas metodológicas. Primeiro, usamos a abordagem de "estatísticas suficientes[41]", pela qual fórmulas fiscais ótimas são derivadas e expressas em termos de estatísticas estimáveis, incluindo pesos de bem-estar social marginal que capturam o valor da sociedade para redistribuição e elasticidades da oferta de mão-de-obra capturando os custos de eficiência da tributação (ver Chetty, 2009). Essa abordagem nos permite entender os principais mecanismos econômicos por trás das fórmulas, ajudando a atender a condição (1) acima. As fórmulas de "estatísticas suficientes" também são frequentemente robustas

[41] Em estatística e econometria, o conceito de estatísticas suficientes está relacionado à ideia de eficiência na estimação de parâmetros. Uma estatística é considerada suficiente para um parâmetro se ela captura toda a informação relevante sobre esse parâmetro que é fornecida pelos dados observados. Em outras palavras, uma estatística suficiente resume os dados de maneira que nenhuma outra informação sobre o parâmetro possa ser obtida a partir dos dados restantes. Mais formalmente, dado um conjunto de dados $X = (X_1, X_2, \ldots, X_n)$ e um parâmetro θ que queremos estimar, uma estatística $T(X)$ é chamada de suficiente para θ se a distribuição condicional de X dado $T(X)$ não depende de θ. Em termos mais técnicos, $T(X)$ é suficiente para θ se a função de verossimilhança dos dados pode ser fatorada em um produto da forma:

$$L(\theta|X) = g(T(X), \theta) \cdot h(X),$$

em que $g(T(X), \theta)$ é uma função que depende da estatística suficiente $T(X)$ e do parâmetro θ, e $h(X)$ é uma função que depende apenas dos dados X e não do parâmetro θ.

Um exemplo clássico de uma estatística suficiente é a soma dos dados em uma amostra quando a amostra é extraída de uma distribuição normal com média conhecida e variância desconhecida. Neste caso, a soma dos dados (ou, mais precisamente, a média) é uma estatística suficiente para a variância da distribuição normal.

Propriedades Importantes de Estatísticas Suficientes:

1. Redução de dados: uma estatística suficiente pode reduzir o conjunto de dados original sem perder informação sobre o parâmetro de interesse.
2. Teorema de Fatorização de Neyman-Fisher esse teorema fornece uma condição necessária e suficiente para que uma estatística seja suficiente. Se a função de verossimilhança pode ser fatorada conforme descrito acima, então a estatística é suficiente.
3. O Princípio de Lehmann-Scheffe: este princípio afirma que, sob certas condições, a estatística suficiente que é também uma estatística de estimativa não viesada (ou seja, que não é tendenciosa) tem as melhores propriedades de estimador entre todos os estimadores não viesados.

Em resumo, a noção de estatística suficiente é crucial para a inferência estatística porque permite simplificar problemas complexos e fazer estimativas mais eficazes sobre parâmetros de interesse com base nos dados disponíveis.

para alterar as primitivas do modelo, que satisfazem a condição (2).

Em segundo lugar, tendemos a nos concentrar em estruturas fiscais simples – por exemplo, um imposto de renda linear - sem sistematicamente tentar derivar o sistema fiscal mais geral possível. Isso ajuda a atender a condição (3), pois as estruturas fiscais que obtemos estarão, por definição, dentro do domínio das estruturas tributárias existentes. Isso contrasta com a abordagem do "desenho de mecanismo" que deriva o imposto ótimo mais geral compatível com a estrutura informacional. Essa abordagem de "desenho de mecanismo" tende a gerar estruturas tributárias altamente complexas e resultados sensíveis às primitivas exatas do modelo. A abordagem do "desenho de mecanismo" recebeu um interesse renovado na nova literatura de finanças públicas dinâmica que se concentra principalmente nos aspectos dinâmicos da tributação.

7.2 História da Tributação Ótima dos Rendimentos do Trabalho

Oferecemos aqui apenas uma breve visão geral cobrindo a tributação ótima de renda.

O surgimento do imposto de renda ocorreu relativamente tarde no desenvolvimento dos povos. A instituição de um real imposto sobre a renda exige um modelo econômico que possa ser avaliado e monitorado, para possibilitar o controle, a fiscalização e a cobrança do tributo.

O sistema econômico de trocas de produtos ou serviços por outros produtos ou serviços dificultava a medição da renda. Com a criação da moeda, houve uma unidade para determinar o acréscimo do patrimônio das pessoas, possibilitando determinar a renda e tributá-la. Em vez de a riqueza ser avaliada apenas pelos bens que o indivíduo possuía, pôde ser medida pelo produto desses bens, isto é, pela renda.

Segundo Seligman, o imposto de renda teve papel insignificante na Idade Média. Ele considerava que os impostos desse período tinham como fato gerador a produção e não a renda.

No século XV, surgiram, em Florença, os primeiros movimentos para uma efetiva tributação sobre a renda. A riqueza não mais decorria só da terra, mas também do comércio e da indústria. Foi criado o tributo conhecido como *catasto*, que transferiu a tributação direta da propriedade para a renda. Os autores divergem quanto à data de criação. Inicialmente não tinha caráter de progressividade, pois havia apenas uma alíquota.

Não demorou muito para que o *catasto* se tornasse progressivo e se denominasse *scala*. Instituía-se o que ficou conhecido como *Decima Scalata*, decima um nome genérico que se dava aos impostos e *scalata*, gradual, progressivo. A *Decima Scalata* não teve longa duração, porque atingia os mais abastados, que não aceitavam arcar com maior carga tributária. Com a reintrodução do regime aristocrático, não só a *Decima Scalata* mas também outros tributos diretos sobre a riqueza desapareceram.

A *Decima Scalata* foi um marco na história do imposto de renda e, segundo alguns estudiosos, a primeira demonstração de uma tributação sobre a renda.

No século XV, na Inglaterra, houve algumas tentativas infrutíferas de instituir um imposto sobre a renda. Alguns pesquisadores chegam a considerar que em 1404 foi criada uma tributação sobre a renda. No entanto, os documentos que tratavam da cobrança foram incinerados, pouco se sabe sobre o tributo e não há registros e provas confiáveis sobre a sistematização do imposto.

A maioria dos historiadores, pesquisadores e estudiosos considera o marco zero do imposto de renda na mesma Inglaterra, mas em 1799. Outros consideram que o pioneirismo coube a Florença no século XV. Registra-se que a documentação existente da experiência inglesa é mais farta, rica e confiável que a de Florença, inclusive com acesso a jornais da época. Há também quem aponte que o início do imposto de renda ocorreu, em 1710, na França.

O imposto de renda no Reino Unido foi introduzido pela primeira vez em 1799 durante o governo do primeiro-ministro William Pitt, o Jovem. O objetivo principal da criação do imposto de renda foi financiar os custos da guerra contra a França durante as Guerras Napoleônicas. No final do século XVIII, o Reino Unido estava envolvido em uma série de conflitos militares com a França, o que gerou grandes pressões financeiras sobre o governo britânico. William Pitt, o Jovem, percebeu que era necessário encontrar novas fontes de receita para financiar o esforço de guerra, uma vez que os impostos indiretos sobre bens de consumo e tarifas alfandegárias não eram suficientes para cobrir os custos crescentes. Em 1799, Pitt introduziu o imposto de renda como uma medida temporária. O imposto foi instituído sob a *Income Tax Act* de 1799, e a alíquota inicial foi de 2 *shillings* por libra (equivalente a 10%) sobre rendas acima de 200 libras anuais. O imposto foi progressivo, ou seja, os rendimentos mais altos eram tributados a uma taxa maior, enquanto rendimentos abaixo de 60 libras estavam isentos.

Os contribuintes foram divididos em três classes:

1. Na primeira, estavam os que possuíam criadagem, carros e cavalos, ou seja, os supostamente mais ricos.

2. Na segunda, a base de cálculo era medida em relógios, cães e janelas. Eram os contribuintes com razoável patrimônio.

3. A terceira se baseava na habitação. Atingia os contribuintes provavelmente mais pobres que os das duas classes anteriores.

O imposto era baseado em rendimentos provenientes de propriedades, profissões, ofícios e comércio. Rendimentos abaixo de 60 libras por ano eram isentos. A alíquota variava de acordo com o nível de renda, com uma taxa máxima de 10% sobre rendas superiores a 200 libras por ano.

Em 1798, com a pouca arrecadação do *assessed taxes*, William Pitt solicitou ao Parlamento transformá-lo num imposto efetivamente sobre a renda, em suma, um imposto geral provisório sobre todas as fontes de rendas mais importantes. Vem daí o nome *income tax*.

Em 3 de dezembro de 1798, na Câmara dos Comuns, Pitt foi defender a instituição do imposto de renda:

> *Na última sessão, aqueles que reconheceram quanto é importante levantar uma considerável parte das contribuições no decorrer do ano, limitaram a criticar os impostos já fixados, taxando-os de injustos e facilmente fraudáveis. Na realidade, parece que os resultados da arrecadação não corresponderam à expectativa, mas isso se deve não a um erro de cálculo dos nossos recursos nem a um exagero na avaliação da nossa riqueza, mas ao fato de se ter tornado muito fácil alterar a lei e por se ter procurado tornar a arrecadação a menos opressiva possível.*
>
> *Não obstante, os resultados obtidos atingiram plenamente a nossa expectativa no tocante aos benefícios decorrentes da medida e animam-nos a permanecer nos mesmos princípios. Deveríamos tomar por norma, antes de tudo, procurar, por meio de uma aplicação justa e rigorosa da lei, levantar a quota de um décimo que os impostos se propõem obter.*
>
> *Para isso, proponho que se ponha de lado uma crítica baseada exclusivamente nos impostos já em vigor e que se imponha um imposto geral sobre todas as fontes de rendas mais importantes.*

A lei continha 124 seções (artigos) e 152 páginas. Para facilitar o entendimento de um novo e complexo tributo, o governo preparou um manual de orientação, o primeiro da história do imposto de renda.

A instituição de um imposto sobre a renda gerou controvérsias. Charles Fox declarou na Câmara dos Comuns: “Quais serão as consequências desta lei? Os seus únicos resultados serão o imediato aniquilamento de nosso comércio, a destruição de nossas fortunas e provavelmente de nossa liberdade pessoal”.

Na mesma Câmara dos Comuns, Sir John Sinclair opunha-se ao novo tributo: “Se nos impuserem esse imposto, será lícito que algum dia nos livremos dele? Enquanto durar a guerra, isso não será possível. Se na paz esse acréscimo às rendas públicas for julgado indispensável? Agora, o ministro Pitt, usando de grande moderação, pede apenas um décimo de nossas rendas, mas o que impedirá de, no futuro, exigir um quinto ou mesmo um terço? Faz-se mister ainda observar que essa lei será pretexto para uma infinidade de exigências altamente vexatórias."

A reação foi veemente e, em alguns casos, agressiva. Alguns temiam que, terminada a guerra, o imposto continuasse a ser cobrado. Havia os que acreditavam que, com o tempo e a necessidade de mais recursos, a base de cálculo fosse aumentada. Outros achavam que interferiria na vida particular do contribuinte.

Os primeiros resultados da arrecadação não corresponderam à expectativa. Alcançaram, mesmo assim, mais do que o dobro do *assessed taxes.* Em 1802, o imposto sobre a renda foi suprimido, não por causa da demissão de Pitt, que ocorrera pouco antes, mas como consequência da paz transitória entre Inglaterra e França. Um ano após, as hostilidades recomeçaram. Novamente, o imposto sobre a renda foi lembrado como fonte de recursos. Addington, que foi o sucessor de Pitt, restabeleceu o imposto em 1803, com uma série de aperfeiçoamentos:

1. Rendimentos classificados e tributados por categoria, de acordo com sua origem. Essa classificação por categoria (cédula) vigorou no Brasil até o exercício de 1989.

2. Implantação da cobrança na fonte.

3. Isenção para pequenos rendimentos.

4. Dedução para encargos de família.

As alterações tributárias afetaram positivamente a arrecadação. Quando Pitt retornou ao governo, em 1804, manteve o sistema de Addington.

Em junho de 1815, Napoleão Bonaparte foi derrotado em Waterloo, Bélgica. Terminava a guerra. Apesar do bom resultado, o imposto sobre a renda havia sido instituído como forma de angariar receita para financiar a guerra. Não havia clima político para mantê-lo e foi novamente suprimido em 1816. O governo inglês foi obrigado a buscar outras fontes de renda.

Apesar de sua natureza temporária inicial, o imposto de renda tornou-se uma característica permanente do sistema tributário britânico quando foi reintroduzido em 1842 pelo então primeiro-ministro Robert Peel, em resposta a um déficit orçamentário crescente.

A primeira sinalização no Brasil sobre o imposto de renda, não com este nome, surgiu no início do segundo reinado com a Lei nº 317 de 21 de outubro de 1843, que fixou a despesa e orçou a receita para os exercícios de 1843/1844 e 1844/1845. O artigo 23 estabeleceu um imposto progressivo sobre os vencimentos percebidos pelos cofres públicos e vigorou por dois anos[42]. Assemelhava-se a uma tributação exclusiva na fonte.

Alcançava, de forma progressiva, apenas os que recebiam vencimentos dos cofres públicos. A reação foi tamanha que foi imediatamente suprimida, mas estabeleceu um movimento pioneiro na instituição do imposto de renda.

A análise moderna da taxação de renda ótima começou com Mirrlees (1971), que formulou e resolveu o problema com rigor. Ele considerou a maximização de uma função de bem-estar social com base em utilidades individuais sujeitas a uma restrição orçamentária do governo e restrições de incentivo decorrentes das respostas da oferta de trabalho dos indivíduos ao sistema tributário. Formalmente, no modelo Mirrlees, as pessoas diferem apenas por sua habilidade (ou seja, sua taxa de salário). O governo quer redistribuir renda dos indivíduos de alta habilidade para indivíduos de

[42] O Art. 23 diz "Fica creada a seguinte contribuição extraordinaria durante o anno desta lei. Também diz "Todas as pessoas que receberem vencimentos dos Cofres Publicos Geraes, por qualquer titulo que seja, ficão sujeitas a uma imposição, que será regulada pela maneira seguinte: De 500$000 a 1:000$000, 2 por cento. 1:000$000 a 2:000$000, 3 por cento. 2:000$000 a 3:000$000, 4 por cento. 3:000$000 a 4:000$000, 5 por cento. 4:000$000 a 5:000$000, 6 por cento. 5:000$000 a 6:000$000, 7 por cento. 6:000$000 a 7:000$000, 8 por cento. 7:000$000 a 8:000$000, 9 por cento. 8:000$000 para cima, 10 por cento. Fonte: https://www2.camara.leg.br/legin/fed/leimp/1824-1899/lei-317-21-outubro-1843-560743-publicacaooriginal-83901-pl.html.

baixa habilidade, mas só pode observar ganhos (e não habilidades). Assim, impostos e transferências são baseados em ganhos, levando a um trade-off eficiência-equidade.

Mirrlees (1971) teve uma enorme influência teórica no desenvolvimento da teoria dos contratos e da informação, mas pouca influência na formulação de políticas reais, já que as lições gerais para uma ótima política tributária eram poucas. O resultado mais surpreendente e discutido foi a famosa taxa de imposto marginal zero no topo.

Stiglitz (1982) desenvolveu a versão discreta do modelo de Mirrlees (1971) com apenas duas habilidades. Neste caso discreto, a taxa de imposto marginal na habilidade superior é zero, fazendo com que o resultado seja ainda maior do que no modelo contínuo de Mirrlees (1971). Isso provavelmente contribuiu para a saliência do resultado. O modelo discreto é útil para entender o problema da tributação ótima como um problema de informação que gera uma restrição de compatibilidade de incentivo para o governo. Ou seja, o sistema de impostos deve ser configurado para que o tipo de alta qualificação não queira trabalhar menos e imitar o tipo de baixa qualificação. Esse modelo discreto também é amplamente usado na teoria dos contratos e na organização industrial. No entanto, esse modelo discreto tem uso limitado para recomendações de políticas fiscais reais, porque é muito mais difícil obter fórmulas expressas em termos de estatísticas suficientes ou colocar números realistas no modelo discreto de duas habilidades do que no modelo contínuo.

Atkinson e Stiglitz (1976) derivaram um resultado muito importante e influente de que, sob as hipóteses de separabilidade e homogeneidade das preferências, a tributação diferenciada de mercadorias não é útil quando os rendimentos podem ser tributados não linearmente. Este famoso resultado foi influente tanto para moldar o campo da teoria tributária ótima quanto para os debates sobre política tributária. Teoricamente, contribuiu grandemente para mudar o foco teórico para a taxação não-linear ótima e para longe do modelo anterior de tributação diferenciada de commodities, de Diamond e Mirrlees (1971) (baseado na contribuição original de Ramsey de 1927). É com base nesse resultado que se deu apoio a um imposto sobre o valor agregado uniforme combinado com transferências baseadas na renda e tributação progressiva do rendimento. Ainda mais importante, o resultado de Atkinson e Stiglitz (1976) tem sido usado para argumentar contra a taxação da renda do capital e em favor da tributação apenas do lucro ou do consumo.

O problema fiscal linear ótimo é tecnicamente mais simples e era conhecido desde pelo menos Ramsey (1927) que a taxa ótima de imposto pode ser expressa em termos de elasticidades. Sheshinski (1972) é o primeiro tratamento moderno do problema do imposto de renda linear ótimo. Foi reconhecido desde cedo que as elasticidades da oferta de mão-de-obra desempenham um papel fundamental na taxa de imposto de renda linear ótima. No entanto, devido à desconexão entre a análise do imposto de renda não-linear e a análise tributária linear, não foi feita nenhuma tentativa sistemática de expressar fórmulas tributárias não-lineares em termos de "estatísticas suficientes" estimadas até há relativamente pouco tempo. Atkinson (1995), Diamond (1998), Piketty (1997) e Saez (2001) mostraram que as fórmulas tributárias não-lineares ideais também podem ser expressas de forma relativamente simples em termos de elasticidades. Isso permitiu conectar a teoria do imposto de renda ideal à grande literatura empírica que estimava as respostas comportamentais à tributação.

Diamond (1980) considerou um modelo tributário ótimo em que há respostas da oferta de trabalho, a chamada margem extensiva (em vez da margem intensiva do trabalho de Mirrlees, isto é, as horas de trabalho). Ele mostrou que a taxa marginal de imposto ótima pode, na verdade, ser negativa nesse caso. Como veremos, esse modelo com amplas margens recebeu atenção renovada na última década. Saez (2002a) desenvolveu fórmulas simples baseadas na elasticidade, mostrando que uma taxa de imposto marginal negativa (ou seja, um subsídio para o trabalho) é ótima na parte inferior da distribuição de oferta de mão-de-obra.

7.3 Equidade e Eficiência

Há duas questões principais envolvidas na tributação de renda. O primeiro é o efeito da tributação sobre a oferta de mão-de-obra. A tributação altera as escolhas que os consumidores fazem ao afetar o *trade-off* entre trabalho e lazer. A esse respeito, uma questão particularmente importante é se um aumento na taxa de imposto reduz necessariamente a oferta de mão-de-obra. Nesse caso, seria fornecido apoio ao argumento de que os impostos deveriam ser reduzidos para atender às necessidades de eficiência. A segunda questão que foi estudada é a determinação do nível ideal de tributação de renda. Por razões que ficarão claras, esse é um problema complexo, pois só pode ser abordado em um modelo com uma troca significativa entre eficiência e patrimônio. Dito isto, a busca pelo *trade-off* correto provou ser um caminho frutífero de investigação.

A ideia essencial que desejamos transmitir aqui é que é um grande erro projetar a estrutura do imposto de renda para atender aos motivos da equidade, sem levar em consideração o impacto no esforço de trabalho. Para entender por que, considere a solução ingênua de definir a taxa de imposto marginal em 100% para todas as rendas acima de algum nível de limite $\overline{z}$ e a uma taxa de zero para todas as rendas abaixo desse limite. Podemos esperar que essa estrutura tributária maximize a redistribuição possível dos ricos (aqueles acima do limite de renda) para os pobres (aqueles abaixo). No entanto, essa conclusão é incorreta quando os contribuintes respondem à estrutura tributária. O imposto confiscatório acima do limite remove o incentivo para ganhar mais de $\overline{z}$, e todos os que estavam acima desse nível optarão por receber exatamente essa quantia de renda. Isso define um círculo vicioso dinâmico. O governo deve abaixar o limiar, induzindo todos acima do novo nível a baixar novamente sua renda, e assim por diante, até que ninguém opte por trabalhar e a renda seja zero. Portanto, é lógico que devemos analisar a equidade da estrutura tributária em conjunto com seu efeito nos incentivos ao trabalho. A ideia é encontrar o esquema tributário que atenda aos objetivos sociais, conforme captado pela função de bem-estar social, dado o ajuste no esforço de trabalho e a participação no mercado de trabalho pelos contribuintes. Diz-se que esse regime tributário é ótimo, dependendo do objetivo estabelecido. Os resultados precisam ser interpretados com cautela, no entanto, porque são muito sensíveis à distribuição de habilidades na população e à forma da função de utilidade. Mais importante, eles dependem do objetivo da equidade incorporado à função de bem-estar social.

7.4 Taxação e Oferta de Trabalho

O efeito do imposto de renda na oferta de mão-de-obra pode ser investigado usando o modelo padrão de escolha do consumidor. A análise começará com a questão geral da oferta de mão-de-obra e passará a uma série de análises específicas sobre o efeito das variações no sistema tributário. O principal *insight* que isso fornecerá será destacar a importância dos efeitos renda e substituição.

Como padrão, supõe-se que o consumidor tenha um determinado conjunto de preferências em relação às alocações de consumo e lazer. O consumidor também possui um estoque fixo de tempo disponível, que pode ser dividido entre a oferta de mão de obra e o tempo gasto como lazer. A função utilidade que representa as preferências pode ser definida por

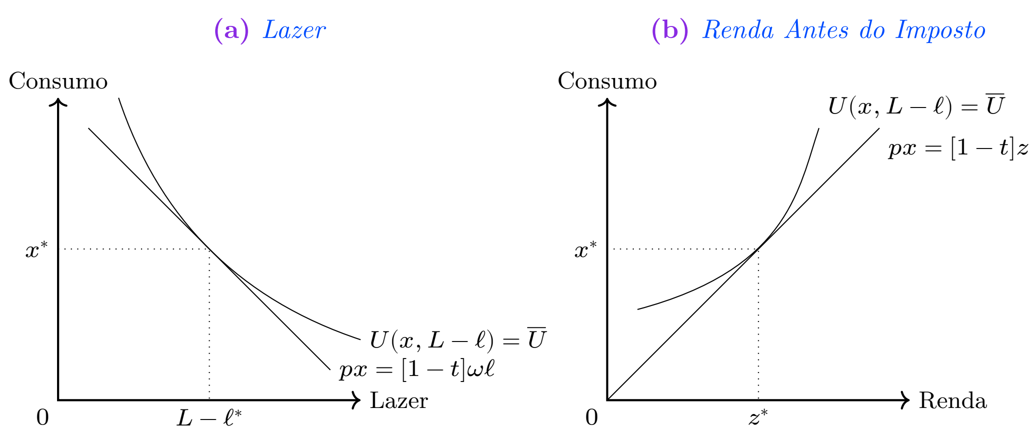
$$U = U(x, L - \ell) = U(x, \ell) \tag{7.1}$$

em que L é a dotação total de tempo, ℓ é a oferta de trabalho e x é o consumo. Consequentemente, o tempo de lazer é $L - \ell$. Presume-se que o trabalho seja desagradável para o trabalhador, de modo que a utilidade é reduzida à medida que mais trabalho é fornecido, o que implica em $\dfrac{\partial U}{\partial \ell} < 0$. Cada hora de trabalho rende uma taxa salarial de ω de modo que a renda, na ausência de tributação, é $\omega\ell$. Deixando que a taxa (constante) de imposto seja t, a restrição orçamentária que o consumidor enfrenta é $px = [1-t]\omega\ell$, em que p é o preço do bem de consumo.

O problema de escolha é mostrado no painel 7.1a da Figura 7.1, que representa graficamente o consumo contra o lazer. As curvas de indiferença e as restrições orçamentárias são padrão para a maximização da utilidade. A escolha ideal está na tangência da restrição orçamentária e na curva de indiferença mais alta possível. Isso resulta no consumo x^* e lazer $L-\ell^*$. Existe uma maneira alternativa de escrever a função de utilidade. Que a renda antes dos impostos seja denotada por z, de modo que $z = \omega\ell$. Como $\ell = \dfrac{z}{\omega}$, a utilidade pode ser escrita em termos de renda antes dos impostos, conforme

$$U = U\left(x, \frac{z}{\omega}\right) \tag{7.2}$$

Figura 7.1 – *Decisão de Oferta de Trabalho*

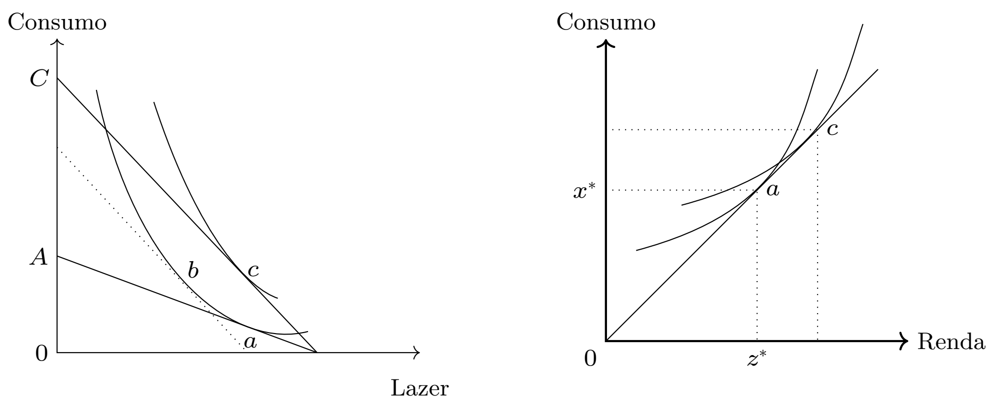


Essas preferências podem ser representadas em um gráfico da receita antes dos impostos em relação ao consumo. Expressa em termos de receita, a restrição orçamentária se torna $px = [1-t]z$. Isso é mostrado no painel 7.1b da Figura 7.1. A escolha ideal ocorre no ponto de tangência entre a curva de indiferença mais alta possível e a restrição orçamentária, com consumo x^* e receita antes dos impostos z^*. A característica importante dessa representação alternativa é que a restrição orçamentária não é afetada à medida que ω muda, portanto é a mesma que o salário que o consumidor ganha, mas as curvas de indiferença mudam, pois $\frac{z}{\omega}$ entra na função de utilidade.

Agora, esse modelo padrão pode ser usado para entender os efeitos de variações na taxa salarial ou tributária. Considere o efeito de um aumento na taxa salarial, que é mostrado no painel 7.2a da Figura 7.2 como uma mudança para a linha orçamentária mais alta e pela nova tangência em c.

Figura 7.2 – *Efeito de um Aumento Salarial*

(a) *Lazer* **(b)** *Renda Antes do Imposto*

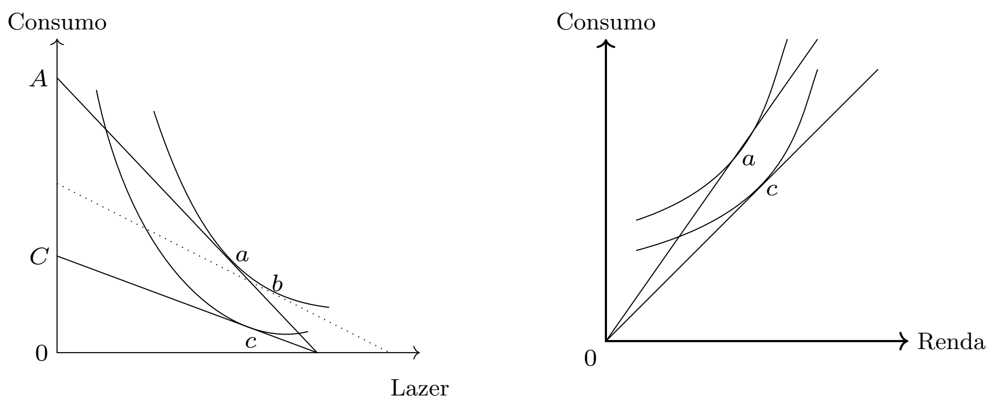


A mudança de a para c pode ser dividida em efeito de substituição (a para b) e efeito de renda (b para c). Com relação à direção do efeito de substituição, sempre é possível conhecê-la, uma vez que é dada por um movimento ao longo da curva de indiferença. Por outro lado, o efeito renda não pode ser definido previamente: pode ser positivo ou negativo. Consequentemente, o efeito líquido é ambíguo: um aumento na taxa salarial pode aumentar ou diminuir a oferta de mão-de-obra. Essa é a ambiguidade básica que ocorre ao longo da análise da oferta de mão-de-obra. O efeito de um aumento salarial quando as preferências são escritas como em 7.2 é mostrado no painel 7.2b da Figura 7.2. Um aumento na taxa salarial significa que menos mão-de-obra adicional é necessária para atingir qualquer aumento no consumo. Essa mudança no *trade-off* entre trabalho e consumo faz com que a curva de indiferença, através de um ponto, gire em volta e fique mais achatada. Esse achatamento das curvas de indiferença faz com que a escolha ideal se mova ao longo da restrição orçamentária. O nível de renda antes dos impostos aumentará, mas o efeito nas horas trabalhadas é ambíguo.

O efeito de um aumento de imposto é agora analisado da mesma maneira. No painel 7.3a da Figura 7.3, o aumento de imposto gira a linha do orçamento para baixo, de modo que a escolha ideal se move de a para c. O efeito substituição do aumento de imposto é a movimentação da curva de indiferença de a para b e o efeito renda afeta a movimentação de b para c. Usando a forma alternativa de preferências, um aumento na taxa de imposto gira a restrição orçamentária no painel b para baixo, de modo que o ponto escolhido se mova de a para c. Em nenhum diagrama, a alteração na taxa de imposto afeta as curvas de indiferença. Também é útil considerar sistemas tributários mais complexos usando essa abordagem. Uma característica comum do imposto de renda em muitos países é que existe um nível limite de renda abaixo do qual a renda não é tributada.

Figura 7.3 – *Efeito de um Aumento de Impostos*

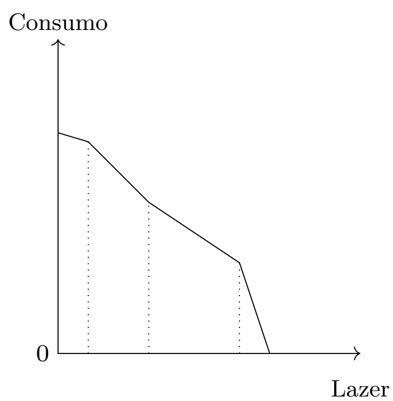
(a) *Lazer* **(b)** *Renda Antes do Imposto*



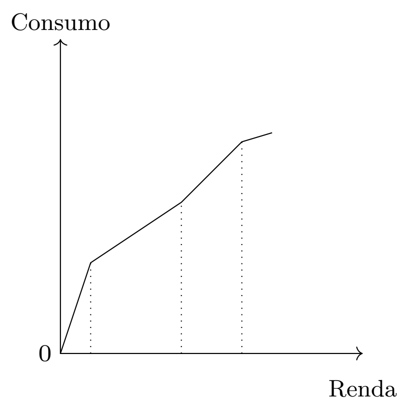
De um modo mais geral, um sistema de imposto de renda pode ter vários limites, com a taxa marginal de imposto subindo a cada um. Esse sistema tributário aparece como na figura 7.4. Novamente, com as preferências variando entre os consumidores, a expectativa é de que haja uma coleção de consumidores em cada ponto crítico. A questão final que vale a pena investigar nessa estrutura é a da participação na força de trabalho. A suposição básica até agora tem sido que o trabalhador pode variar continuamente o número de horas de trabalho para chegar ao resultado mais preferido. Na prática, geralmente é o caso em que as horas são fixas ou há um mínimo que deve ser realizado com a possibilidade de mais. Qualquer um dos casos leva a uma descontinuidade na restrição de orçamento no ponto de horas mínimas. A escolha para o consumidor é, então, entre não realizar trabalho e trabalhar pelo menos o mínimo. Esta é a decisão de participação: se deve ou não ingressar na força de trabalho.

Figura 7.4 – *Múltiplos Thresholds*

(a) *Lazer*



(b) *Renda Antes do Imposto*



O efeito de um aumento na tributação é diminuir a restrição orçamentária. Um consumidor que antes era indiferente entre trabalhar e não (ambos os pontos estão na mesma curva de indiferença) agora prefere estritamente não fazê-lo. Nesta margem, não há conflito entre renda e efeitos substituição. Um aumento na tributação reduz estritamente a participação na força de trabalho.

7.5 Tributação Ótima

A análise até este ponto considerou a questão positiva de como o imposto de renda afeta a oferta de trabalho. Tendo entendido isso, agora é possível voltar à questão normativa de como a estrutura do imposto de renda deve ser determinada. Esta é, por natureza, uma questão complexa. Na prática, os sistemas de imposto de renda geralmente têm vários limites nos quais a taxa marginal de imposto aumenta. Uma investigação do sistema ótimo deve ser pelo menos flexível o suficiente para considerar esses sistemas tributários sem limitar o número de limites ou as taxas de imposto em cada um. De fato, deve fazer mais do que isso. O modelo de tributação de renda introduzido por Mirrlees (1971) possui vários atributos importantes. Primeiro, há uma distribuição desigual de renda, portanto existem motivações patrimoniais para a tributação. Segundo, o imposto de renda afeta as decisões de oferta de mão-de-obra dos consumidores, de modo que tenham consequências de eficiência. Terceiro, tendo em vista os comentários acima, a estrutura é suficientemente flexível para que não sejam impostas restrições prévias às funções fiscais ideais que possam surgir.

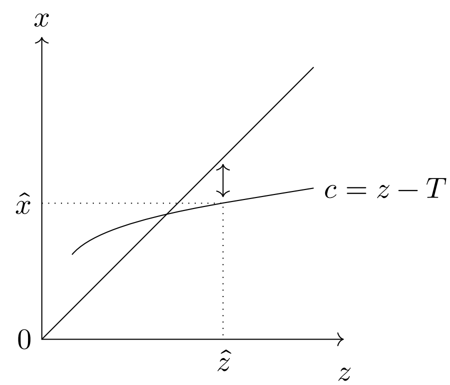
No modelo, todos os consumidores têm preferências idênticas, mas diferem em seu nível de habilidade no emprego. O salário por hora recebido por cada consumidor é determinado pelo seu nível de habilidade. Isso combina com a decisão de oferta de mão-de-obra para determinar a renda. Como a economia é competitiva, a taxa salarial também é igual ao produto marginal do trabalho e as empresas precificam seu produto com base no custo marginal. Um imposto cobrado sobre a habilidade seria uma política *first-best*, pois seria um imposto fixo sobre a característica inalterável que diferencia os consumidores. Mas não é viável, uma vez que se supõe que o nível de habilidade seja uma informação privada e não observável pelo governo. Isso torna impossível tributar diretamente a habilidade. Como o governo não pode observar o nível de habilidade de um consumidor (que é essencialmente a dotação inicial do consumidor), emprega um imposto de renda como a segunda melhor política. A função de imposto de renda é escolhida para maximizar o bem-estar social sujeito a gerar receita suficiente para atender aos requisitos do governo.

Existem duas mercadorias: um bem de consumo e trabalho. A oferta de trabalho de um consumidor é denotada por ℓ e o consumo por x. Cada consumidor é caracterizado por um nível de habilidade s. O valor de s mede a produção por hora do consumidor e, como a economia é competitiva, é igual à taxa salarial. Se um consumidor de habilidade s fornece ℓ horas de trabalho, esse consumidor ganha um salário de $s\ell$ antes de impostos. Seja a renda do consumidor com habilidade$s$ denotada por $z(s) = s\ell(s)$. O valor do imposto pago sobre a renda z é dado por $T(z)$. Essa é a função tributária que a análise pretende determinar. Equivalentemente, denote a função de consumo por $c(z)$ para que um consumidor que obtém renda z possa consumir

$$x = c(z) = z - T(z) \tag{7.3}$$

A relação entre renda, função tributária e consumo é mostrada na Figura 7.5. Na ausência de tributação, a renda seria igual ao consumo e isso é representado pela linha de 45 graus. Onde a função de consumo está acima da linha de 45 graus, o pagamento do imposto é negativo. É positivo quando a função de consumo está abaixo da linha. Por exemplo, o consumidor que ganha $\hat{z}$ na figura paga uma quantia de imposto $T(\hat{z})$ e pode consumir $\hat{x}$. O gradiente da função de consumo é igual a um menos a taxa marginal de imposto.

Figura 7.5 – *Taxação e Função Consumo*



Todos os consumidores têm a mesma função de utilidade (para excluir a possibilidade de trabalhadores com aversão diferente ao trabalho):

$$U = U\left(x, \frac{z}{s}\right) \tag{7.4}$$

As curvas de indiferença dependem do nível de habilidade do consumidor, pois um consumidor de alto nível de habilidade leva menos tempo de trabalho para atingir um determinado nível de renda do que um consumidor de baixa qualificação. Isso se reflete no fato de que, em qualquer par de renda e consumo $\{\hat{x}, \hat{z}\}$, a curva de indiferença de um consumidor de alta habilidade que passa por esse ponto é mais plana que a curva de indiferença de um consumidor de baixa qualificação. Essa propriedade de cruzamento único é denominada monotonicidade das preferências do agente.

Uma consequência imediata da monotonicidade do agente é que os consumidores de alta habilidade nunca ganharão menos renda do que a baixa habilidade. Geralmente, eles ganharão estritamente mais. Isso ocorre porque, no ponto em que a curva de indiferença do consumidor de baixa qualificação é tangencial à função de consumo (e, portanto, determina a escolha ideal para esse consumidor), a curva de indiferença do consumidor de alta habilidade é mais plana e, portanto, não pode ser uma tangência. Lembre-se de que todos os consumidores enfrentam a mesma função tributária e, portanto, a mesma função de consumo, independentemente de suas habilidades. A escolha ideal para o consumidor altamente qualificado deve, portanto, estar à direita do ponto ótimo para o trabalhador de baixa habilidade, o que implica um nível de renda mais alto.

Uma propriedade da função tributária ideal está relacionada à taxa máxima de imposto que será cobrada. Se a função de consumo se inclinar para baixo, o formato das curvas de indiferença garantirá que nenhum consumidor escolha localizar na seção de inclinação para baixo. Esta parte da função de consumo é, portanto, redundante. Economicamente, ao longo da seção inclinada para baixo, o aumento do esforço de trabalho é alcançado com menor consumo. Portanto, não há incentivo para trabalhar mais e esses pontos não serão escolhidos. Como $c(z) = z - T(z)$, segue-se que $c'(z) = 1 - T'(z)$. O argumento mostrou que $c'(z) \geq 0$, o que implica $T'(z) \leq 1$, portanto a taxa marginal de imposto é inferior a 100%.

Para fornecer mais informações sobre o imposto de renda ideal, vale a pena considerar uma especificação para o modelo. Vamos considerar uma forma especial da função de utilidade. Supõe-se que a utilidade é quase-linear em relação à renda do trabalho, isto é,

$$U\left(x, \frac{z}{s}\right) = u(x) - \frac{z}{s} \tag{7.5}$$

de modo que a desutilidade marginal do trabalho é $\ell = \dfrac{z}{s}$ é constante. A utilidade do consumo, $u(x)$, é crescente e côncava (então $u' > 0$ e $u'' < 0$). Para esta função de utilidade, a taxa marginal de substituição entre consumo e renda é $TMS_{x,z} = \dfrac{1}{su'(x)}$. Como a taxa marginal de substituição é decrescente em s, o gradiente da curva de indiferença para qualquer valor de x cai à medida que s aumenta. Isso torna a função de utilidade consistente com a monotonicidade do agente.

Simplificamos ainda mais assumindo que existem apenas dois consumidores, um com alto nível de habilidade, s_h, e outro com baixo nível, s_l. Supõe-se que $s_l < s_h < 3s_l$ (a razão para isso é explicada mais adiante). Com apenas dois consumidores, o problema de escolher a função ideal de imposto (ou consumo) pode receber a seguinte formulação: com qualquer função de consumo selecionada pelo governo, o fato de haver apenas dois consumidores garante que, no máximo, dois pontos serão escolhidos. Por exemplo, a_l é a alocação escolhida pelo consumidor de baixa qualificação e a_h é a alocação escolhida da alta habilidade. Tendo observado isso, é evidente que selecionar a função de consumo é equivalente a especificar as duas alocações. O restante da função de consumo pode ser escolhido para garantir que nenhum ponto seja melhor para os consumidores do que as duas alocações escolhidas. Essencialmente, a função de consumo só precisa vincular os dois pontos, enquanto em outros lugares permanece abaixo das curvas de indiferença através dos pontos. Seguir esse raciocínio reduz a escolha da função tributária a um simples problema de maximização envolvendo os dois locais.

Um consumidor só escolherá a alocação destinada a ele se preferir sua própria localização à do outro consumidor. Em outras palavras, as alocações devem ser compatíveis com incentivos. Como o consumidor de alta habilidade pode imitar a baixa habilidade, mas não vice-versa, a restrição de compatibilidade de incentivos deve ser vinculativa para o consumidor de alta habilidade. Denotando o local destinado ao consumidor de baixa qualificação por $\{x_l, z_l\}$ e o local destinado a alta qualificação por $\{x_h, z_h\}$, a restrição de compatibilidade de incentivo é

$$u(x_h) - \frac{z_h}{s_h} = u(x_l) - \frac{z_l}{s_h} \tag{7.6}$$

O consumidor de alta habilidade é indiferente entre as duas alocações (a_l e a_h estão na mesma curva de indiferença para a alta habilidade), enquanto o consumidor de baixa qualificação prefere estritamente a alocação a_l.

A otimização enfrentada por um governo que maximiza uma função utilidade de bem-estar social é

$$\max_{x_l, x_h, z_l, z_h} U = u(x_l) - \frac{z_l}{s_l} + u(x_h) - \frac{z_h}{s_h} \tag{7.7}$$

$$\text{sujeito a } x_l + x_h = z_l + z_h \tag{7.8}$$

A restrição de recursos supõe simplificadamente que nenhuma receita deve ser gerada para que o sistema tributário seja puramente redistributivo. O que é mostrado agora é que a quase linearidade da utilidade permite que esse problema de maximização seja consideravelmente simplificado. A simplificação permite então uma solução explícita.

Reescrevendo a restrição de compatibilidade de incentivo temos:

$$z_h = s_h[u(x_h) - u(x_l)] + z_l \tag{7.9}$$

Combinando essa equação com a restrição de recursos e eliminando z_h, a renda do consumidor de baixa qualificação pode ser escrita como

$$\begin{aligned} z_l &= x_l + x_h - z_h \\ z_l &= x_l + x_h - [s_h[u(x_h) - u(x_l)] + z_l] \\ 2z_l &= x_l + x_h - [s_h[u(x_h) - u(x_l)]] \\ z_l &= \frac{1}{2}\left[x_l + x_h - s_h[u(x_h) - u(x_l)]\right] \end{aligned} \tag{7.10}$$

De forma semelhante,

$$z_h = \frac{1}{2}\left[x_l + x_h + s_h[u(x_h) - u(x_l)]\right] \tag{7.11}$$

Substituindo essas expressões na função objetivo, encontramos:

$$
\begin{aligned}
\max_{\{x_l, x_h\}} U &= u(x_l) - \frac{z_l}{s_l} + u(x_h) - \frac{z_h}{s_h} \\
&= u(x_l) - \frac{1/2\,[x_l + x_h - s_h[u(x_h) - u(x_l)]]}{s_l} \\
&+ u(x_h) - \frac{1/2\,[x_l + x_h + s_h[u(x_h) - u(x_l)]]}{s_h} \\
&= u(x_l)\left[1 - \frac{1}{2}\frac{s_h}{s_l} + \frac{1}{2}\frac{s_h}{s_h}\right] + u(x_h)\left[1 + \frac{1}{2}\frac{s_h}{s_l} - \frac{1}{2}\right] - \frac{s_h s_l}{2 s_h s_l}(x_l + x_h) \\
&= \left[\frac{3s_l - s_h}{2s_l}\right] u(x_l) + \left[\frac{s_l + s_h}{2s_l}\right] u(x_h) - \frac{s_h s_l}{2 s_h s_l}(x_l + x_h) \\
&= \beta_l u(x_l) + \beta_h u(x_h) - \frac{s_h + s_l}{2 s_h s_l}(x_l + x_h) \qquad (7.12)
\end{aligned}
$$

A otimização enfrentada por um governo que maximiza uma função utilidade de bem-estar social é

$$
\max_{\{x_l, x_h\}} U = \beta_l u(x_l) + \beta_h u(x_h) - \frac{s_h + s_l}{2 s_h s_l}(x_l + x_h) \qquad (7.13)
$$

A suposição $s_h < 3s_l$ garante que β_l seja maior que zero, para que o consumidor de baixa qualificação tenha um peso social positivo. Sem essa suposição, a análise se torna mais complexa.

A comparação das duas funções objetivo permite uma nova interpretação do problema tributário ideal. A construção empreendida transformou a maximização da função de utilidade de bem-estar social sujeita a restrições na maximização de uma função ponderada de bem-estar sem restrições. A compatibilidade de incentivos e as restrições de recursos foram incorporadas, colocando um peso maior no bem-estar do consumidor de alta qualificação (desde que $\beta_h > \beta_l$), o que, por sua vez, garante que seu nível de consumo seja maior no ponto ótimo. Isso gera um nível de renda mais alto para o consumidor de alta qualificação. Também é possível observar que, à medida que a diferença de habilidade entre os dois consumidores aumenta, o peso relativo atribuído à alta habilidade aumenta também.

Vamos resolver o problema de otimização:

$$\frac{\partial U}{\partial x_l} = 0 \quad \Longleftrightarrow \quad \beta_l u'(x_l) - \frac{s_h + s_l}{2s_h s_l} = 0 \tag{7.14}$$

$$\frac{\partial U}{\partial x_h} = 0 \quad \Longleftrightarrow \quad \beta_h u'(x_h) - \frac{s_h + s_l}{2s_h s_l} = 0 \tag{7.15}$$

Para os consumidores de alta habilidade, substituindo os valores de β_h temos:

$$\begin{aligned} \beta_h &= \frac{1}{u'(x_h)} \frac{s_h + s_l}{2s_h s_l} \\ \frac{s_l + s_h}{2s_l} &= \frac{1}{u'(x_h)} \frac{s_h + s_l}{2s_h s_l} \\ u'(x_h) &= \frac{1}{s_h} \end{aligned} \tag{7.16}$$

Consequentemente, a utilidade marginal do consumidor de alta habilidade é inversamente proporcional ao seu nível de habilidade. Com $u''(x) < 0$ (utilidade marginal decrescente), isso implica que o consumo é proporcional à habilidade. Combinando este resultado com o fato de que $TMS^h_{x,z} = \frac{1}{su'(x)}$, segue-se que na alocação ótima temos $TMS^h_{x,z} = 1$. A constatação de que a taxa marginal de substituição é igual a 1 mostra que o consumidor de alta qualificação enfrenta uma taxa de imposto marginal zero. Este é o resultado sem distorção que já vimos.

Para o consumidor de baixa qualificação, temos:

$$\begin{aligned} \beta_l &= \frac{1}{u'(x_l)} \frac{s_h + s_l}{2s_h s_l} \\ \frac{3s_l - s_h}{2s_l} &= \frac{1}{u'(x_l)} \frac{s_h + s_l}{2s_h s_l} \\ u'(x_l) &= \frac{s_l + s_h}{s_h(3s_l - s_h)} \end{aligned} \tag{7.17}$$

e $TMS^l_{x,z} = \dfrac{s_h(3s_l - s_h)}{s_l + s_h} < 1$. Isso mostra que o consumidor do tipo l enfrenta uma taxa marginal positiva de imposto.

O uso da utilidade quase-linear permite a construção de uma solução explícita para o problema ótimo do imposto de renda, que mostra como as descobertas gerais da seção anterior se traduzem nesse caso especial. É interessante observar a simples dependência dos níveis de consumo das habilidades relativas e a maneira pela qual as restrições se traduzem em um maior peso efetivo do bem-estar para o consumidor de alta habilidade. Isso mostra que esse consumidor precisa ser incentivado a fornecer mais mão-de-obra através da recompensa do consumo adicional.

Exemplo 7.1. *Suponha que a função de utilidade seja representada por*

$$U = x(1 - \ell) \tag{7.18}$$

em que x denota consumo e ℓ é o tempo de trabalho. O preço do bem de consumo é normalizado para 1. Ao trabalhar, o consumidor recebe um salário por hora ω, considerado exógeno. Sua renda do trabalho é assim $\omega\ell$. O governo tem duas soluções para aumentar algumas receitas: definir um imposto fixo T ou definir um imposto linear sobre a renda do trabalho à taxa t.

- *Determine ℓ_T, a oferta de mão-de-obra do indivíduo sob o imposto fixo.*

  O consumidor representativo maximiza a sua função de utilidade sujeito a restrição orçamentária:

$$\begin{aligned} &\max_{\ell} x(1-\ell) \\ &\textit{sujeito a } x \leq \omega\ell - T \end{aligned} \tag{7.19}$$

que pode ser reescrito como

$$\max_{\ell}\ (\omega\ell - T)(1-\ell) \tag{7.20}$$

A condição de primeira ordem é

$$\frac{\partial U}{\partial \ell} = 0 \quad \Longleftrightarrow \quad \omega - 2\omega\ell + T = 0 \tag{7.21}$$

que resulta em

$$\ell_T = \frac{1}{2} + \frac{T}{2\omega} \tag{7.22}$$

- *Determine* ℓ_t*, a oferta de trabalho do indivíduo sob o imposto linear sobre a renda do trabalho.*

  O consumidor representativo maximiza a sua função de utilidade sujeito a restrição orçamentária:

$$\begin{aligned} &\max_{\ell} x(1-\ell) \\ &\textit{sujeito a } x \leq \omega\ell(1-t) \end{aligned} \tag{7.23}$$

  que pode ser reescrito como

$$\max_{\ell}\ [\omega\ell(1-t)](1-\ell) \tag{7.24}$$

  A condição de primeira ordem é

$$\frac{\partial U}{\partial \ell} = 0 \quad \Longleftrightarrow \quad (1-2\ell)(1-t)w = 0 \tag{7.25}$$

que resulta em

$$\ell_t = \frac{1}{2} \tag{7.26}$$

- *Determine R, a receita gerada pelo imposto de renda do trabalho. Qual das duas soluções leva a uma maior oferta de mão-de-obra: o imposto de renda da mão-de-obra à taxa t ou o montante fixo T que gera a mesma receita que o imposto de renda do trabalho? Qual solução deve ser escolhida pelo governo?*

  De acordo com o imposto de renda do trabalho, a receita tributária pode ser expressa como:

$$R = \ell_t \omega t = \frac{1}{2}\omega t \tag{7.27}$$

Se o imposto fixo gera a mesma receita, então:

$$T = \frac{1}{2}\omega t \tag{7.28}$$

Sob esse imposto, a oferta de trabalho é:

$$\ell_T = \frac{1}{2} + \frac{\frac{1}{2}\omega t}{2\omega} = \frac{1}{2} + \frac{t}{4} \tag{7.29}$$

Para a mesma receita tributária total, a oferta de mão-de-obra é, portanto, mais alta sob o imposto global do que sob imposto de renda trabalhista. Esse resultado advém do fato de o uso do imposto de renda do trabalho distorcer os preços relativos (aqui, os preços relativos de lazer e consumo). Por outro lado, este não é o caso do imposto fixo. Consequentemente, o governo deve optar por usar o imposto fixo T em vez do imposto sobre o rendimento do trabalho t.

7.6 Evasão Fiscal

A análise teórica da relação entre tributação e tomada de risco tem se concentrado principalmente no efeito dos impostos sobre as decisões de portfólio dos consumidores, como em Mossin (1968b) e Stiglitz (1969). No entanto, existem alguns problemas que não se enquadram naturalmente nessa categoria e que, embora tenham considerável interesse prático, foram deixados de fora das discussões teóricas. Um desses problemas é a evasão fiscal. Essa evasão assume várias formas, e dificilmente se pode esperar uma análise completamente geral de todas elas. O objetivo é analisar a decisão do contribuinte individual sobre se deve, e em que medida, evitar impostos por meio da subdeclaração deliberada. Por um lado, nossa abordagem está relacionada aos estudos sobre a economia da atividade criminosa, como, por exemplo, nos trabalhos de Becker (1968) e de Tulkens e Jacquemin (1971). Por outro lado, está relacionada à análise de políticas ótimas de portfólio e seguro na economia da incerteza, como nos trabalhos de Arrow (1970), Mossin (1968a) e vários outros.

A decisão sobre a declaração de imposto é uma decisão tomada sob incerteza. Isso ocorre porque a omissão de parte da renda à autoridade fiscal não provoca automaticamente uma reação na forma de uma penalidade. O contribuinte tem a escolha entre duas estratégias principais: (1) pode declarar sua renda real; (2) pode declarar menos do que sua renda real. Se escolher a segunda estratégia, seu resultado dependerá de ser ou não investigado pela autoridade fiscal. Se não for investigado, estará claramente em melhor situação do que na estratégia (1). Se for investigado, estará em pior situação. Portanto, a escolha da estratégia não é trivial.

Assumiremos que o comportamento do contribuinte está de acordo com os axiomas de Von Neumann-Morgenstern para comportamento sob incerteza. Sua função de utilidade cardinal tem a renda como único argumento; isso deve ser entendido como a função de utilidade indireta com preços constantes. Supõe-se que a utilidade marginal seja sempre positiva e estritamente decrescente, de modo que o indivíduo é avesso ao risco.

A renda real, W, é dada exogenamente e é conhecida pelo contribuinte, mas não pelo fiscal do governo. O imposto é cobrado a uma alíquota constante, θ, sobre a renda declarada, X, que é a variável de decisão do contribuinte. Contudo, com certa probabilidade p, o contribuinte será submetido a uma investigação pelas autoridades fiscais, que então conhecerão o valor exato da sua renda real. Se isso ocorrer, o contribuinte terá que pagar imposto sobre a quantia não declarada, $W - X$, a uma alíquota de penalidade π, que é maior que θ.

Essa representação formal da situação de escolha do contribuinte é, de certa forma, uma simplificação significativa da sua situação real; em particular, a formulação atual ignora alguns elementos de incerteza. Primeiro, abstrai do fato que as leis fiscais, em certa medida deixam a critério dos tribunais determinar se a penalidade será do tipo discutido aqui ou se assumirá a forma de uma pena de prisão; também pode ser uma combinação de ambas. Em segundo lugar, mesmo que a prisão não seja uma alternativa, a alíquota da penalidade π pode ser, do ponto de vista do contribuinte, incerta. Embora ignoremos esses pontos, esperamos ter mantido estrutura suficiente do problema para que a análise teórica seja válida.

O contribuinte escolherá agora X de forma a maximizar

$$\mathbb{E}[U] = (1-p)U(W - \theta X) + pU(W - \theta X - \pi(W - X)) \tag{7.30}$$

Por conveniência notacional, definimos

$$Y = W - \theta X$$

$$Z = W - \theta X - \pi(W - X)$$

A condição de primeira ordem para um máximo interior de (7.30) pode então ser escrita como

$$-\theta(1-p)U'(Y) - (\theta - \pi)pU'(Z) = 0 \tag{7.31}$$

A condição de segunda ordem

$$D = \theta^2(1-p)U''(Y) + (\theta - \pi)^2 pU''(Z) \tag{7.32}$$

é satisfeita pela suposição da concavidade da função de utilidade.

Nesta análise, as condições para que um máximo interior exista são de particular importância. Claramente, não se pode assumir *a priori* que $0 < X < W$, pois isso deve depender dos valores dos parâmetros. Para verificar quais condições sobre os valores dos parâmetros são necessárias para que haja uma solução interior, avaliamos a utilidade esperada em $X = 0$ e em $X = W$. Como a utilidade marginal esperada decresce com X, devemos ter que

$$\left.\frac{\partial \mathbb{E}[U]}{\partial X}\right|_{X=0} = -\theta(1-p)U'(W) - (\theta - \pi)pU'(W(1-\pi)) > 0 \tag{7.33}$$

e

$$\left.\frac{\partial \mathbb{E}[U]}{\partial X}\right|_{X=W} = -\theta(1-p)U'(W(1-\theta)) - (\theta - \pi)pU'(W(1-\theta)) < 0 \tag{7.34}$$

Essas condições podem ser reescritas como

$$p\pi > \theta\left[p + (1-p)\frac{U'(W)}{U'(W(1-\theta))}\right] \tag{7.35}$$

$$p\pi < \theta \tag{7.36}$$

A condição (7.36) implica que o contribuinte declarará menos do que sua renda real se o pagamento esperado de imposto sobre a renda não declarada for menor que a alíquota regular. Como o fator entre colchetes em (7.35) é obviamente positivo e menor que um, essas duas condições nos fornecem um conjunto de valores positivos dos parâmetros que garantem uma solução interior. É com essas soluções que estaremos preocupados mais à frente.

Esta é uma teoria muito simples, e talvez possa ser criticada por dar pouca atenção aos fatores não pecuniários na decisão do contribuinte sobre evadir ou não impostos. Nem é preciso enfatizar que, além da perda de renda, podem existir outros fatores que afetam a utilidade caso a tentativa de evasão fiscal seja detectada. Esses fatores podem ser sumariamente caracterizados como afetando negativamente a reputação do indivíduo como cidadão da comunidade; podemos representar isso por uma variável adicional, s, na função de utilidade. Agora escrevemos a utilidade esperada como

$$\mathbb{E}[U] = (1-p)U(Y, s_0) + pU(Z, s_1) \tag{7.37}$$

Assim, a variável s assume diferentes valores dependendo do estado do mundo (se a evasão é detectada ou não). Como convenção, assumimos que $U(Y, s_0) > U(Z, s_1)$. A condição de primeira ordem é então

$$-\theta(1-p)U_1(Y, s_0) - (\theta - \pi)pU_1(Z, s_1) = 0 \tag{7.38}$$

em que U agora denota a derivada de U em relação à variável renda. De especial interesse é agora a condição sobre os valores dos parâmetros que deve ser satisfeita para que $X < W$. Procedendo como nos casos estudados acima, obtemos essa condição como

$$p\pi < \theta \left[p + (1-p) \frac{U_1(W(1-\theta), s_0)}{U_1(W(1-\theta), s_1)} \right] \tag{7.39}$$

Observe primeiramente que (7.39) se reduz a (7.36) se $U_1(W(1-\theta), s_0) = U_1(W(1-\theta), s_1)$, ou seja, quando uma mudança na variável de estado não afeta a utilidade marginal da renda. A suposição mais natural talvez seja $U_1(W(1-\theta), s_0) < U_1(W(1-\theta), s_1)$; uma melhor reputação diminui a utilidade marginal da renda, de modo que "reputação" e renda são substitutos no sentido cardinal. Isso faria com que a expressão entre colchetes em (7.39) fosse menor que um e o lado direito da desigualdade fosse menor que zero, tornando a condição para evasão fiscal "lucrativa" mais rigorosa. Dependendo do valor de $\dfrac{U_1(W(1-\theta), s_0)}{U_1(W(1-\theta), s_1)}$ pode-se observar diferentes valores "ponto de equilíbrio" dos parâmetros para diferentes contribuintes.

Agora vamos examinar de que forma a renda declarada depende dos parâmetros do modelo, W, θ, π e p. Faremos isso usando o modelo mais simples dos dois apresentados acima, no qual o único argumento na função de utilidade do contribuinte é sua renda líquida. Isso representa uma simplificação do argumento em comparação com o modelo alternativo, na medida em que as várias derivadas em relação à renda dependerão do valor de s.

Faremos uso das conhecidas medidas de aversão ao risco de Arrow-Pratt para avaliar nossos resultados. Estas são as funções de aversão absoluta e relativa ao risco, definidas como

$$R_A(Y) = -\frac{U''(Y)}{U'(Y)}, \quad R_R(Y) = -\frac{U''(Y)Y}{U'(Y)} \tag{7.40}$$

Parece haver uma presunção geral de que a aversão absoluta ao risco diminui com a renda; o caso da aversão relativa ao risco é mais complicado, e não nos comprometeremos com nenhuma hipótese específica quanto ao seu formato. Derivando a equação (7.31) com respeito a W e resolvendo para $\frac{\partial X}{\partial W}$, obtemos:

$$\frac{\partial X}{\partial W} = \frac{1}{D}\left[\theta(1-p)U''(Y) + (\theta-\pi)(1-\pi)pU''(Z)\right] \tag{7.41}$$

Substituindo a partir da equação (7.31), podemos reescrever isso como:

$$\frac{\partial X}{\partial W} = -\frac{1}{D}\theta(1-p)U'(Y)\left[-\frac{U''(Y)}{U'(Y)} + (1-\pi)\frac{U''(Z)}{U'(Z)}\right] \tag{7.42}$$

ou, usando (7.40),

$$\frac{\partial X}{\partial W} = -\frac{1}{D}\theta(1-p)U'(Y)\left[R_A(Y) - (1-\pi)R_A(Z)\right] \tag{7.43}$$

Assumindo que a aversão absoluta ao risco é decrescente, temos $R_A(Y) < R_A(Z)$. No entanto, o sinal da expressão entre colchetes depende do valor de π. Apenas no caso de $\pi \geq 1$ podemos concluir que a derivada é inequivocamente positiva.

Talvez seja de maior interesse estudar o sinal da derivada $\frac{\partial\left(\frac{X}{W}\right)}{\partial W}$, isto é, como varia a fração da renda real que é declarada à medida que a renda real muda? Como temos:

$$\frac{\partial\left(\frac{X}{W}\right)}{\partial W} = \frac{1}{W^2}\left(\frac{\partial X}{\partial W}W - X\right) \tag{7.44}$$

Podemos substituir a partir das equações (7.41) e (7.32) para obter:

$$\frac{\partial\left(\dfrac{X}{W}\right)}{\partial W} = \frac{1}{W^2}\frac{1}{D}\left[\theta(1-p)U''(Y)W + (\theta-\pi)(1-\pi)pU''(Z)W\right] \\ - \frac{1}{W^2}\frac{1}{D}\left[\theta^2(1-p)U''(Y)X + (\theta-\pi)^2pU''(Z)X\right] \tag{7.45}$$

Reagrupando os termos e substituindo a partir da equação (7.31), podemos escrever:

$$\frac{\partial\left(\dfrac{X}{W}\right)}{\partial W} = \frac{1}{W^2}\frac{1}{D}\left[\theta(1-p)U''(Y) + (\theta-\pi)pU''(Z)Z\right] \tag{7.46}$$

Podemos agora substituir nesta expressão a partir da condição de primeira ordem. Isso resulta em:

$$\frac{\partial\left(\dfrac{X}{W}\right)}{\partial W} = -\frac{1}{W^2}\frac{1}{D}\theta(1-p)U'(Y)\left[R_R(Y) - R_R(Z)\right] \tag{7.47}$$

Podemos então concluir que, quando a renda efetiva varia, a fração declarada aumenta, permanece constante ou diminui, conforme a aversão relativa ao risco seja uma função crescente, constante ou decrescente da renda. Não é fácil escolher qual dessas hipóteses sobre a função de aversão relativa ao risco é a mais realista. Portanto, contentar-nos-emos em adicionar esse resultado àqueles de natureza semelhante que já existem na literatura sobre economia da incerteza. No entanto, é interessante por si só observar que, mesmo um modelo tão simples como o presente, não gera um resultado simples sobre a relação entre renda e evasão fiscal.

Agora diferenciamos a equação (7.31) em relação a θ. Isso resulta em:

$$\frac{\partial X}{\partial \theta} = -\frac{1}{D} X \left[\theta(1-p)U''(Y) + (\theta - \pi)pU''(Z)\right] + \frac{1}{D}\left[(1-p)U'(Y) + pU'(Z)\right] \tag{7.48}$$

Substituindo a partir da equação (7.31), podemos reescrever isso como

$$\frac{\partial X}{\partial \theta} = \frac{1}{D} X\theta(1-p)U'(Y)\left[R_A(Y) - R_A(Z)\right] + \frac{1}{D}\left[(1-p)U'(Y) + pU'(Z)\right] \tag{7.49}$$

O segundo dos dois termos no lado direito é inequivocamente negativo. O primeiro termo é positivo, nulo ou negativo, conforme a aversão absoluta ao risco seja decrescente, constante ou crescente. Dentre essas possibilidades, a hipótese de aversão absoluta ao risco decrescente parece ser a mais atraente; mas, mesmo assim, devemos concluir que não surge uma hipótese clara sobre a relação entre a alíquota regular do imposto e a renda declarada.

O significado econômico desse resultado é melhor compreendido se considerarmos os dois termos na equação (7.49) como o efeito-renda e o efeito-substituição, respectivamente. O último é negativo porque um aumento na alíquota do imposto torna mais lucrativo evadir impostos na margem. O primeiro é positivo porque um aumento na alíquota torna o contribuinte menos rico, reduzindo tanto Y quanto Z para qualquer nível de X, e isso, sob aversão absoluta ao risco decrescente, tende a reduzir a evasão.

A próxima questão que investigamos é como a renda declarada depende da taxa de penalidade. A partir da equação (7.31), obtemos:

$$\frac{\partial X}{\partial \pi} = -\frac{1}{D}(W - X)(\theta - \pi)pU''(Z) - \frac{1}{D}pU'(Z) \tag{7.50}$$

Estes termos são ambos positivos, de modo que um aumento na taxa de penalidade sempre aumentará a fração da renda real declarada.

Por fim, diferenciamos a equação (7.31) em relação a p para obter

$$\frac{\partial X}{\partial p} = \frac{1}{D}\left[-\theta U'(Y) + (\theta - \pi)U'(Z)\right] \tag{7.51}$$

Essa derivada é positiva, isto é, um aumento na probabilidade de detecção levará sempre a uma maior renda declarada.

Resumindo a análise estática comparativa do nosso modelo, podemos notar que, embora não forneça resultados definitivos na análise das mudanças na renda real e na alíquota do imposto, resultados inequívocos podem ser obtidos para dois parâmetros do modelo que são de particular interesse para fins de política pública, a saber, a taxa de penalidade e a probabilidade de detecção.

O primeiro é um parâmetro sobre o qual a autoridade fiscal exerce controle direto; o segundo pode ser assumido como controlado indiretamente por meio da quantidade e eficiência dos recursos destinados à detecção da evasão fiscal.

O modelo implica que essas duas ferramentas de política são substitutas uma da outra. Enquanto a arrecadação fiscal esperada diminuiria com a redução de p, a perda na receita tributária poderia ser compensada por um aumento de π.

8 Tributação do Capital

Valid economic theory exists and is applicable to all economic systems and countries. There is not a special economic theory for capitalism and another for communism, although significant differences exist in the institutions and legal frameworks to which the theory is applied.

Alchian & Allen, 1972

8.1 Introdução

É claro que um imposto tão importante quanto o imposto de renda das empresas e com ramificações em tantos setores da economia deve ser analisado em termos de equilíbrio geral, em vez de termos parciais. A principal característica do arcabouço teórico que apresento é sua natureza de equilíbrio geral. Foi inspirado por uma longa tradição de escritos no campo do comércio internacional, em que os nomes de Heckscher, Ohlin, Stolper, Samuelson, Metzler e Meade estão entre os mais proeminentes. Esses escritores investigaram os efeitos do comércio internacional, ou de determinadas políticas comerciais, sobre os preços relativos dos fatores e a distribuição de renda. Aqui examinaremos os efeitos do imposto de renda sobre essas mesmas variáveis.

Nosso modelo divide a economia em dois setores (indústria), um corporativo e outro não corporativo, cada um empregando dois fatores de produção, trabalho e capital. O imposto de renda das empresas é visto como um imposto que atinge os ganhos de capital no setor corporativo, mas não no setor não corporativo. Supõe-se que ambas as indústrias sejam competitivas, com a produção em cada uma delas governada por uma função de produção homogênea de primeiro grau (incorporando retornos constantes de escala). Não investigamos os efeitos a curto prazo da imposição do imposto sobre as sociedades, supondo que são os efeitos a longo prazo que são de maior interesse teórico e prático. No curto prazo, o imposto será necessariamente suportado pelos ganhos de equipamentos de capital fixo no setor afetado, desde que se aplique a nossa premissa de concorrência. Mas isso implicará em um desequilíbrio no mercado de capitais, com a taxa líquida de retorno aos proprietários de capital no setor tributado sendo menor que a taxa líquida de retorno recebida pelos proprietários de capital no setor não tributado. Uma redistribuição dos recursos da economia resultará, caminhando em direção a um equilíbrio de longo prazo no qual as taxas líquidas de retorno ao capital sejam iguais nos dois setores. Nesse equilíbrio de longo prazo, os salários do trabalho também serão iguais nos dois setores e as quantidades disponíveis de trabalho e capital serão totalmente empregadas.

Supõe-se também que as quantidades disponíveis de trabalho e capital na economia não sejam afetadas pela existência do imposto. Essa suposição é bastante inócua no caso do trabalho, mas, no caso do capital, é certamente questionável. É altamente provável que, como resultado da imposição do imposto sobre as sociedades, a taxa de retorno líquida recebida pelos proprietários de capital seja menor do que seria na ausência desse imposto. Essa redução no retorno ao capital pode influenciar a poupança de duas maneiras: primeiro, porque agora os proprietários do capital têm menor renda total e, segundo, porque a taxa de retorno é menor. No primeiro, devemos ter em mente que qualquer forma alternativa de aumentar a mesma receita implicaria a mesma redução de renda no setor privado; o impacto na economia do imposto sobre as sociedades seria, portanto, diferente daquele, digamos, de um imposto de renda proporcional que produza a mesma receita, apenas como resultado das diferenças que possam existir entre os grupos econômicos em suas propensões de poupança. No segundo, devemos investigar a elasticidade da oferta de poupança em relação à taxa de juros. Se essa elasticidade for zero, a alteração na taxa líquida de juros enfrentada pelos poupadores não influenciará o tamanho do estoque de capital em um dado momento, nem o caminho pelo qual o estoque de capital cresce ao longo do tempo.

Sejam as seguintes suposições do modelo:

- Oferta fixa de fatores (curto prazo, economia fechada)
- Livre mobilidade de fatores entre setores
- Pleno emprego de fatores
- Retornos constantes de escala nos dois setores de produção
- Competição perfeita

8.2 Resultados da Incidência: O Caso Cobb-Douglas

Enquanto o mercado de capitais trabalhar para equilibrar as taxas de retorno líquidas de impostos e prêmios de risco, e enquanto a imposição de um imposto de renda corporativo não tiver, por si só, um efeito significativo no (padrão de) prêmios de risco associados a diferentes tipos de atividades, é inevitável que, a longo prazo, o imposto sobre as sociedades seja incluído no preço do produto. Ou seja, de dois setores, um corporativo e um não corporativo, cada um usando a mesma combinação de trabalho e capital para produzir uma unidade de produto, o preço de equilíbrio do produto corporativo será maior que o preço de equilíbrio do produto não corporativo precisamente pelo valor do imposto sobre as sociedades pago por unidade de produto. Este resultado é considerado por algumas pessoas como evidência de que a carga do imposto sobre as sociedades é suportada pelos consumidores, ou seja, que o imposto é deslocado para a frente. Essa inferência é muito ampla. Talvez a maneira mais fácil de demonstrar o erro da inferência acima seja apresentar um simples contra-exemplo. Considere uma economia que produz apenas dois produtos, o produto X, produzido por empresas na forma corporativa, e o produto Y, produzido por empresas não corporativas. Suponha que as características de demanda da economia sejam tais que os consumidores sempre gastem metade de sua renda disponível em X e metade em Y. Suponha que as funções de produção nos dois setores sejam do tipo Cobb-Douglas, com coeficientes $1/2$ para ambos os insumos, ou seja, $X = L_X^{1/2} K_X^{1/2}$ e $Y = L_Y^{1/2} K_Y^{1/2}$, em que L_X e L_Y representam as quantidades de trabalho utilizadas nas indústrias X e Y, e K_X e K_Y as quantidades correspondentes de capital. Presume-se que os montantes totais de trabalho e capital disponíveis para a economia sejam fixos, nos níveis L e K, respectivamente.

Sob concorrência perfeita, a produção em cada setor será levada ao ponto em que o valor do produto marginal de cada fator é igual ao preço pago pelos empreendedores pelos serviços do fator. Assim, na ausência de impostos, temos

$$L_X P_L = \frac{1}{2} X P_X \tag{8.1}$$

$$K_X P_K = \frac{1}{2} X P_X \tag{8.2}$$

$$L_Y P_L = \frac{1}{2} Y P_Y \tag{8.3}$$

$$K_Y P_K = \frac{1}{2} Y P_Y \tag{8.4}$$

Se a renda total da economia for de R$ 1.200, dividida igualmente entre X e Y, o trabalho na indústria X ganhará R$ 300, o trabalho na indústria Y ganhará R$ 300, o capital na indústria X ganhará R$ 300 e o capital na indústria Y ganhará R$ 300. É claro que tanto a força de trabalho quanto o estoque de capital terão que ser igualmente divididos entre as indústrias X e Y. Escolhendo nossas unidades de trabalho e capital para que, nessa posição de equilíbrio $P_L = P_K = R\$1$, teríamos o resultado de que, sem nenhum imposto, haverá 300 unidades de trabalho na indústria X e 300 na indústria Y, e o estoque de capital será distribuído da mesma forma.

Suponha agora que um imposto de 50% seja cobrado sobre os ganhos de capital da indústria X. O trabalho na indústria X será R$ 300, assim como o trabalho na indústria Y. Como o preço pago pelos empreendedores pela mão de obra também é o preço recebido pelos trabalhadores e, como é assumido o equilíbrio no mercado de trabalho, a distribuição de equilíbrio da mão-de-obra será a mesma neste caso como no anterior, ou seja, 300 trabalhadores em cada setor.

A situação é diferente, no entanto, quando chegamos ao capital. O preço pago pelos empreendedores pelo capital, multiplicado pela quantidade de capital utilizada, será novamente de R$ 300 em cada setor. Mas o preço pago pelos empresários da indústria X incluirá o imposto, enquanto o preço pago pela indústria Y não. Com um imposto de 50% sobre o valor total pago, o capital na indústria X receberá, líquido de impostos, apenas R$ 150, enquanto o capital na indústria Y receberá R$ 300.

Para obter o equilíbrio no mercado de capitais, deve haver duas vezes mais capital na indústria Y do que na indústria X. Assim, como resultado do imposto, a distribuição do capital muda: em vez de ter 300 unidades de capital em cada setor, agora temos 200 unidades no setor X e 400 unidades no setor Y. Do total de R$ 600 que os empresários estão pagando pelo capital de ambos os setores, metade será destinada ao capital do setor Y, sobre o qual nenhum imposto será pago, um quarto irá para o capital na indústria X, líquido de impostos, e um quarto irá para o governo como pagamento de impostos. O preço do capital cairá de R$ 1,00 para R$ 0,75.

Um cálculo bruto é suficiente para sugerir a incidência tributária resultante. De uma renda nacional de R$ 1.200, a mão-de-obra obteve R$ 600 antes da imposição do imposto e depois dela, mas o capital obteve apenas R$ 450 após o imposto contra R$ 600 antes do imposto. A diferença de R$ 150 é pago ao governo. O capital está claramente suportando o peso do imposto, apesar de, na situação tributária, o imposto ser incluído no que os consumidores estão pagando pela mercadoria X. É claro que isso não conta toda a história da incidência do imposto. Como o preço da mercadoria X aumenta e o preço da mercadoria Y cai, os consumidores com preferências particularmente fortes por um ou outro dos dois bens serão prejudicados ou beneficiados em seu papel de consumidores, além de qualquer benefício que obtenham. É importante perceber, no entanto, que o preço de Y diminui e que isso traz para os consumidores, como um grupo, um benefício que contrabalança a carga que eles carregam como resultado do aumento no preço de X.

Veja que:

$$
\begin{aligned}
&P_Y = \frac{2L_Y P_L}{Y} && \\
&P_Y = \frac{2 \cdot 300 \cdot 1}{(300)^{1/2}(300)^{1/2}} = 2 && \Longrightarrow \quad \text{Antes do imposto} \\
&P_Y = \frac{2 \cdot 300 \cdot 1}{(400)^{1/2}(300)^{1/2}} = 1,73 && \Longrightarrow \quad \text{Depois do imposto} \\
&P_X = \frac{2L_X P_L}{Y} &&
\end{aligned}
$$

$$P_X = \frac{2 \cdot 300 \cdot 1}{(300)^{1/2}(300)^{1/2}} = 2 \qquad \Longrightarrow \quad \text{Antes do imposto}$$

$$P_X = \frac{2 \cdot 300 \cdot 1}{(200)^{1/2}(300)^{1/2}} = 2\sqrt{6} \qquad \Longrightarrow \quad \text{Depois do imposto}$$

No equilíbrio pós-imposto, o valor do produto marginal do capital na indústria X excede o valor na indústria Y pelo valor do imposto, enquanto a alocação eficiente de capital exigiria que esses dois valores fossem iguais. Além disso, o padrão de consumo na economia também é tornado "ineficiente" pelo imposto, porque a taxa marginal de substituição de X por Y no consumo (que é dada pelas proporções de seus preços brutos de imposto) é diferente da taxa marginal de substituição técnica de X por Y na produção (que é dada pela razão de seus preços líquidos de impostos). O resultado dessa dupla ineficiência é que os mesmos recursos, embora totalmente empregados, produzam menos renda nacional na presença do imposto do que na sua ausência. Se negligenciarmos o "excesso de carga", poderemos tratar os efeitos das mudanças nos preços de X e Y como tendo exatamente compensações sobre o bem-estar do consumidor e determinar a incidência do imposto observando o que acontece com os preços do trabalho e do capital. Essa abordagem não exclui a carga total do imposto suportado pelos consumidores, pois nos casos em que os preços (líquidos de impostos) da mão-de-obra e do capital se movem nas mesmas proporções como resultado do imposto, é igualmente correto dizer que o imposto é suportado pelos consumidores, como também dizer que a carga tributária é compartilhada por trabalho e capital na proporção de suas contribuições iniciais para a renda nacional; exemplos de tais casos são dados abaixo.

Poderíamos resumir a análise da incidência do imposto assumido sobre o capital na indústria X da seguinte forma: essa redução na renda do capital se espalha por todo o capital, seja empregado na indústria X ou na indústria Y, assim que o mercado de capitais é novamente colocado em equilíbrio após a imposição do imposto. Na medida em que os consumidores individuais têm o mesmo padrão de despesa que a média de todos os consumidores, eles não ganham nem perdem seu papel de consumidores. Na medida em que os consumidores individuais diferem da média, eles ganham se gastam uma fração maior de seu orçamento em Y que a média e perdem se gastam uma fração maior de seu orçamento em X que a média. Os ganhos daqueles consumidores que preferem Y, no entanto, são contrabalançados pelas perdas daqueles que preferem X. Se estamos preparados para aceitar esse cancelamento de ganhos e perdas como base para uma declaração de que os consumidores como um grupo não sofrem como um consequência do imposto, então podemos concluir que o capital suporta o imposto. Caso contrário, devemos nos contentar em observar que as transferências brutas de indivíduos como capitalistas e consumidores de X excedem o rendimento do imposto em uma quantidade igual à transferência bruta para os consumidores de Y.

O exemplo acima é representativo de toda a classe de casos em que as despesas são divididas entre os bens em determinadas proporções, e a produção de cada bem é determinada por uma função Cobb-Douglas. Os expoentes das funções Cobb-Douglas podem diferir de indústria para indústria, e até as taxas de imposto sobre ganhos de capital podem ser diferentes em diferentes setores tributados; no entanto, a conclusão de que o capital suporta o imposto, no sentido indicado acima, permanece. É fácil demonstrar a verdade da afirmação acima. Seja A_i a fração da renda nacional gasta no produto da indústria i, seja B_i o coeficiente do insumo trabalho na i-ésima indústria (igual à fração das receitas da i-ésima indústria que é paga em salários ao trabalho), e seja $C_i(= 1 - B_i)$ o coeficiente do insumo capital no i-ésimo setor (igual à fração dos recebimentos do i-ésimo setor que é pago [bruto de imposto] ao capital). Então $\sum A_i B_i$ será a fração da renda nacional que vai para o trabalho, tanto na situação com tributação quanto no caso em que os impostos estão ausentes. Imediatamente, pode-se concluir que a participação do trabalho na renda nacional permanecerá a mesma nos dois casos. Além disso, a distribuição do trabalho entre as indústrias também permanecerá inalterada, pois cada indústria empregará a fração

$\dfrac{A_iB_i}{\sum A_iB_i}$ da força de trabalho em ambos os casos. Da mesma forma, o capital receberá uma fração fixa da renda nacional (bruta de imposto) igual a $\sum A_iC_i$. Quando um imposto é cobrado sobre o capital, o capital receberá $\sum A_iC_i(1-t_i)$ líquido de impostos, e o governo receberá $\sum A_iC_it_i$, em que t_i é a taxa percentual do imposto aplicável à renda do capital no i-ésimo setor. Assim, o capital como um todo perderá uma fração da renda nacional exatamente igual à recebida pelo governo em receitas fiscais. Como no caso apresentado no exemplo acima, a distribuição de capital entre as indústrias mudará como resultado da imposição do imposto, sendo a fração do estoque total de capital da i-ésima indústria $\dfrac{A_iC_i}{\sum A_iC_i}$ na ausência de impostos e $\dfrac{A_iC_i(1-t_i)}{\sum A_iC_i(1-t_i)}$ na presença do imposto. Exceto quando a alíquota do imposto de renda sobre o capital for igual em cada setor, haverá efeitos nos preços relativos e nas transferências de renda entre os consumidores, da mesma natureza geral que as descritas acima para o caso mais simples. Mas, como antes, os ganhos daqueles consumidores beneficiados com as mudanças nos preços relativos serão, em uma primeira aproximação, compensados pelas perdas daqueles consumidores prejudicados pelas mudanças de preços relativos; portanto, se aceitarmos essa compensação como um cancelamento de efeitos no que diz respeito às pessoas em seu papel de consumidor, podemos dizer que o capital suporta toda a carga do imposto.

8.3 Resultados da Incidência: O Caso de Proporções Fixas

Voltando agora a um exemplo em que existem apenas duas indústrias, suponhamos que a indústria tributada não seja caracterizada por uma função de produção Cobb Douglas, mas por uma função de produção na qual os fatores se combinem em proporções estritamente fixas. Vamos manter todas as outras suposições do exemplo anterior – que as despesas são divididas igualmente entre os dois produtos, que a produção na indústria Y é governada pela função $Y = L_Y^{1/2}K_Y^{1/2}$, que existem 600 unidades de cada fator e que os preços dos dois fatores são inicialmente cada um R$

1,00. Essas premissas determinam que o equilíbrio inicial antes dos impostos será o mesmo de antes, com 300 unidades de cada fator ocupado em cada setor. A função de produção de proporções fixas para a indústria X, que é consistente com essas suposições, é $X = \min\{L_X, K_X\}$.

O que acontece quando um imposto de 50% é imposto sobre a receita de capital na indústria X? É claro que qualquer redução na produção que possa ocorrer na indústria X, os dois fatores de produção serão liberados para a indústria Y em quantidades iguais. Como a indústria Y já está usando uma unidade de capital por unidade de trabalho, ela pode absorver incrementos nesses dois fatores na mesma proporção sem alterar a produtividade marginal de qualquer fator em termos físicos. O preço de Y terá que cair, no entanto, para criar uma demanda maior por ele. Qualquer que seja essa queda no preço de Y, induzirá uma queda proporcional no preço de cada um dos fatores (uma vez que suas produtividades físicas marginais permanecem inalteradas). Temos, portanto, o resultado de que, no equilíbrio final após o imposto, R$ 600 serão gastos no produto da indústria Y, com metade destinada ao capital e metade ao trabalho, e R$ 600 serão gastos no produto da indústria X, com R$ 200 indo para o trabalho, R$ 200 para o capital (líquido de impostos) e R$ 200 para o governo. O preço da mão-de-obra caiu de R$ 1,00 para R$ $\frac{5}{6}$. O imposto incidirá sobre capital e mão-de-obra na proporção de suas contribuições iniciais para a renda nacional.

Deveria ser evidente que o resultado recém-obtido, de trabalho e capital sofrendo a mesma carga percentual, depende criticamente do fato de que, no exemplo acima, a indústria Y estava em posição de absorver capital e trabalho exatamente nas proporções em que foram expulsos da indústria X sem alteração nos preços relativos dos dois fatores. Se a indústria X tivesse ejetado duas unidades de trabalho para cada unidade de capital, enquanto a indústria Y usasse inicialmente quantidades iguais dos dois fatores, o preço da mão-de-obra teria que cair em relação ao preço do capital, a fim de induzir o aumento da proporção de trabalho em relação ao capital na indústria

Y. Nesse caso, o trabalho suportaria mais impostos, em relação à sua participação na renda nacional, do que capital.

8.4 Modelo Geral

Embora os exemplos apresentados nas seções anteriores forneçam algumas dicas sobre a natureza do problema de incidência e sobre os fatores que provavelmente governarão a incidência do imposto de renda da empresa, eles sofrem com o defeito de se basearem em premissas restritivas específicas sobre a natureza das funções de demanda e produção.

Nesta seção, apresentarei um modelo de generalidade substancialmente maior. Suponha que haja dois produtos na economia, X e Y, com suas quantidades escolhidas para que seus preços sejam inicialmente iguais à unidade. A demanda por cada produto dependerá de seu preço relativo e do nível de renda dos demandantes. A renda dos consumidores cairá naturalmente como resultado da imposição do imposto e, com a consequente restrição de sua demanda por bens, parte dos recursos será destinado ao governo.

A demanda final dependerá de como os consumidores reagirão às mudanças de renda e de preço e de como o governo escolherá gastar o produto do imposto. Suponha, por uma questão de simplicidade, que a maneira pela qual o governo gastaria os impostos, se os preços iniciais continuassem a prevalecer, apenas contrabalançaria as reduções nas despesas privadas dos dois bens. Essa suposição, mais a suposição adicional de que as redistribuições de renda entre os consumidores não alteram o padrão de demanda, nos permitem tratar as mudanças na demanda em função das mudanças apenas nos preços relativos. Como o pleno emprego também é assumido, as funções de demanda para X e Y não são independentes; uma vez conhecido o nível de demanda por X, dados determinados preços e renda total do emprego, o nível de demanda por Y pode ser derivado das informações disponíveis.

Do lado da demanda, as demandas agregadas são geradas pela maximização de uma única função de utilidade sujeita a uma restrição orçamentária agregada. A função de utilidade é:

$$U(X, Y) \tag{8.5}$$

A restrição orçamentária reconhece que o único indivíduo possui todo o trabalho e capital na economia e também possui as indústrias em ambos os setores. No entanto, com retornos constantes de escala, não há lucros, portanto, a posse das indústrias não gera retorno por si só. Logo,

$$P_X X + P_Y Y = P_L L_0 + P_K K_0 \tag{8.6}$$

Em geral, teríamos:

$$X = X^D(P_X, P_Y, P_L, P_K) \tag{8.7}$$

$$Y = Y^D(P_X, P_Y, P_L, P_K) \tag{8.8}$$

Harberger, de fato, propõe:

$$X = X^D\left(\frac{P_X}{P_Y}\right) \tag{8.9}$$

A dependência apenas da razão de preços significa que a utilidade é homotética. A ausência dos termos de renda é mais complicada. As suposições de Harberger têm a intenção de implicar que isso não é afetado pelo imposto. Portanto, é constante e pode ser suprimido.

Harberger assume que o governo impõe um imposto infinitesimal em uma economia sem distorções. Assim, não há ônus excessivo. Em segundo lugar, a receita do imposto é devolvida de forma *lump-sum* ao indivíduo. Portanto, o imposto não tem efeito direto sobre as rendas. Como o dinheiro extra é arrecadado e gasto, isso é uma incidência de orçamento equilibrado.

Diferenciando totalmente (8.9), obtemos:

$$dX = X' \frac{P_Y dP_X - P_X dP_Y}{P_Y^2} \tag{8.10}$$

Normalizando os preços, chegamos a:

$$\frac{dX}{X} = \frac{X'}{X}(dP_X - dP_Y) \tag{8.11}$$

Definindo

$$\varepsilon = \frac{X'}{X}\frac{P_X}{P_Y} = \frac{X'}{X} \tag{8.12}$$

Portanto, podemos resumir as condições de demanda em nosso modelo por uma equação na qual a quantidade demandada de X depende da razão $\dfrac{P_X}{P_Y}$:

$$\frac{dX}{X} = \varepsilon \left(\frac{d\left(^{P_X}/_{P_Y}\right)}{\left(^{P_X}/_{P_Y}\right)} \right) = \varepsilon(dP_X - dP_Y) \tag{8.13}$$

em que ε é a elasticidade-preço da demanda por X e onde é usada a suposição de que os preços iniciais eram iguais a 1 para obtermos a expressão final.

Suponha que a função de produção para X seja homogênea de grau 1[43]. Isso

[43] Uma função de produção homogênea de grau 1 é uma função que descreve a relação entre os insumos utilizados na produção e a quantidade de produto produzido, mantendo uma propriedade

nos permite escrever a oferta de X como

$$\frac{dX}{X} = f_L \frac{dL_X}{L_X} + f_K \frac{dK_X}{K_X} \tag{8.14}$$

específica chamada homogeneidade de grau 1. Uma função de produção é uma representação matemática que relaciona a quantidade de insumos utilizados (como trabalho, capital, e outros recursos) com a quantidade de produto gerado. A forma geral de uma função de produção pode ser expressa como:

$$Q = f(L, K)$$

em que Q é a quantidade de produto, L é a quantidade de trabalho, K é a quantidade de capital, f é a função que descreve a relação entre insumos e produto.

Uma função de produção é dita ser homogênea de grau 1 se ela satisfaz a seguinte condição:

$$f(\lambda L, \lambda K) = \lambda f(L, K)$$

para qualquer fator de escala $\lambda > 0$. Isso significa que se você multiplicar todos os insumos por um fator λ, a quantidade de produto produzido será multiplicada pelo mesmo fator λ.

As principais características de funções homogêneas de grau 1:

1. Retornos Constantes de Escala: uma função de produção homogênea de grau 1 implica que a produção tem retornos constantes de escala. Isso significa que dobrar todos os insumos resultará em dobrar a produção. Por exemplo, se você dobrar a quantidade de trabalho e capital, a produção também dobrará.
2. Eficiência de escala: a propriedade de homogeneidade de grau 1 indica que a função de produção é eficiente em termos de escala. Em outras palavras, a função não apresenta aumentos ou diminuições desproporcionais na produção quando todos os insumos são aumentados ou reduzidos na mesma proporção.
3. Linearidade na escala: graficamente, uma função de produção homogênea de grau 1 apresenta uma superfície de produção que é uma expansão linear das entradas de produção.

Um exemplo simples de uma função de produção homogênea de grau 1 é a função Cobb-Douglas com parâmetros que somam 1:

$$Q = AL^{\alpha} K^{\beta}$$

em que A é uma constante, α e β são os coeficientes dos insumos trabalho e capital, respectivamente, e $\alpha + \beta = 1$.

Neste caso, se você aumentar tanto o trabalho quanto o capital em uma proporção λ, a produção aumentará na mesma proporção λ, mostrando que a função é homogênea de grau 1. Em resumo, uma função de produção homogênea de grau 1 reflete uma situação em que a produção é proporcional aos insumos utilizados, o que é uma característica importante em muitos modelos econômicos e teóricos de eficiência produtiva.

em que f_L e f_K são as proporções iniciais de trabalho e capital, respectivamente, nos custos totais de produção de X.

Em um setor caracterizado pela competição perfeita e por uma função de produção homogênea, a variação percentual na proporção em que dois fatores de produção são usados será igual à elasticidade de substituição (S) entre esses fatores vezes a variação percentual na proporção de seus preços. Assim, temos, para a indústria Y

$$\frac{d\left({}^{K_Y}/_{L_Y}\right)}{\left({}^{K_Y}/_{L_Y}\right)} = S_Y \frac{d\left({}^{P_K}/_{P_L}\right)}{\left({}^{P_K}/_{P_L}\right)} \tag{8.15}$$

Se escolhermos unidades de trabalho e capital para que seus preços iniciais sejam iguais à unidade, isso pode ser simplificado para

$$\frac{dK_Y}{K_Y} - \frac{dL_Y}{L_Y} = S_Y(dP_K - dP_L) \tag{8.16}$$

Podemos seguir um procedimento análogo para obter uma equação para a resposta dos fatores na indústria X, mas devemos perceber aqui que o retorno ao capital está sendo sujeito a um imposto em X, mas não em Y. Se dP_K for a alteração no preço do capital relevante para as decisões de produção na indústria Y, é claramente a mudança no preço do capital líquido de impostos. A mudança no preço do capital, incluindo o imposto, será $dP_K + T$, em que T é o valor do imposto por unidade de capital. A equação da resposta fatorial para X será, portanto,

$$\frac{dK_X}{K_X} - \frac{dL_X}{L_X} = S_X(dP_K + T - dP_L) \tag{8.17}$$

em que S_X é a elasticidade da substituição entre trabalho e capital na indústria X.

As quatro equações (8.13), (8.14), (8.16) e (8.17) contêm as seguintes nove incógnitas: dX, dP_X, dP_Y, dL_X, dL_Y, dK_X, dK_Y, dP_L e dP_K. Estes podem ser reduzidos para quatro pelo uso das cinco equações adicionais a seguir:

$$dK_Y = -dK_X \tag{8.18}$$

$$dL_Y = -dL_X \tag{8.19}$$

$$dP_X = f_L dP_L + f_K(dP_K + T) \tag{8.20}$$

$$dP_Y = g_L dP_L + g_K dP_K \tag{8.21}$$

$$dP_L = 0 \tag{8.22}$$

As equações (8.18) e (8.19) vêm do pressuposto de que a oferta de fatores é fixa: a quantidade de qualquer fator reduzida em uma das duas indústrias deve ser absorvida pela outra. As equações (8.20) e (8.21) provêm dos pressupostos de funções homogêneas de produção nos dois setores e da concorrência perfeita. Essas premissas garantem que os pagamentos por cada fator esgotam a receita total em cada setor. Assim, para a indústria Y, temos como uma aproximação de primeira ordem

$$P_Y dY + Y dP_Y = P_L dL_Y + L_Y dP_L + P_K dK_Y + K_Y dP_K \tag{8.23}$$

Como o produto marginal do trabalho em Y é $\frac{P_L}{P_Y}$ e o produto marginal do capital em Y é $\frac{P_K}{P_Y}$, também temos a seguinte aproximação de primeira ordem, $P_Y dY = P_L dL_Y + P_K dK_Y$. Usando esse fato, temos que $Y dP_Y = L_Y dP_L + K_Y dP_K$, que, dividindo por Y e lembrando que os preços iniciais de fatores e produtos são considerados igual a unidade, encontramos (8.21), em que g_L e g_K representam os *shares* iniciais de trabalho e capital, respectivamente, no produto da indústria Y. Um procedimento exatamente análogo aplicado à indústria X produz a equação (8.20); aqui, no entanto, deve-se ter em mente que a mudança no preço do capital, vista pelos empreendedores da indústria X, não é dP_K, mas $dP_K + T$.

A equação (8.22) é de uma variedade diferente das demais. As equações do modelo contêm variações absolutas de preços como variáveis, enquanto na teoria econômica subjacente são apenas os preços relativos que importam. Precisamos de algum tipo de numerário, cujo preço é expresso pelos outros preços, e a equação (8.22) expressa o preço do trabalho como esse numerário. Essa escolha não restringe a generalidade de nossos resultados. O governo invariavelmente ganhará $K_X T$ em receita tributária. Se o preço do capital, líquido de impostos, cair em $\frac{K_X T}{K_X + K_Y}$ como resultado do imposto, podemos concluir que o capital suporta todo o imposto. A mudança na renda nacional, medida em unidades do preço da mão-de-obra, é $K_X T + (K_X + K_Y) dP_K$, portanto o resultado assumido acima deixaria inalterada a parte do trabalho na renda nacional, enquanto a parte do capital cairia pelo valor ganho pelo governo. Se a solução de nossas equações nos dissesse que dP_K era zero, por outro lado, teríamos que concluir que trabalho e capital estavam pagando o imposto na proporção de suas contribuições iniciais para a renda nacional. Os preços relativos do trabalho e do capital (líquidos de impostos) permaneceriam os mesmos de antes, e, portanto, ambos os fatores teriam sofrido o mesmo declínio percentual na renda real como resultado do imposto. O caso em que o trabalho suporta toda a carga do imposto surge quando a variação percentual no preço líquido do capital (medido em unidades salariais) é igual à variação percentual na renda nacional (também medido em unidades salariais). Como dP_K já está em termos percentuais porque o preço inicial do capital é igual à unidade, essa condição pode ser escrita $dP_K = \frac{K_X T + (K_X + K_Y) dP_K}{L_X + L_Y + K_X + K_Y}$, que se reduz a $dP_K = \frac{K_X T}{L_X + L_Y}$. Assim, a escolha do preço da mão-de-obra como numerário de modo algum predestina a mão-de-obra a não suportar o ônus do imposto, como se poderia supor a princípio; de fato, essa suposição não restringe a solução do problema de incidência.

Substituindo as equações (8.18) a (8.22) em (8.13), (8.14), (8.16) e (8.17), obtemos:

$$\frac{dX}{X} = \varepsilon \left[f_K (dP_K + T) - g_K dP_K \right] \tag{8.24}$$

$$\frac{dX}{X} = f_L \frac{dL_X}{L_X} + f_K \frac{dK_X}{K_X} \tag{8.25}$$

$$\frac{K_X(-dK_X)}{K_Y K_X} - \frac{L_X(-dL_X)}{L_Y L_X} = S_Y dP_K \tag{8.26}$$

$$\frac{dK_X}{K_X} - \frac{dL_X}{L_X} = S_X(dP_K + T) \tag{8.27}$$

Igualando (8.24) e (8.25), obtemos um sistema 3×3:

$$\begin{cases} \varepsilon f_K T = \varepsilon(g_K - f_K)dP_K + f_L \dfrac{dL_X}{L_X} + f_K \dfrac{dK_X}{K_X} \\ \\ 0 = S_Y dP_K - \dfrac{L_X}{L_Y}\dfrac{dL_X}{L_X} + \dfrac{K_X}{K_Y}\dfrac{dK_X}{K_X} \\ \\ S_X T = -S_X dP_K - \dfrac{dL_X}{L_X} + \dfrac{dK_X}{K_X} \end{cases} \tag{8.28}$$

Escrevendo o sistema em formato matricial, temos:

$$\begin{pmatrix} \varepsilon(g_K - f_K) & f_L & f_K \\ \\ S_Y & -\dfrac{L_X}{L_Y} & \dfrac{K_X}{K_Y} \\ \\ -S_X & -1 & 1 \end{pmatrix} \begin{pmatrix} dP_K \\ \\ \dfrac{dL_X}{L_X} \\ \\ \dfrac{dK_X}{K_X} \end{pmatrix} = \begin{pmatrix} \varepsilon f_K T \\ \\ 0 \\ \\ S_X T \end{pmatrix} \tag{8.29}$$

Podemos usar a regra de Cramer para resolver para dP_K:

$$dP_K = \frac{\begin{vmatrix} \varepsilon f_K & f_L & f_K \\ 0 & -\dfrac{L_X}{L_Y} & \dfrac{K_X}{K_Y} \\ S_X & -1 & 1 \end{vmatrix}}{\begin{vmatrix} \varepsilon(g_K - f_K) & f_L & f_K \\ S_Y & -\dfrac{L_X}{L_Y} & \dfrac{K_X}{K_Y} \\ -S_X & -1 & 1 \end{vmatrix}} \cdot T \tag{8.30}$$

Podemos reescrever o problema como:

$$dP_K = \frac{\varepsilon f_K \left(\dfrac{K_X}{K_Y} - \dfrac{L_X}{L_Y}\right) + S_X \left(f_L \dfrac{K_X}{K_Y} + f_K \dfrac{L_X}{L_Y}\right)}{\varepsilon(g_K - f_K)\left(\dfrac{K_X}{K_Y} - \dfrac{L_X}{L_Y}\right) - S_Y - S_X \left(f_L \dfrac{K_X}{K_Y} + f_K \dfrac{L_X}{L_Y}\right)} \cdot T \tag{8.31}$$

Para resolver o determinante é utilizado o fato de que $(f_L + f_K) = 1$.

Antes de examinarmos algumas das implicações econômicas dessa solução, vamos estabelecer o fato de que o denominador é necessariamente positivo. S_X é necessariamente negativo; a expressão entre parênteses que multiplica no denominador é necessariamente positiva; e S_X é precedido por um sinal de menos; portanto, todo o terceiro termo no denominador é positivo. $(-S_Y)$ também é positivo. No primeiro termo, ε é negativo, de modo que, se for possível demonstrar que $(g_K - f_K)\left(\dfrac{K_X}{K_Y} - \dfrac{L_X}{L_Y}\right)$ é negativo ou zero, será estabelecido que todo o denominador é positivo (ou, no caso limite, zero). Se g_K for maior que f_K, a indústria Y consumirá mais capital do que a indústria X e $\left(\dfrac{K_X}{K_Y} - \dfrac{L_X}{L_Y}\right)$ deverá ser negativo; portanto, o produto indicado deve ser negativo. Da mesma forma, se $(g_K - f_K)$ for negativo, o setor X será o mais intensivo em capital dos dois setores e $\left(\dfrac{K_X}{K_Y} - \dfrac{L_X}{L_Y}\right)$ será positivo. Todo o primeiro termo no denominador é, portanto, positivo, e o denominador também.

Proposição 8.1: Utilização Relativa dos Insumos

Somente se a indústria tributada for relativamente mais intensiva em trabalho do que capital, o trabalho poderá suportar mais impostos, proporcionalmente à sua participação inicial na renda nacional.

Demonstração. Lembre-se de que quando dP_K é zero, trabalho e capital arcam com o imposto exatamente na proporção de suas participações iniciais na renda nacional. Para que o trabalho suporte mais do que isso, dP_K deve ser positivo. Como o denominador é positivo, o sinal de dP_K será determinado pelo sinal do numerador. O segundo termo no numerador é necessariamente negativo; portanto, dP_K pode ser positivo apenas se o primeiro termo for positivo e maior em magnitude absoluta que o segundo termo. Como ε é negativo, o primeiro termo pode ser positivo apenas se $\left(\dfrac{K_X}{K_Y} - \dfrac{L_X}{L_Y}\right)$ for negativo, e isso só poderá ocorrer se o setor X for relativamente mais intensivo em trabalho do que o setor Y. ■

Proposição 8.2: Elasticidade de Substituição e Elasticidade da Demanda

Se a elasticidade de substituição entre trabalho e capital (S_X) no setor tributado é tão grande ou maior em valor absoluto do que a elasticidade da demanda (ε) pelo produto do setor tributado, o capital deve arcar mais com o imposto do que com o trabalho, em relação às suas participações de renda iniciais .

Demonstração. Nesse caso, o termo $-\varepsilon f_K \dfrac{L_X}{L_Y}$, que é o único termo que pode dar ao numerador um sinal positivo, é dominado pelo termo $S_X f_K \dfrac{L_X}{L_Y}$. ∎

Proposição 8.3: Elasticidade de Substituição

Se a elasticidade de substituição entre trabalho e capital no setor tributado é tão grande em valor absoluto quanto a elasticidade de substituição entre os dois produtos finais, o capital deve suportar mais impostos do que mão-de-obra, em relação às suas participações iniciais nos rendimentos.

Demonstração. Isso vale do acima exposto, uma vez que a elasticidade de substituição entre X e Y deve ser maior em valor absoluto do que a elasticidade da demanda por X. A fórmula que relaciona a elasticidade de substituição entre X e Y, que denotarei por V, e a elasticidade da demanda por X, ε, é $\varepsilon = V\left(\dfrac{Y}{X+Y}\right)$. ∎

Proposição 8.4: Elasticidade de Substituição no Setor Não Tributado

Quanto maior a elasticidade de substituição entre trabalho e capital no setor não tributado, maior será a tendência do trabalho e do capital de arcar com o imposto proporcionalmente às suas quotas de renda iniciais.s

Demonstração. Essa elasticidade, S_Y, aparece apenas no denominador. Não altera o sinal, mas a magnitude da expressão para dP_K. Quanto maior for S_Y, em valor absoluto, menor será o valor absoluto de dP_K. No limite, em que S_Y é infinito, dP_K deve ser zero: nesse caso, os preços relativos do trabalho e do capital são determinados na indústria não tributada; o imposto não pode afetá-los. ∎

Proposição 8.5: Elasticidade de Substituição no Setor Tributado

Quanto maior a elasticidade de substituição entre trabalho e capital no setor tributado, mais próximas, *ceteris paribus*, será a taxa de retorno pós-imposto sobre o capital com a taxa inicial de retorno menos a taxa unitária aplicada ao capital na indústria X.

Demonstração. Essa elasticidade, S_X, aparece no numerador e no denominador de com coeficientes iguais, mas com sinais opostos. Quando S_X é infinito, e as outras elasticidades finitas, a expressão para dP_K é igual a $-T$. O preço do capital na indústria tributada deve, nesse caso, ter a mesma relação com o preço da mão de obra que existia no período anterior à situação tributária. O preço líquido do capital deve, portanto, cair no valor do imposto por unidade de capital em X. Como essa queda no preço se aplica ao capital empregado em Y e também em X, a redução na renda do capital deve exceder a quantidade de receita obtida pelo governo; a renda real do trabalho deve, portanto, aumentar. Quando S_X não é infinito, sua contribuição é mover o valor de dP_K em direção a $-T$. ■

Proposição 8.6: Elasticidade de Substituição e Proporção de Fatores

Quando as proporções de fatores são inicialmente as mesmas em ambas as indústrias, o capital suportará a carga total do imposto se as elasticidades de substituição entre trabalho e capital forem as mesmas em ambas as indústrias, suportará menos que a carga total do imposto se a elasticidade de substituição entre trabalho e capital é maior no setor não tributado do que no setor tributado e suportará mais do que toda a carga tributária se a elasticidade de substituição for maior no setor tributado

Demonstração. Quando $\frac{K_X}{K_L} = \frac{L_X}{L_K}$, os primeiros termos no numerador e no denominador desaparecem e a expressão é simplificada para $dP_K = -T\left(\dfrac{S_X K_X}{S_Y K_Y + S_X K_X}\right)$. Quando, adicionalmente, $S_X = S_Y$, isso se reduz a $-T\left(\dfrac{K_X}{K_Y + K_X}\right)$, que foi indicado anteriormente como a condição para o capital suportar exatamente a carga total do imposto. Quando S_X for maior que S_Y, a carga de capital será maior do que no caso em que as duas elasticidades são iguais e, inversamente. ■

Proposição 8.7: Elasticidade da Demanda da Mercadoria Tributada

Quando a elasticidade da demanda pela mercadoria tributada é zero, os resultados são um pouco semelhantes aos que acabamos de alcançar. Nesse caso, no entanto, o capital não suporta necessariamente toda a carga tributária, mesmo quando as elasticidades de substituição são iguais nas duas indústrias. É um pouco mais importante se a indústria tributada exige muito trabalho e um pouco menos se a indústria tributada exige muito capital.

Proposição 8.8: Elasticidade de Substituição nas Indústrias

Quando a elasticidade de substituição entre trabalho e capital é zero nas duas indústrias, a incidência do imposto dependerá unicamente das proporções relativas em que os fatores são usados nas duas indústrias, sendo que a mão-de-obra suportará o imposto mais do que na proporção da sua contribuição inicial à renda nacional quando o setor tributado é relativamente intensivo em trabalho e vice-versa.

Demonstração. Neste caso $dP_K = \dfrac{f_K T}{g_K - f_K}$, que será positivo quando g_K for maior que f_K (indústria tributada relativamente intensiva em mão-de-obra) e negativo quando f_K for maior que g_K (indústria tributada relativamente intensiva em capital). ■

Exemplo 8.1. *Suponha que estimamos os seguintes dados:*

Tabela 8.1 – *Valores iniciais*

	$\frac{K_X}{K_Y}$	$\frac{L_X}{L_Y}$	f_K	f_L	g_K
Conjunto 1	1	10	$\frac{1}{11}$	$\frac{10}{11}$	0,5
Conjunto 2	2	10	$\frac{1}{6}$	$\frac{5}{6}$	0,5

Substituindo na expressão para dP_K, obtemos:

$$\begin{aligned} dP_K^I &= \frac{\varepsilon\left(\frac{1}{11}\right)(1-10)+S_X\left(1\cdot\frac{10}{11}+10\cdot\frac{1}{11}\right)}{\varepsilon\left(\frac{1}{2}-\frac{1}{11}\right)(1-10)-S_Y-S_X\left(1\cdot\frac{10}{11}+10\cdot\frac{1}{11}\right)}\cdot T \\ &= \frac{T(-9\varepsilon+20S_X)}{-40,5\varepsilon-11S_Y-20S_X} \qquad (8.32) \\ dP_K^{II} &= \frac{\varepsilon\left(\frac{1}{6}\right)(2-10)+S_X\left(1\cdot\frac{5}{6}+10\cdot\frac{1}{6}\right)}{\varepsilon\left(\frac{1}{2}-\frac{1}{6}\right)(2-10)-S_Y-S_X\left(1\cdot\frac{5}{6}+10\cdot\frac{1}{6}\right)}\cdot T \\ &= \frac{T(-8\varepsilon+20S_X)}{-16\varepsilon-6S_Y-20S_X} \qquad (8.33) \end{aligned}$$

A incidência final depende de todo o conjunto de elasticidades.

Assim, sendo um aumento no imposto sobre a renda corporativa é possível depreender um efeito substituição e um efeito produto.

O efeito substituição implica que o capital suporta o imposto:

- depende da elasticidade de substituição entre capital e trabalho
- o imposto desloca a produção no setor X para longe do capital
- a demanda agregada por capital diminui

- como o capital é fixo, r diminui
- o capital suporta o ônus do imposto

O efeito produto implica que o capital pode não suportar o imposto:

- deslocamento da demanda para o outro setor Y
- a consequência para a demanda por fatores depende das intensidades relativas dos fatores
  - se X é intensivo em capital: reduz a demanda por capital e o capital suporta mais do imposto
  - se X é intensivo em trabalho: aumenta a demanda por capital e o trabalho pode suportar parte ou todo o imposto

Como consequência, a composição dos efeitos implica *overshifting*. Assim, se o setor corporativo é intensivo em capital, pode levar a uma incidência superior a 100% (sobrecarga). Porém, se o setor corporativo é intensivo em trabalho, pode levar a toda a incidência sobre o trabalho.

8.5 Extensões

Talvez o resultado mais proeminente dos modelos dinâmicos de tributação ótima seja que a tributação da renda do capital deve ser evitada. Esse resultado, controverso desde o início em meados da década de 1980, foi modificado de algumas maneiras sutis e desafiado diretamente em outras, mas sua forte lógica subjacente a tornou a referência.

A intuição para um imposto zero sobre o capital pode ser desenvolvida de várias maneiras. Primeiro, como o equipamento de capital é um insumo intermediário para a produção do produto futuro, o resultado de Diamond e Mirrlees (1971) sugere que ele não deve ser tributado. Segundo, porque um imposto sobre capital é efetivamente um imposto sobre o consumo futuro, mas não sobre o consumo atual, viola a prescrição de Atkinson e Stiglitz (1976) para tributação uniforme. De fato, um imposto sobre o capital impõe um imposto cada vez maior ao consumo ainda mais no futuro, portanto sua violação do princípio da tributação uniforme das mercadorias é extrema.

Uma terceira intuição para um imposto zero sobre o capital deriva da elaboração do problema tributário considerado por Frank Ramsey (1928). Em artigos importantes, Chamley (1986) e Judd (1985) examinam a tributação ótima do capital nesse modelo. Eles descobriram que, no curto prazo, um imposto positivo sobre o capital pode ser desejável, porque é um imposto sobre o capital antigo e, portanto, não é distorcido. A longo prazo, no entanto, um imposto zero sobre o capital é ideal. No modelo de Ramsey, pelo menos algumas famílias são modeladas como tendo um horizonte infinito de planejamento (por exemplo, podem ser dinastias cujas gerações são altruisticamente conectadas como em Barro, 1974). Essas famílias determinam quanto economizar com base no desconto do futuro e no retorno ao capital na economia. No equilíbrio de longo prazo, suas decisões de poupança são perfeitamente elásticas em relação à taxa de retorno após os impostos. Assim, qualquer imposto sobre a renda do capital deixará inalterado o retorno após o imposto para o capital, o que significa que o retorno antes do imposto ao capital deve aumentar, reduzindo o tamanho do estoque de capital e a produção agregada na economia. Essa distorção é tão grande que torna qualquer tributação sobre a renda do capital subótima em comparação com a tributação sobre a renda do trabalho, mesmo na perspectiva de um indivíduo sem poupança (Mankiw, 2000). Essa mensagem é fortalecida na economia moderna pela crescente globalização do mercado de capitais, que pode levar a respostas altamente elásticas dos fluxos de capital às mudanças tributárias, mesmo no curto prazo.

Pode-se encontrar razões para questionar o imposto zero sobre o capital. Se todos os indivíduos tiverem horizontes de planejamento relativamente curtos, como nos modelos de gerações sobrepostas, a tributação do capital poderá fornecer redistribuição sem os efeitos dramáticos na acumulação de capital identificados na literatura de Ramsey. Alternativamente, se os indivíduos acumulam poupança para se proteger contra choques, pode haver superacumulação agregada de capital, justificando a tributação do capital. Apesar dessas possíveis exceções, no entanto, a lógica para baixos impostos sobre o capital é poderosa: a oferta de capital é altamente elástica, os impostos sobre o capital produzem grandes distorções nos planos de consumo intertemporais e desencorajam a economia, e a acumulação de capital é central para a produção agregada da economia.

VICTOR OLIVEIRA

9 Federalismo Fiscal

Economics studies the competitive and the cooperative behavior of people in resolving conflicts of interests that arise because wants exceed what is available.

ALCHIAN & ALLEN, 1972

9.1 Introdução

O federalismo fiscal é a divisão das responsabilidades de arrecadação e despesa entre os diferentes níveis de governo. A maioria dos países tem um governo central (ou federal), governos estaduais ou municipais, conselhos municipais e, no nível mais baixo, conselhos. Juntos, eles constituem a administração multinível e sobreposta que governa um país desenvolvido típico.

O governo central geralmente pode escolher quaisquer instrumentos tributários que desejar e, embora tenha liberdade em seus gastos, geralmente se concentra na defesa nacional, no fornecimento de lei e ordem, infraestrutura e pagamentos de transferências.

Os poderes tributários dos governos estaduais são mais restritos. No Reino Unido, os governos locais podem cobrar apenas impostos sobre a propriedade. Nos Estados Unidos e no Brasil, são permitidos impostos sobre mercadorias e impostos locais. Os níveis de governo estão conectados pela sobreposição de responsabilidades e pelos pagamentos de transferência feitos entre eles.

Os fundamentos teóricos modernos da descentralização têm origem em Oates (1972). O teorema da descentralização postula que, assumindo que não há economia de custos da centralização, o bem-estar agregado em um conjunto de jurisdições será superior quando cada jurisdição puder selecionar seu próprio pacote de consumo público em oposição a quando o consumo uniforme for fornecido em todas as jurisdições. Obviamente, esta simples declaração do teorema não leva em conta a perda de economias de escala ou uma miríade de outros fatores que podem intervir à medida que a descentralização fiscal aumenta.

O federalismo fiscal é apontado pela literatura como um instrumento capaz de conter a tendência natural dos governos a crescer. A existência de vários governos dentro de uma nação, competindo pela atração de eleitores e investimentos, levaria à redução da carga tributária e à maior eficiência nos serviços públicos, resultando em governos menores. A redução do tamanho do setor público não é, contudo, uma consequência natural da adoção do sistema federativo. Buchanan (1995) e (1996) vê os sistemas de governo como uma distribuição contínua, que vai desde a extrema centralização (governo unitário) até à extrema descentralização (confederação ou tratado de cooperação entre governos autônomos).

O federalismo capaz de gerar competição e equilíbrio de forças políticas estaria no centro desta distribuição, estabelecendo, além da competição fiscal horizontal (entre governos subnacionais), um sistema de balanço de poder vertical (governo central versus governos locais), em que cada nível de governo policia o outro, para não ter seus poderes usurpados.

Nada impede, porém, que se tenha um federalismo desequilibrado, situado em um ponto da distribuição mais próximo do governo unitário (federalismo com centralização) ou com excessiva descentralização. Nesses casos é possível que a organização federativa leve ao crescimento do governo.

Primeiro, por que deveria haver mais de um nível de governo? Usando a lógica do raciocínio econômico, o governo multinível só pode ser justificado se puder alcançar algo que um único nível não pode. As explicações sobre o que o governo multinível pode alcançar devem girar em torno do acesso à informação e como a informação pode ser melhor utilizada. Se esse argumento for aceito, surge uma segunda questão. Como as funções do governo são melhor alocadas entre os níveis? Um breve esboço de como essa alocação funciona na prática já foi dado: esse resultado é eficiente ou reflete algum outro fator?

Os argumentos econômicos para a existência do governo se baseiam em dois princípios:

1. se houver falha de mercado, o governo pode intervir na economia para aumentar a eficiência.

2. também pode intervir para melhorar a equidade, independentemente de a eco-

nomia ser eficiente.

Esses argumentos justificam a intervenção; para justificar o governo multinível, deve-se argumentar que os objetivos de eficiência e equidade são melhor atendidos por uma combinação de governo local e central.

Se as decisões corretas forem tomadas sobre o nível de provisão de bem público e sobre os impostos, não importa em que nível de governo elas sejam tomadas. Desde que não haja desperdício de recursos na sobreposição de responsabilidades, o número de níveis de governo é indiferente. Um caso de governo multinível deve, portanto, ser buscado em diferenças na informação e no processo político que permitem que algumas estruturas alcancem melhores resultados do que outras.

As decisões devem ser tomadas em nível nacional se envolverem bens públicos que sirvam a toda a economia. O exemplo óbvio aqui seria a defesa, cujos benefícios não podem ser atribuídos a nenhuma comunidade específica dentro do país. É comum argumentar que todos os cidadãos devem ter o mesmo acesso à lei, ter os mesmos direitos perante a lei e estar sujeitos às mesmas restrições.

Uma aplicação desse argumento de equidade dá suporte a um sistema jurídico que é organizado e administrado a partir do centro. Dada a prestação central desses serviços, é natural apoiá-los por meio de impostos organizados centralmente.

Os outros bens públicos, nomeadamente os bens públicos locais, beneficiam apenas os residentes numa área geográfica definida.

O nível de oferta desses bens poderia ser determinado e financiado em nível nacional, mas há três argumentos que sugerem que uma decisão de nível subnacional é preferível.

1. Primeiro, a determinação no nível local pode levar em conta informações mais precisas disponíveis sobre as preferências locais. Nesse contexto, o conhecimento das circunstâncias locais podem ajudar a chegar a uma decisão mais eficiente.

2. Em segundo lugar, se uma decisão for tomada em nível nacional, as pressões

políticas podem impedir que haja qualquer diferenciação de provisão entre as localidades, ao passo que pode ser eficiente ter diferentes níveis de provisão de bens públicos locais em diferentes áreas.

3. Finalmente, se a hipótese de Tiebout, de que os consumidores têm preferências heterogêneas, é válida, então a eficiência requer que numerosas localidades se formem e ofereçam diferentes níveis de provisão de bens públicos. Isso não será possível se as decisões forem tomadas em nível nacional.

9.2 Teorias de Competição Fiscal

Há diferentes explicações para o fenômeno do federalismo fiscal.

1. Tiebout (1956)

    - Plena mobilidade dos indivíduos
    - Pleno conhecimento
    - Número atomizado de governos locais
    - Não existem efeitos *spillovers* espaciais
    - A população local possui um tamanho ótimo
    - Comportamento governamental: governo local compete por indivíduos usando pacotes de serviços e impostos
    - Resultados: oferta de bens públicos eficiente e impostos são benéficos

2. Beck (1983)

    - Trabalho não tem mobilidade
    - Capital tem mobilidade perfeita
    - Dois bens públicos (para as famílias e as firmas)
    - Os impostos incidem sobre o capital industrial e das famílias (propriedade)

- Comportamento governamental: os governos locais competem por investimentos via redução de impostos sobre o capital e maximizam a utilidade das famílias em sua jurisdição
- Resultados: oferta ineficiente de bens públicos (sub-provisão), capital é sub-tributado e famílias super-tributadas

3. Mintz e Tulkens (1986)

- Duas jurisdições
- Três bens (público, privado e trabalho)
- Tributação de bens privados pelo princípio da origem
- Trabalho não é tributado
- Não existem efeitos *spillovers* espaciais
- Os preços dos bens são afetados pelos impostos locais e custos de transporte
- Comportamento governamental: as estratégias competitivas dos governos seguem um jogo não cooperativo (equilíbrio de Nash), onde os instrumentos são a política tributária e de gastos; os governos regionais buscam maximizar o bem-estar regional
- Resultados: o equilíbrio de Nash existe sob determinadas condições, mas não é um ótimo de Pareto, e a oferta de bens públicos é ineficiente

4. Zodrow e Mieszkowski (1986)

- Dois bens públicos (para famílias e firmas), sem efeitos *spillovers*
- Dois impostos (sobre capital e impostos per capita neutros), sem externalidade fiscal
- Governos locais atomizados e independentes
- Terra e capital são os fatores de produção

- A população e o fator terra (fixo) são distribuídos simetricamente entre as regiões
- Mobilidade perfeita do capital
- Comportamento governamental: a competição interjurisdicional segue um jogo não cooperativo. Os governos maximizam o bem-estar regional, consideram que as outras jurisdições não reagem ao imposto sobre a propriedade e que suas ações não afetam a taxa líquida de retorno do capital
- Resultados: oferta local de bens públicos ineficiente, e existe uma relação inversa entre a variação do imposto sobre o capital e a variação dos *head taxes*

5. Wilson (1986)

- Número elevado de regiões
- Dois fatores primários (capital e trabalho)
- Capital possui mobilidade perfeita e trabalho é imóvel
- Capital e trabalho possuem perfeita mobilidade intersetorial, mas a oferta nacional é fixa
- Um bem privado nacional comercializável, um bem privado e um bem público, ambos não comercializáveis
- Um imposto sobre a propriedade (capital)
- Tecnologia de produção regional idêntica e população regional simétrica
- Comportamento governamental: governo local define o imposto sobre a propriedade e a demanda dos fatores de produção que maximiza o bem-estar regional. A competição tributária é definida como uma situação em que o governo federal poderia aumentar o bem-estar regional exigindo uma pequena elevação uniforme na oferta de bens públicos locais (definida pela ineficiência na oferta de bens públicos)

- Resultados: a oferta local de bens públicos e a relação capital-trabalho na produção pública são ineficientes. A competição tributária existe se e somente se a produção pública é trabalho-intensiva em relação à produção privada e a elasticidade substituição entre trabalho e capital na produção privada excede um limite inferior menor que 1.

6. Wildasin (1988)

- Duas ou mais regiões
- Um bem privado homogêneo e um bem público local
- A função de produção usa capital e outro fator local fixo (terra ou trabalho)
- O bem privado pode ser consumido ou usado como insumo na produção do bem público local
- A oferta do bem público é financiada por um imposto sobre o capital
- Capital possui mobilidade perfeita, mas a oferta nacional de capital é fixa
- Comportamento governamental: os governos locais engajam-se em estratégias não cooperativas e maximizam o bem-estar regional. As variáveis estratégicas são os impostos e os gastos, analisados separadamente: equilíbrio de Nash para competição através de impostos (ENCI) e para competição através de gastos (ENCG).
- Resultados: quando o número de jurisdições é pequeno, os efeitos de bem estar de ENCI e ENCG não são iguais e ENCI gera provisão de bens públicos superior a ENCG. Os efeitos de bem-estar de ENCI e ENCG são iguais somente quando o número de jurisdições é muito elevado.

7. Oates e Schwab (1988)

- Existem grandes regiões

- Um bem privado e um mal privado (poluição)
- Capital tem mobilidade perfeita
- Trabalho não se move
- A poluição não gera efeitos *spillover*
- Os indivíduos consomem o bem privado e a poluição
- A política da jurisdição envolve imposto sobre o capital e legislação ambiental
- Comportamento governamental: as regiões competem via redução de impostos sobre o capital ou atenuação da legislação ambiental. Os governos maximizam bem-estar regional considerando uma abordagem de escolha pública (modelo do eleitor mediano).
- Resultados: a competição inter-jurisdicional é eficiente quando o imposto sobre o capital é nulo (*race-to-the-bottom*) e a legislação é dura. O resultado não é robusto quando o imposto é positivo devido alguma ineficiência governamental, quando o governo busca auto-benefício e quando há conflito de interesse entre os eleitores.

8. Burbidge e Myers (1994)

   - Duas regiões
   - Terra é um fator fixo
   - Capital tem mobilidade perfeita
   - População tem mobilidade imperfeita
   - Os indivíduos são heterogêneos na preferência por uma residência local
   - Um bem privado e um bem público
   - Oferta de trabalho é inelástica

- Três impostos (terra, capital e impostos per capita)
- As regiões podem fazer transferências fiscais
- Comportamento governamental: os governos engajam-se em estratégias não cooperativas (equilíbrio de Nash) via política tributária e maximizam o bem estar individual para os bens privado e público. A análise considera três regimes tributários: (i) imposto direto, (ii) imposto direto e sobre o capital e (iii) imposto sobre o capital.
- Resultados: o equilíbrio do regime 1 é eficiente, independente do grau de mobilidade populacional. O equilíbrio do regime 2 é ineficiente, exceto quando as regiões são simétricas ou quando há alta mobilidade populacional. O equilíbrio do regime 3 geralmente é ineficiente. As transferências podem controlar a migração e neutralizar as externalidades fiscais.

9. Rauscher (1998)

- Grande número de pequenas jurisdições
- Três fatores de produção (terra, capital e trabalho)
- Um bem privado agregado e um bem público
- Capital tem mobilidade perfeita
- Um imposto *lump-sum* e uma taxa sobre o bem público
- Comportamento governamental: os governos são como um Leviatã[44], buscam maximizar o bem estar dos cidadãos mas também seu próprio bem-

[44] Leviatã é uma obra seminal na filosofia política escrita por Thomas Hobbes, publicada pela primeira vez em 1651. O livro é uma das principais contribuições ao pensamento político e social do período moderno e continua a ser estudado e discutido até hoje. Aqui estão alguns pontos-chave sobre o Leviatã: Thomas Hobbes (1588–1679) foi um filósofo inglês que viveu durante um período de grande agitação política na Inglaterra, incluindo a Guerra Civil Inglesa. Sua experiência e observações sobre a natureza humana e o caos político influenciaram profundamente sua obra. Os temas principais incluem:

(a) Natureza Humana: Hobbes tem uma visão bastante pessimista da natureza humana. Ele acredita que, sem um governo forte para manter a ordem, os seres humanos agiriam de maneira egoísta e conflituosa. Segundo Hobbes, o estado de natureza, ou a condição préestar se apropriando de parte da receita tributária. Dois cenários: o governo oferece um pacote de imposto e gastos ou fixa os impostos e taxas e deixa o mercado decidir sobre a demanda.

- Resultados: no primeiro cenário, a oferta de bens públicos é eficiente, os impostos são maiores que no equilíbrio sem o governo Leviatã, o governo se apropria de parte da receita e sua decisão independe da mobilidade do capital. No segundo, a mobilidade gera taxas menores e maior oferta do bem público (doma-se o Leviatã).

10. Cardarelli, Taugourdeau e Vidal (2002)

- Duas regiões
- O consumidor representativo tem dois períodos de vida

---

social, é um "estado de guerra de todos contra todos" (*bellum omnium contra omnes*), onde a vida é "solitária, pobre, desagradável, brutal e curta".

(b) Contrato Social: a solução para o estado de natureza, segundo Hobbes, é o estabelecimento de um contrato social. No contrato social, os indivíduos concordam em ceder certos direitos e poderes a um soberano absoluto em troca de segurança e ordem. Esse soberano é chamado de Leviatã, uma metáfora para um poder soberano forte e indomável que mantém a ordem e evita o caos.

(c) Soberania Absoluta: o Leviatã representa a autoridade suprema e centralizada. Hobbes argumenta que para evitar o caos, o soberano deve ter poder absoluto e inquestionável. A soberania não deve ser dividida ou contestada, pois isso poderia levar à desordem e à fragmentação social.

(d) Natureza do Poder: Hobbes explora a natureza do poder e da autoridade, discutindo como o poder é obtido e mantido. Ele vê o poder como essencial para a criação e manutenção de uma sociedade estável.

(e) Crítica ao Direito Natural: embora Hobbes se baseie em princípios de direito natural para justificar a necessidade do Leviatã, ele critica outros pensadores que acreditam na possibilidade de uma lei moral ou um estado de natureza ideal. Para Hobbes, a moralidade é uma construção social que surge do contrato social.

O Leviatã teve um impacto profundo na teoria política moderna. A ideia de que o poder absoluto é necessário para evitar o caos influenciou muitos pensadores e políticos, e o conceito de contrato social foi fundamental para o desenvolvimento das teorias políticas subsequentes. A visão de Hobbes sobre o absolutismo e o papel do estado tem sido amplamente debatida. Críticos argumentam que sua visão é excessivamente pessimista e que o poder absoluto pode levar à tirania. Pensadores como John Locke e Jean-Jacques Rousseau ofereceram visões alternativas sobre o contrato social e o papel do governo. O Leviatã é estudado não apenas por seu conteúdo político, mas também por suas implicações para a filosofia social e a teoria do poder. Suas ideias sobre a natureza humana e a necessidade de ordem ainda são relevantes em debates contemporâneos sobre política e governança.

- Capital tem perfeita mobilidade
- Um bem público
- Um bem privado que pode ser armazenado para consumo futuro
- A dotação de riqueza pode ser investida na região doméstica ou externa
- A taxa de juros é nula
- Tributação da renda pelo princípio da origem
- O governo garante a alíquota tributária para o período $t+1$ anunciada em t
- Os consumidores têm previsão perfeita
- O custo de mobilidade do investimento externo é quadrático
- Comportamento governamental: os governos agem estrategicamente e maximizam o valor presente da função de bem-estar regional. Analisa os resultados de estratégias ótimas definidas a partir de um jogo *one-shot* ou com repetição, pelo qual os governos cooperam.
- Resultados: no equilíbrio de Nash (*one-shot game*), a oferta de bens é ineficiente, sejam as regiões simétricas ou assimétricas. Existe harmonização tributária somente quando as regiões são simétricas. O equilíbrio do jogo com repetição pode gerar harmonização e eficiência, mas sob a restrição de simetria regional (dotações e preferência por bens públicos) e paciência dos governos.

11. Cassette, Jayet e Paty (2005)

    - Poucas regiões
    - Os residentes são homogêneos e imóveis
    - Um bem privado e um bem público
    - Existe imposto sobre o capital e a renda privada

- O governo desvia parte da receita para seu próprio bem-estar
- Existe processo eleitoral
- A probabilidade de reeleição é função crescente da utilidade dos residentes
- Capital é perfeitamente móvel
- Comportamento governamental : os governos são como um Leviatã. Analisa o equilíbrio em diversos ambientes, inclusive num jogo não cooperativo.
- Resultados: no equilíbrio de Nash, a oferta de bens públicos é eficiente. O imposto sobre o capital é positivo apenas se a região atrai mais capital que a quantidade de capital dos residentes (exportação de imposto). Os resultados dependem da efetividade dos impostos *lump-sum*.

9.3 Uniformidade

A provisão uniforme de bens e serviços públicos por todas as jurisdições só atenderá exatamente às necessidades de toda a população quando as preferências forem homogêneas.

Quando não o são, qualquer forma de provisão uniforme deve ser um meio-termo entre níveis concorrentes de demanda. Como tal, deve envolver alguma perda de bem-estar em relação à provisão diferenciada.

Esse argumento pode ser ilustrado considerando uma economia em que há dois grupos de consumidores com gostos diferentes pelo bem público único da economia.

O bem público é financiado por um imposto de renda uniforme.

Denomine os dois grupos por A e B e suponha que os membros do grupo B tenham uma preferência relativamente mais forte pelo bem público do que os do grupo A, levando em consideração a taxa de imposto mais alta que isso implica.

As escolhas preferenciais de provisão de bens públicos são indicadas como G_A e G_B (com $G_A < G_B$).

Agora considere a escolha de um nível uniforme de provisão, e seja esse nível G_0.

Suponha que esse nível esteja entre G_A e G_B (o argumento se estende facilmente a casos em que está fora desses limites).

A perda de bem-estar para a sociedade é então dada por
$L = n_A\,[u_A(G_A) - u_A(G_0)] + n_B[u_B(G_B) - u_B(G_0)]$, em que n é o tamanho do grupo, em comparação com o que seria alcançado se cada grupo pudesse ser abastecido com sua quantidade preferida.

O valor da perda pode ser minimizado definindo a localização de G_0 de modo que o benefício marginal para o grupo B de ter mais bem público, $n_B u'_B(G_0) > 0$, apenas compensa a perda marginal do grupo, A, $n_A u_A(G_0) < 0$, mas o ponto essencial é que a perda continue positiva.

Além disso, a perda aumenta quanto mais dispersas forem as preferências e quanto mais membros houver em cada grupo.

Esta análise mostra como a uniformidade pode custar caro em termos de perda de bem-estar.

Uma política de uniformidade só pode ser mantida se os custos de diferenciação excederem o benefício.

Tais custos podem surgir na coleta de informações para determinar a diferenciação e nos custos de administração de um sistema diferenciado.

9.4 Hipótese de Tiebout

Embora os custos da uniformização ilustrados acima sejam indiscutíveis, é mais um passo para mostrar que a descentralização é justificada.

O caminho para fazer isso é explorar a hipótese de Tiebout.

Cada comunidade pode ser tratada como fornecedora independente de bens públicos locais.

Se os consumidores na economia tiverem gostos heterogêneos, haverá claras vantagens para jurisdições com diferentes níveis de provisão.

Cada um pode projetar o que oferece (suas taxas de imposto, nível de provisão e tipo de provisão) para atrair grupos específicos dentro da sociedade.

Ao escolher a jurisdição em que viver (ou seja, ao “votar com os pés”), os consumidores revelam seus gostos por bens públicos.

Na ausência de custos de transação ou outros impedimentos à liberdade de movimento, um equilíbrio eficiente deve ocorrer.

Os limites desse argumento explorado no contexto dos bens públicos locais também são aplicáveis aqui.

Os custos de transação são relevantes na prática, e surgirá o problema de dividir de forma otimizada uma população finita em um número limitado de jurisdições.

O fato de que a primeira melhor alocação não será alcançada não enfraquece necessariamente o argumento de correspondência de preferência para a descentralização.

Claramente ainda há benefícios na descentralização, mesmo quando ela não pode ser levada ao nível exigido pela hipótese de Tiebout.

Partindo de um nível uniforme de serviços que é muito pouco para alguns consumidores e muito para outros, um afastamento desse nível uniforme por parte de algumas jurisdições deve beneficiar alguns dos consumidores.

Dessa forma, mesmo a descentralização restrita pode aumentar em eficiência.

9.5 Estrutura Ótima: Eficiência versus Estabilidade

Os argumentos anteriores exploraram uma série de vantagens da descentralização fiscal.

Estes envolveram aspectos de eficiência e equidade.

A questão que permanece é qual é a estrutura ideal ou o número correto de níveis de administração.

A dificuldade que surge aqui é que a divisão ótima pode diferir de um bem público para outro.

Serviços de bombeiros são organizados em um nível muito local, a educação em um nível acima e a defesa em um nível ainda mais alto.

Existem muitos outros bens públicos fornecidos pelo governo federal.

Se cada um fosse alocado no nível correto de descentralização, isso implicaria um número igualmente grande de níveis de governo.

Esse argumento agora é ilustrado em um modelo espacial simples que troca economias de escala por diversidade de preferência.

O ponto de partida é que a tomada de decisão centralizada produz um resultado de "tamanho único" que não reflete a heterogeneidade de gostos.

A provisão uniforme decorre de considerações de economia política que impedem que a votação por maioria centralizada aloque níveis diferentes de bens públicos a localidades diferentes.

É apenas descentralizando a votação majoritária no nível distrital que é possível diferenciar a provisão de bem público, mas com algum custo de duplicação.

Suponha que haja um bem público que pode ser fornecido tanto no nível federal quanto no nível regional.

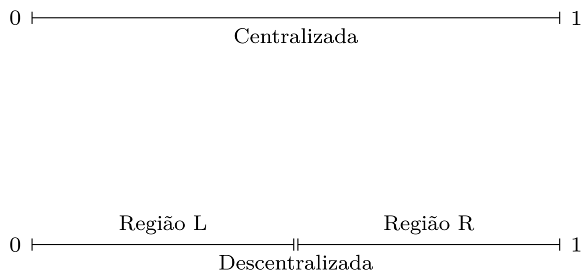
Modelamos a federação como o segmento de linha $[0, 1]$, com pontos na linha representando diferentes localizações geográficas.

O bem público pode estar localizado em qualquer lugar ao longo da linha, e os indivíduos são caracterizados por sua localização ideal para o bem público.

Com provisão central, o bem público está localizado em $\frac{1}{2}$, o ponto médio do segmento de linha.

Quanto mais longe desse ponto os indivíduos estiverem localizados, menos eles gostarão do bem público provido no nível federal. Isso é mostrado na Figura 9.1.

Figura 9.1 – *Centralização e Descentralização*



Alternativamente, a provisão pode ser descentralizada.

Cada região é então representada por um intervalo no segmento de linha.

A região L é o intervalo esquerdo $\left[0, \frac{1}{2}\right]$ e região R o intervalo direito $\left[\frac{1}{2}, 1\right]$.

Supõe-se que os indivíduos estejam distribuídos uniformemente, de modo que ambas as regiões tenham o mesmo tamanho.

Com provisão descentralizada, o bem público está localizado no ponto médio de cada intervalo, ou seja, $\frac{1}{4}$ na região esquerda e $\frac{3}{4}$ na região direita.

Há um custo fixo C (per capita) de fornecer o bem público ao nível central e, devido à duplicação, o custo é 2C com provisão descentralizada (ou seja, o número de indivíduos pelos quais o custo da provisão de bem público é distribuído é reduzido pela metade).

A função de utilidade do indivíduo i sob centralização é

$$u_i^c = 1 - \alpha \left| \frac{1}{2} - i \right| - C \tag{9.1}$$

A função de utilidade do indivíduo i sob descentralização é

$$u_i^d = \begin{cases} 1 - \alpha \left| \dfrac{1}{4} - i \right| - 2C, \text{ se } i \in \left[0, \dfrac{1}{2}\right] \\ \\ 1 - \alpha \left| \dfrac{3}{4} - i \right| - 2C, \text{ se } i \in \left[\dfrac{1}{2}, 1\right] \end{cases} \tag{9.2}$$

$\left|\dfrac{1}{2} - i\right|$, $\left|\dfrac{1}{4} - i\right|$ e $\left|\dfrac{3}{4} - i\right|$ denotam a distância entre a localização do bem público e a localização ideal do indivíduo i.

A taxa na qual a utilidade diminui com a distância é dada pelo parâmetro α.

Podemos definir quando a descentralização é socialmente ótima considerando o *trade-off* entre duplicar o custo da provisão do bem público contra tomar a provisão mais próxima das preferências individuais.

A solução socialmente ótima maximiza a soma de todas as utilidades individuais.

Como as utilidades individuais diferem apenas na distância até a localização do bem público (devido ao compartilhamento igual de custos), a soma das utilidades dependerá da distância média.

Sob provisão centralizada, a distância média de $\dfrac{1}{2}$ é, devido à distribuição uniforme, exatamente igual a $\dfrac{1}{4}$.

Observe que essa distância é realmente minimizada ao localizar o bem público no ponto médio $\dfrac{1}{2}$.

A descentralização reduz a distância média em qualquer região para $\dfrac{1}{8}$.

Esse efeito positivo da descentralização deve ser contrabalançado com o custo extra C de fornecer o bem público duas vezes.

Portanto, a descentralização é a solução ótima se e somente se o custo extra C for menor que a vantagem de reduzir a distância média por $\dfrac{1}{4} - \dfrac{1}{8}$ avaliada na taxa a na qual a utilidade diminui com a distância; isto é $C \leq \alpha \left[\dfrac{1}{4} - \dfrac{1}{8}\right]$.

Em resumo, a descentralização é ótima se e somente se $C \leq \frac{\alpha}{8}$.

Esses argumentos mostram que a quantidade ideal de descentralização será alcançada quando os benefícios de uma maior descentralização, em termos de combinar a diversidade de gostos, superarem o custo de diferenciar a provisão de bem público.

Na prática, a extensão da descentralização é determinada pelo processo político.

Usando esse modelo básico, podemos agora ilustrar a tendência de a votação por maioria levar à descentralização excessiva.

Para observar o incentivo à descentralização na votação por maioria, assumimos que a provisão descentralizada prevalece quando a maioria dos eleitores é favorável em pelo menos uma região.

Essa suposição é inócua dada a simetria entre as regiões: se há maioria a favor da descentralização em uma região, também deve haver maioria equivalente na outra região.

Concentramo-nos no incentivo da esquerda para a descentralização.

A maioria na região-esquerda é formada por aqueles que estão à esquerda ou à direita do indivíduo no ponto médio regional $\frac{1}{4}$.

É fácil ver que, se esse indivíduo central prefere a descentralização, todos os que estão à sua esquerda também têm as mesmas preferências porque estão localizados mais longe da provisão centralizada, mas compartilham o custo igualmente.

Portanto, haverá uma maioria a favor da descentralização na região esquerda se o indivíduo decisivo $i = \frac{1}{4}$ prefere descentralização, isto é, se

$$u_i^c = 1 - \alpha \left|\frac{1}{2} - i\right| - C \leq u_i^d = 1 - \alpha \left|\frac{1}{4} - \frac{1}{4}\right| - 2C \tag{9.3}$$

Portanto, pode-se concluir que a descentralização é um equilíbrio de votação majoritária se e somente se $C \leq \frac{\alpha}{4}$.

Este resultado sugere descentralização excessiva sob votação por maioria porque o nível de custo crítico sob votação por maioria é maior do que o nível de custo crítico para otimalidade.

Em particular, para qualquer custo C entre $\frac{\alpha}{8}$ e $\frac{\alpha}{4}$, a votação por maioria leva à descentralização, $C < \frac{\alpha}{4}$, embora não seja socialmente ótima, $C > \frac{\alpha}{8}$.

Portanto, sob votação majoritária, há descentralização excessiva: os eleitores localizados nos extremos têm um incentivo para apoiar a provisão descentralizada para obter um bem público mais próximo do que desejam, mas o processo democrático não internaliza as externalidades negativas impostas aos eleitores localizados no centro que sofrem com o custo extra, e pouca ou pior combinação de preferências.

9.6 Federalismo Brasileiro

Uma característica central do federalismo brasileiro é que a descentralização não foi feita através da descentralização do poder de tributar, mas sim mediante transferências fiscais do governo central para os estados e municípios e dos estados para os municípios.

Isso cria o conhecido problema do "poço comum": se vários governos locais retiram suas receitas de uma única fonte (a arrecadação dos níveis superiores de governo), cada governo local tentará extrair o máximo que puder.

O resultado é o excesso de gastos dos governos locais e a necessidade de o governo central e os governos estaduais aumentarem a carga tributária para financiar essas despesas crescentes.

De início, convém destacar dois aspectos fundamentais

1. as enormes disparidades regionais

2. a forte tradição municipalista do País

Como de hábito, em regimes federativos cabe à União envidar esforços para reduzir as disparidades regionais de desenvolvimento, e os instrumentos fiscais costumam desempenhar um importante papel a respeito.

No Brasil, a marcante presença do Estado na vida econômica contribuiu para dar ao sistema tributário um papel de relevo na política de desenvolvimento regional, colocando em conflito, com frequência, as demandas por maior autonomia tributária dos estados mais desenvolvidos com as pressões por aumento das transferências compensatórias executadas pelos estados de menor grau de desenvolvimento.

Outro condicionamento importante é o que resulta da forte tradição municipalista.

A força dessa tradição está hoje refletida no caráter singular assumido pela federação brasileira após a promulgação da Constituição de 1988.

Nela, os municípios foram reconhecidos como membros da federação, em pé de igualdade com os estados no que diz respeito a direitos e deveres ditados pelo regime federativo.

As principais consequências desse fenômeno, do ponto de vista do federalismo fiscal, são o largo campo de competência impositiva dos municípios e a instituição de transferências compensatórias federais semelhantes às que beneficiam os estados.

A relativa independência dos municípios em relação ao poder público estadual, conferida pela posição singular que ocupam no sistema tributário brasileiro, é causa de importantes distorções que acentuam os desequilíbrios verticais e horizontais na repartição da receita tributária nacional, tomando mais difícil a negociação de reformas capazes de sedimentar propostas de um novo equilíbrio federativo.

10 Dívida Pública

Não é a vitória da ciência que distingue o nosso século dezenove, mas a vitória do método científico sobre a ciência.

Nietzsche

[...] the real problem is the premise that you can split content from style. It's wrong. They are yolk and white in a scrambled egg. Economically speaking, the production function for thinking cannot be written as the sum of two sub-functions, one producing "results" and the other "writing them up". The function is not separable. You do not learn the details of an argument until writing it in detail, and in writing the details you uncover flaws in the fundamentals. Thinking requires detail. Good thinking is accurate, symmetrical, relevant to the thoughts of the audience, concrete yet usefully abstract, concise yet usefully full; above all it is self-critical and honest. So too is good writing. Writing resembles mathematics. Mathematics is a language, an instrument of communication. But so too language is a mathematics, an instrument of thought.

McCloskey, 2000

10.1 Introdução

O setor público abrange a administração direta, as autarquias e as fundações das três esferas de governo (federal, estadual e municipal) e suas respectivas empresas estatais, o Banco Central e o Instituto Nacional de Seguridade Social (INSS).

O conceito de setor público, para fins de apuração dos indicadores de dívida pública, considera as instituições públicas não financeiras, bem como os fundos públicos que não possuem características de intermediários financeiros, isto é, aqueles cujas fontes de recursos advêm de contribuições fiscais ou parafiscais, além da empresa Itaipu Binacional.

A dívida é uma obrigação de determinada entidade com terceiros, gerada pela diferença entre despesas e receitas dessa entidade.

Em outras palavras, só há dívida quando há déficit (despesas maiores que receitas), embora muitas vezes ocorra defasagem entre a realização do déficit e a contabilização da dívida.

O resultado nominal do setor público, também conhecido como Necessidade de Financiamento do Setor Público (NFSP), é o conceito mais amplo de resultado fiscal e corresponde à diferença entre receitas e despesas nominais no período, incluídas as financeiras.

É calculado com base na variação da Dívida Fiscal Líquida, que exclui, dentre outros, o efeito dos passivos contingentes registrados (chamados “esqueletos”) e das receitas de privatização, visto que estes não representam resultado de esforço fiscal no exercício em que foram contabilizados.

O Resultado Primário do Setor Público é o resultado nominal, excluído o efeito dos juros nominais incidentes sobre a dívida pública interna e externa.

Esse indicador mede o efetivo esforço determinado pela política fiscal, “descontaminada” dos efeitos da taxa de juros nominal sobre o estoque da dívida existente, que é função dos déficits acumulados no passado.

O resultado primário pode ser mensurado de duas formas:

1. pela diferença entre receitas (exceto aplicações financeiras) e despesas (exceto juros), a qual é denominada "acima da linha" e é calculada e divulgada mensalmente pelo Tesouro Nacional;

2. pela variação da Dívida Fiscal Líquida – resultado nominal calculado pelo Banco Central segundo o conceito "abaixo da linha" – descontada dos valores referentes aos juros nominais.

Um ponto importante a destacar é que o conceito de dívida líquida é o que mais comumente se utiliza para fins de acompanhamento da sustentabilidade fiscal de um país.

Dívida Líquida do Setor Público =

(−) Ajuste de Privatização

(−) Ajuste Patrimonial

(−) Ajuste Metodológico sobre a Dívida Externa

(−) Dívida Fiscal Líquida com Câmbio

(−) Ajuste Metodológico sobre a Dívida Interna

(−) Dívida Fiscal Líquida

A Dívida Líquida do Setor Público (DLSP) refere-se ao total das obrigações do setor público não financeiro, deduzido dos seus ativos financeiros junto aos agentes privados não financeiros e aos agentes financeiros, públicos e privados.

10.2 Processo de Planejamento Estratégico da DPF

O planejamento estratégico da Dívida Pública Federal (DPF) no Brasil envolve vários aspectos que podem ser didaticamente agrupados em três fases:

1. Definição da estrutura desejada no longo prazo (*benchmark*)
2. Planejamento de médio prazo (estratégia de transição)
3. Elaboração, divulgação e execução da estratégia de curto prazo (Plano Anual de Financiamento – PAF), cuja primeira publicação ocorreu em 2001.

O ponto de partida e a referência principal para todo o processo de planejamento da dívida pública é a definição clara de seus objetivos.

No Brasil, o objetivo estabelecido para a gestão da DPF, divulgado em todos os seus Planos Anuais de Financiamento desde 2001, consiste em "suprir de forma eficiente as necessidades de financiamento do governo federal, ao menor custo de financiamento no longo prazo, respeitando-se a manutenção de níveis prudentes de risco. Adicionalmente, busca-se contribuir para o bom funcionamento do mercado brasileiro de títulos públicos".

Tendo por referência o objetivo supracitado, o PAF apresenta um conjunto de diretrizes que norteiam a elaboração das estratégias de financiamento da DPF. São elas:

1. Aumento do prazo médio do seu estoque
2. Suavização da estrutura de vencimentos, com especial atenção para a dívida que vence no curto prazo
3. Substituição gradual dos títulos remunerados pela taxa de juros Selic por títulos com rentabilidade prefixada ou vinculada a índices de preços

4. Aperfeiçoamento do perfil da Dívida Pública Federal externa (DPFe), por meio de emissões de títulos com prazos de referência (*benchmarks*), programa de resgate antecipado e operações estruturadas

5. Desenvolvimento da estrutura a termo de taxas de juros nos mercados interno e externo e o aumento da liquidez dos títulos públicos federais no mercado secundário

6. Ampliação da base de investidores

A cada ano, tendo em vista o objetivo e as diretrizes do PAF, os cenários macroeconômicos, a estimativa da necessidade de financiamento do Governo Federal e as diversas estratégias de emissão dos títulos da dívida pública, o Tesouro Nacional calcula os valores esperados para os principais indicadores da DPF: estoque, composição por indexador, prazo médio e percentual vincendo em 12 meses.

A partir daí, são divulgados por meio do PAF limites indicativos dos valores superior e inferior que cada um desses indicadores deve assumir ao final do ano.

Base para elaboração do PAF, o planejamento estratégico da DPF define uma "estratégia de transição" da composição atual da dívida pública para o *benchmark* de longo prazo.

A estratégia de transição procura responder à seguinte questão: quais devem ser a trajetória e a velocidade de convergência para a composição de longo prazo desejada, respeitando-se as condições iniciais (isto é, o atual perfil da dívida) e as restrições de curto e médio prazos (especialmente, restrições macroeconômicas e de desenvolvimento dos mercados financeiros locais).

A escolha das estratégias de transição para o longo prazo também explora os *trade-offs* entre custos e riscos da dívida pública.

Já a composição ótima de longo prazo (*benchmark*) é a primeira etapa a ser discutida e aprovada pelo Comitê da Dívida Pública, a partir da qual é elaborada a estratégia de transição e aprovado o Plano Anual de Financiamento para cada ano.

No Brasil, o desenvolvimento do modelo de composição ótima da dívida pública foi uma consequência natural de um longo processo de melhoria do arcabouço institucional utilizado para avaliar os custos e riscos da DPF.

Inicialmente, implantou-se o modelo de gestão ativos e passivos do governo.

A seguir, surgiram os instrumentos gestão de riscos utilizados pelo Tesouro Nacional na gestão da DPF.

Só então, partiu-se para o estudo de um modelo de composição ótima para a dívida pública que considerasse todas as variáveis relevantes.

10.3 O Arcabouço Analítico do Benchmark da DPF

Especificamente com respeito à definição de uma composição ótima (*benchmark*) de longo prazo para a dívida pública, esta representa o perfil desejado para a estrutura da dívida e constitui um guia para o delineamento de estratégias de financiamento de curto e médio prazo do governo.

No caso brasileiro, o *benchmark* é expresso por um conjunto de indicadores relevantes para a dívida, sendo eles a composição do estoque por tipo de remuneração, o prazo médio e a estrutura de vencimentos, particularmente a proporção de dívida a vencer nos próximos 12 meses.

Na definição da composição ótima (*benchmark*) da dívida pública, um conjunto de modelos descreve como as variáveis macroeconômicas e financeiras relevantes para a trajetória da dívida pública (taxas de juros, taxa de câmbio, inflação e PIB) evoluem ao longo do tempo.

Com base em cenários simulados, a evolução da razão dívida/PIB é avaliada para se derivar medidas de custo e risco de uma dada estrutura de dívida.

Assim, após o exame de múltiplas alternativas possíveis, obtém-se a fronteira eficiente em termos de custos e riscos da dívida pública.

A partir daí, escolhe-se aquela estrutura na fronteira que possui o perfil desejado para o longo prazo, de acordo com as preferências da sociedade entre custos e riscos.

Uma questão importante no modelo se refere a qual deve ser o conceito de dívida relevante para avaliação do custo e dos riscos.

No caso brasileiro, considera-se que a razão entre a Dívida Líquida do Setor Público e o PIB (DLSP/PIB) é a medida mais relevante para esse propósito, pois este é o indicador mais comumente utilizado, tanto pelo Governo Federal para definir suas metas de endividamento e o superávit primário necessário para atingi-las, quanto pelos analistas para avaliar a sustentabilidade fiscal.

Apesar de o instrumento de trabalho do Tesouro Nacional ser a DPF, busca-se uma comunicação clara entre esta dívida e a DLSP, que é mais ampla e também é referência de política econômica.

Os Riscos na Gestão da DPF são

1. **Risco de financiamento**

   Esse risco está relacionado à capacidade do governo de rolar (refinanciar) sua dívida que vence no curto prazo. Caso o ambiente econômico ou financeiro se deteriore, pode haver um aumento abrupto no custo dos financiamentos ou mesmo a impossibilidade de captar recursos no mercado.

   - **Impactos:** Um aumento no custo do financiamento pode elevar significativamente os gastos com juros da dívida, pressionando o orçamento público e limitando investimentos e gastos sociais.
   - **Causas:** Mudanças no cenário macroeconômico, aumento da percepção de risco país pelos investidores, crise financeira global ou local, e deterioração da confiança na capacidade fiscal do governo.
   - **Medidas para mitigar:** Alongar o perfil da dívida para reduzir a concentração de vencimentos no curto prazo, diversificar as fontes de financiamento e manter reservas financeiras para contingências.

2. **Risco de mercado**

Esse risco decorre da volatilidade dos preços dos ativos e das variáveis econômicas que influenciam o custo da dívida, como taxas de juros, inflação e câmbio.

- **Exemplo:** Se a taxa de juros subir, o custo do serviço da dívida indexada a essas taxas aumenta; se o real se desvalorizar, o custo da dívida em moeda estrangeira também sobe.
- **Consequências:** Oscilações inesperadas podem aumentar os encargos financeiros, afetando a sustentabilidade fiscal e a credibilidade do país.
- **Gestão:** Uso de instrumentos financeiros como swaps e contratos futuros para proteger a carteira da dívida contra variações adversas, além do monitoramento constante dos mercados financeiros.

3. **Risco estratégico**

Esse risco está associado à possibilidade de a estratégia adotada para gerir a dívida pública não alcançar os resultados esperados.

- **No Tesouro Nacional:** Pode envolver, por exemplo, a escolha inadequada de indexadores para a dívida (entre taxa fixa, indexada à inflação ou ao câmbio), que não esteja alinhada às condições de mercado ou às metas fiscais.
- **Impacto:** Pode comprometer o equilíbrio fiscal e a confiança dos investidores, gerando custos adicionais e aumentando o risco de refinanciamento.
- **Mitigação:** Planejamento estratégico rigoroso, análise constante das condições econômicas e financeiras, e flexibilidade para ajustar a estratégia conforme as mudanças no cenário.

4. **Risco operacional**

Refere-se a falhas internas ou externas que possam prejudicar as operações relacionadas à gestão da dívida.

- **Exemplos:** Erros em processos administrativos, falhas em sistemas de TI, ataques cibernéticos, ou problemas em eventos como leilões de títulos.
- **Consequências:** Interrupção de operações, perda financeira direta ou indireta, e danos à reputação do órgão gestor da dívida.
- **Medidas:** Implantação de controles internos rigorosos, planos de contingência, treinamento constante da equipe, e investimentos em segurança da informação e infraestrutura tecnológica.

5. **Risco legal**

   Refere-se à possibilidade de descumprimento de normas legais que regulam os limites e parâmetros da dívida pública.

   - **Aspectos:** Pode envolver ultrapassar limites de emissão, desrespeitar regras fiscais ou normativas relacionadas à gestão da dívida.
   - **Consequências:** Penalidades legais, restrições ao acesso a mercados financeiros, perda de credibilidade institucional, e consequências políticas.
   - **Prevenção:** Monitoramento constante da legislação vigente, conformidade regulatória rigorosa, e transparência nas operações.

As Tabelas 10.1 e 10.2, a seguir, apresentam uma breve descrição dos principais indicadores utilizados para monitorar os riscos da DPF no Brasil.

Tabela 10.1 – *Indicadores de Risco da Dívida Pública Federal (DPF): Risco de Refinanciamento*

Indicador	Descrição	Comentários
Percentual vincendo em 12 meses da DPF	Indica a concentração de curto prazo de vencimentos da dívida.	Uma descrição completa do perfil de maturação da dívida pode ser feita a partir da análise de toda a estrutura de vencimentos da DPF (isto é, avaliando-se também a porcentagem da dívida vincendo no médio e no longo prazos).
Prazo Médio da DPF	Indica o período de tempo que, em média, toda a dívida deverá ser paga ou refinanciada.	No Brasil, no cálculo do prazo médio da DPF utiliza-se o valor presente dos fluxos (principal e juros) da dívida como fator de ponderação dos prazos de cada fluxo (Conceito de duration).
Cash-Flow-at-Risk (CFaR)	Indica o aumento máximo que pode ocorrer nos fluxos de pagamentos da DPF projetados para um dado período, considerando-se um dado intervalo de confiança (e.g. probabilidade de 95%).	O CFaR e o CaR são indicadores de risco baseados em simulações estocásticas, que têm a vantagem de indicar uma distribuição de probabilidades do valor dos fluxos de pagamentos ou do valor do estoque da dívida. Esse tipo de medida de risco permite estimar perdas para a dívida pública decorrentes de eventos negativos que ocorrem na economia, além de quantificar a probabilidade de tais eventos.

Tabela 10.2 – *Indicadores de Risco da Dívida Pública Federal (DPF): Risco de Refinanciamento*

Indicador	Descrição	Comentários
Composição da DPF	Indica a porcentagem do estoque da dívida por tipo de remuneração (Prefixados, Flutuantes, Índices de Preços e Cambiais).	Os tipos de remuneração que categorizam o estoque da DPF nesse indicador são definidos de acordo com classes de risco que, por sua vez, dependem dos indexadores da dívida. Na DPF há 4 classes: Juros Prefixados, Juros Flutuantes, Indexados à Inflação (Índices de Preços), Atrelados à Taxa de Câmbio (Cambiais).
Risco de Repactuação (ou risco de taxas de juros)	Indica a parcela da DPF sujeita a aumentos no seu custo devido a flutuações nas taxas de juros no curto prazo.	Corresponde à dívida exposta à flutuações nas taxas de juros, seja porque a dívida deve ser refinanciada (por novas taxas) ou porque a dívida é remunerada por taxas de juros flutuantes (por exemplo: taxa de juros Selic). Assim, o indicador é dado pelo somatório da porcentagens de DPF a vencer em 12 meses e da remunerada por juros flutuantes a vencer após 12 meses.
Análise de Sensibilidade da DPF	Indica o aumento no estoque (custo) da dívida devido a uma variação de 1% em um determinado indexador (Taxa de juros de curto prazo ou taxa de câmbio).	Trata-se de uma análise que procura responder "O que acontece se determinado choque ocorre?". Alternativamente, esse indicador pode ser calculado assumindo variação equivalente a um desvio-padrão no indexador de referência.
Teste de Estresse	Mede o impacto negativo no estoque (ou no custo) da DPF devido a uma pressão forte e persistente sobre a taxa de juros reais ou sobre a taxa de câmbio.	É equivalente à análise de sensibilidade, mas, neste caso, aplica-se um choque equivalente a 3 desvios-padrão da taxa de juros real ou da desvalorização cambial real acumuladas em 12 meses sobre o estoque e composição da DPF.
Cost-at-Risk (CaR)	Indica o valor máximo que o estoque da DPF pode atingir ao final de um determinado período (e.g.: 1 ano) para um dado intervalo de confiança (e.g. probabilidade de 95%).	No Brasil, o cost-at-risk é usado para medir a incerteza com relação ao montante de dívida ao final de um período. Apesar de ser definido em termos de estoque, ao invés de custos (juros), essas duas abordagens são diretamente relacionadas, uma vez que quanto maior o custo maior o estoque da dívida para um dado resultado primário do governo.

A importância de uma composição ótima (*benchmark*) é amparada na literatura teórica, que preconiza a relevância da gestão da dívida pública para a atividade econômica, especialmente as literaturas sobre tributação ótima (*tax smoothing*) e consistência temporal, que levam à defesa de uma gestão ativa da dívida.

Os argumentos teóricos em favor da busca de uma adequada composição de dívida se ampliam quando são considerados elementos providos pelas literaturas sobre credibilidade das políticas macroeconômicas, sinalização, e efeitos reais de um *default* soberano, dentre outros.

Neste debate deve se registrar também a contribuição de instituições multilaterais, tais como o Banco Mundial e o Fundo Monetário Internacional.

Estas duas instituições descrevem o *benchmark*, em sua publicação *Guidelines for Public Debt Management*, como uma poderosa ferramenta para representar o perfil de dívida que o governo deseja atingir, com base em suas preferências diante do *trade-off* entre custos e riscos.

Finalmente, a experiência internacional documenta que diversos países empreenderam esforços para a definição de uma composição ótima para suas dívidas.

Este é o caso de Portugal, um dos pioneiros na formulação e adoção de um modelo de composição ótima de longo prazo para quantificar o objetivo da gestão de sua dívida pública, bem como para aumentar a consistência entre as decisões diárias e o objetivo de longo prazo.

Dinamarca, Suécia, Canadá e Reino Unido também desenvolveram modelos para auxiliar na definição de um portfólio de referência para guiar a elaboração de estratégias de financiamento.

10.4 Testes de Estacionariedade

Para cada período, a dívida pública evolui de acordo com a seguinte relação:

$$B_{t+1} = (1 + r_t)B_t + G_{t+1} - T_{t+1} \tag{10.1}$$

em que B_t é o valor da dívida do governo no momento t, r_t é o valor da taxa de juros em t, e T_t e G_t são as receitas e as despesas do governo em t. De modo semelhante,

$$B_{t+2} = (1 + r_{t+1})B_{t+1} + G_{t+2} - T_{t+2} \tag{10.2}$$

Substituindo recursivamente em (10.1), temos:

$$B_t = \frac{B_{t+2}}{(1 + r_{t+1})(1 + r_t)} + \frac{T_{t+2} - G_{t+2}}{(1 + r_{t+1})(1 + r_t)} + \frac{T_{t+1} - G_{t+1}}{(1 + r_t)} \tag{10.3}$$

Assim, retrocedendo até um período $t + s$:

$$B_t = \frac{B_{t+s}}{\prod_{v=1}^{s} (1 + r_{t+v-1})} + \sum_{v=0}^{s} \frac{T_{t+v} - G_{t+v}}{r_v} \tag{10.4}$$

Usamos r_t para denotar a taxa de juros de um título comprado em t, a ser honrado em $t + 1$. Da mesma forma, $G_t - T_t$ representa o déficit primário em t. A restrição-fluxo (10.1) representa um requerimento mínimo.

O que torna interessante o conceito de sustentabilidade é a condição de transversalidade.

Notando que o preço em t do consumo em $t + s$ é dado por

$$P_t = \left(\prod_{v=1}^{s} (1 + r_{t+v-1}) \right)^{-1} \tag{10.5}$$

A condição de transversalidade é

$$\lim_{t\to\infty} P_t B_t \leq 0 \tag{10.6}$$

isto é, o valor presente da dívida deve se aproximar de um valor não positivo quando um horizonte de tempo suficientemente longo for considerado.

Essa condição elimina os chamados jogos Ponzi, em que uma dívida é sempre “rolada” e nunca paga.

Ou seja, corresponde à hipótese de que governos não podem endividar-se permanentemente.

É natural admitir que as pessoas também não possam endividar-se contra o governo indefinidamente, o que justifica a imposição da restrição (10.6) como uma igualdade, em cujo caso tem-se

$$B_t = \sum_{v=0}^{\infty} \frac{T_{t+v} - G_{t+v}}{\prod_{v=1}^{s} (1 + r_{t+v-1})} \tag{10.7}$$

A imposição da condição de transversalidade (10.6) com igualdade garante, portanto, que o valor presente dos superávits primários seja igual ao valor da dívida. Isso significa que o governo, em algum momento do tempo, deve arrecadar recursos suficientes não somente para cobrir seus gastos correntes, mas também para honrar integralmente seus compromissos financeiros acrescidos dos respectivos juros. Assim, a solvência do setor público é assegurada pela capacidade futura de gerar superávits primários suficientes para extinguir o estoque de dívida acumulada.

No entanto, a dinâmica da dívida pública está sujeita a importantes fontes de incerteza, que podem comprometer essa trajetória ideal de equilíbrio fiscal. Podemos destacar duas dimensões principais em que a incerteza exerce papel fundamental:

1. **Incerteza sobre receitas e despesas públicas**

   O valor das receitas (arrecadação tributária, receitas não tributárias, transferências) e das despesas públicas (investimentos, gastos correntes, transferências, juros da dívida) é afetado por variáveis econômicas, políticas e institucionais que são intrinsecamente incertas. Eventos imprevistos, como recessões econômicas, choques externos, mudanças políticas ou desastres naturais, podem afetar significativamente o saldo primário do governo, alterando a trajetória projetada da dívida. Essa variabilidade dificulta o planejamento e o cumprimento das metas fiscais, aumentando o risco de desequilíbrio fiscal.

2. **Incerteza sobre as taxas de desconto aplicadas aos superávits ou déficits futuros**

   Além da incerteza quanto aos valores nominais dos superávits ou déficits futuros, há também incerteza sobre a taxa pela qual esses fluxos são descontados para determinar seu valor presente. As taxas de juros reais, expectativas inflacionárias e o risco-país influenciam esse fator de desconto. Variações nessas taxas alteram o valor presente da dívida e, consequentemente, a avaliação da sustentabilidade fiscal. Cenários de alta volatilidade nas taxas de juros podem fazer com que, mesmo com superávits primários constantes, o valor presente da dívida se altere substancialmente, impactando a percepção de solvência e a estratégia de gestão da dívida.

Portanto, o enfrentamento dessas incertezas exige não apenas a formulação de políticas fiscais prudentes e a manutenção de *buffers* financeiros[45] adequados, mas

[45] Os *buffers* financeiros referem-se a reservas ou margens de segurança que uma organização ou indivíduo mantém para lidar com imprevistos financeiros ou períodos de incerteza econômica. Esses *buffers* servem como uma proteção contra riscos financeiros, como flutuações na receita, aumentos inesperados de custos ou crises econômicas. Esses *buffers* podem ser usados para cobrir despesas inesperadas, cumprir obrigações financeiras ou manter a estabilidade financeira em tempos difíceis. Em um contexto corporativo, por exemplo, uma empresa pode manter um fundo de reserva ou uma linha de crédito como um *buffer* financeiro. Já no contexto pessoal, ter uma reserva de emergência é uma forma de *buffer* financeiro. Em resumo, *buffers* financeiros são mecanismos de proteção

também o desenvolvimento de modelos de análise e previsão que incorporem choques econômicos e variações nas condições de mercado. A gestão eficaz da dívida pública depende da capacidade de antecipar e mitigar os efeitos adversos dessas incertezas, assegurando a estabilidade fiscal e a confiança dos investidores no longo prazo.

Consideremos o primeiro caso. Se o fluxo de superávit fosse variável, mas pudéssemos descontá-lo por uma taxa que independesse do cenário (o que quer dizer que o valor de uma unidade de poder de compra é igual em todos os diferentes cenários), então a nova condição de sustentabilidade seria de que o valor presente esperado dos superávits do governo fosse o mesmo (igual ao valor da dívida) para todos os cenários.

Em um mundo com incerteza, ativos diferentes com características de risco distintas pagam retornos distintos. Há várias taxas de desconto. Qual delas é a relevante? Como o pagamento somente se dará no futuro, e o futuro é incerto, o pagamento tem de ser ajustado em duas dimensões: temporal e de risco. Na dimensão temporal, poder de compra amanhã tem menos valor do que poder de compra hoje. Assim, deve-se pagar uma taxa de juros positiva. No que concerne ao risco, pessoas avessas a ele são aquelas que atribuem maior valor à renda quando a têm menos. Uma política fiscal em que os superávits são gerados principalmente em momentos de recessão, ou seja, em que a dívida é abatida nesses momentos, está associada, porém, a uma dívida pública mais barata. O menor custo dessa dívida advém da imposição de um maior custo social da tributação.

Somente olhar para o custo médio da dívida pode induzir a uma política fiscal socialmente perversa, em que a redução da oferta de bens públicos e a elevação da carga tributária ocorrem exatamente nos momentos de recessão. Assim, ainda que enfatizemos aqui a questão da sustentabilidade, é importante ter em mente que aquelas políticas que parecem fazer mais sentido do ponto de vista de reduzir o custo

que ajudam a garantir que, mesmo diante de adversidades financeiras, a saúde econômica de uma entidade ou pessoa seja preservada.

financeiro da dívida podem ser as mais custosas do ponto de vista social, sendo, de fato, politicamente insustentáveis.

A condição de sustentabilidade é

$$B_t = \mathbb{E}_t \left[\sum_{v=1}^{\infty} m_{t+v} \left[T_{t+v} - G_{t+v} \right] \right] \tag{10.8}$$

em que m_t é o chamado fator estocástico de desconto (ou *pricing kernel*[46]), uma variável aleatória que desconta os fluxos incertos para incorporar as dimensões tempo e risco.

Ou seja, a taxa de desconto relevante para o desconto dos fluxos de superávits é ajustada para o risco.

[46] O *pricing kernel*, também conhecido como *stochastic discount factor* (fator estocástico de desconto), é uma variável aleatória que relaciona os valores dos ativos em diferentes momentos no tempo, levando em consideração tanto o valor temporal do dinheiro quanto o risco associado a fluxos futuros. Matematicamente, o *pricing kernel* é um fator que, ao ser multiplicado pelo *payoff* futuro de um ativo, permite calcular seu valor presente esperado ajustado pelo risco. Em outras palavras, ele transforma valores futuros incertos em valores presentes certos, considerando a preferência dos agentes econômicos por consumo presente e sua aversão ao risco.

Formalmente, para um ativo que paga um valor X_{t+1} no futuro, seu preço no tempo t é dado por:

$$P_t = \mathbb{E}_t \left[m_{t+1} X_{t+1} \right]$$

em que

- P_t é o preço do ativo no tempo t,
- X_{t+1} é o *payoff* (pagamento) no tempo futuro $t + 1$,
- m_{t+1} é o *pricing kernel*, que incorpora as preferências temporais e a aversão ao risco,
- $\mathbb{E}_t[\cdot]$ representa a expectativa condicional à informação disponível em t.

O *pricing kernel* é fundamental na teoria financeira, pois ele define o modo como os agentes descontam valores futuros para o presente, ajustando pelos riscos que enfrentam. Por exemplo, em um ambiente sem risco, o *pricing kernel* seria simplesmente o fator de desconto determinístico baseado na taxa de juros livre de risco. Em ambientes com risco, ele varia aleatoriamente para refletir a incerteza. No contexto das finanças públicas, como no modelo da dívida pública sustentável, o *pricing kernel* ajusta o valor dos superávits ou déficits futuros para refletir tanto o tempo quanto o risco, permitindo que se avalie corretamente o valor presente da dívida e das obrigações futuras do governo.

A natureza do conceito de sustentabilidade impede a definição de uma medida objetiva que determine se uma dívida é sustentável. Os testes são indicadores capazes de auxiliar a formação de crenças sobre a trajetória futura de superávits, suas associadas taxas de desconto e sua compatibilidade com a satisfação da restrição orçamentária do governo.

Dizemos que um processo estocástico é estacionário quando tende a reverter à sua média ou à sua tendência depois de um choque aleatório. Imaginemos, então, que a postura fiscal do governo seja tal que, em seguida a um choque que mude o valor da dívida, os superávits sejam elevados para fazer com que a dívida lentamente retorne a seu valor (ou, se a dívida tiver uma taxa de crescimento lenta, por exemplo, igual à taxa de crescimento do PIB, que retorne a essa tendência). Então, é fácil ver que essa postura fiscal faz com que o valor da dívida respeite a condição de transversalidade, isto é, que a dívida seja sustentável. Os ajustes são quase sempre obtidos por meio de elevação de impostos e que a receita de senhoriagem precisa ser somada à receita tributária para que a receita e a despesa convirjam no longo prazo. Seria necessário testar todas as ordens de integração. Não garante nada. Uma dívida integrada de qualquer ordem arbitrária é sustentável.

Toda avaliação de sustentabilidade requer a formação de crenças acerca da capacidade do país de fazer o necessário sacrifício para gerar os superávits que garantam que a equação de transversalidade seja satisfeita. O real custo desse sacrifício depende diretamente de que proporção da riqueza será empregada para esse fim.

Seja

$$\frac{B_{t+1}}{Y_{t+1}} = (1 + r_t)\frac{B_t}{Y_t}\frac{Y_t}{Y_{t+1}} + \frac{G_{t+1}}{Y_{t+1}} - \frac{T_{t+1}}{Y_{t+1}} \tag{10.9}$$

isto é,

$$b_{t+1} - b_t = \frac{r_t - \gamma_t}{1 + \gamma_t} b_t + g_{t+1} - \tau_{t+1} \tag{10.10}$$

Supomos $\gamma_t < r_t$: evitar a ineficiência dinâmica da economia. Uma economia dinamicamente ineficiente é aquela em que existe um acúmulo excessivo de capital. Nesse caso, há espaço para ampliação do consumo sem sacrifício da renda disponível para as gerações futuras.

Os estudos que usam a relação dívida/PIB como indicador de sustentabilidade, em sua maioria, exploram o comportamento dessa variável ao longo do tempo, avaliando se ela tem uma tendência de estabilidade ou decréscimo. Uma dívida estável com relação ao PIB tem seu valor presente decrescente com o tempo. Portanto, a estabilidade da relação dívida/PIB é uma condição suficiente para garantir a sustentabilidade da dívida pública. Isto é, $b_{t+1} = b_t = b$:

$$0 = \frac{r_t - \gamma_t}{1 + \gamma_t} b + g_{t+1} - \tau_{t+1} \Longrightarrow \tau_{t+1} - g_{t+1} = \frac{r_t - \gamma_t}{1 + \gamma_t} b \tag{10.11}$$

O lado direito da expressão anterior nos dá o superávit (como proporção do PIB) necessário para estabilizar a relação dívida/PIB em função da relação dívida/PIB atual, da taxa de juros e da taxa de crescimento da economia.

Suponha, por exemplo, um país cuja relação dívida/PIB se encontre em 40%, cujo custo de carregamento (taxa real) dessa dívida seja de 7% e esteja crescendo a 5% ao ano. Nesse caso, o superávit necessário para estabilizar a relação dívida/PIB seria igual a

$$(r_t - \gamma_t) \times b_t = (0,07 - 0,05) \times 0,4 = 0,008 \tag{10.12}$$

Ou seja, 0,8% do PIB.

Forçar ano a ano o superávit primário a satisfazer tal regra implica eliminar o papel fundamental do endividamento público: dissociar temporalmente gastos públicos do seu financiamento, escolhendo de forma independente o melhor momento de produzir um e outro.

Assim, é preciso apresentar formas de avaliar a sustentabilidade, considerando trajetórias alternativas das variáveis que não impliquem necessariamente a constância da relação dívida/PIB.

De acordo com a metodologia VaR (*Value-at-Risk*), avalia-se qual o maior valor, tal que a relação dívida/PIB não se situe acima dele com uma probabilidade predefinida. Essa metodologia permite também a incorporação de não linearidades, o que pode ter consequências interessantes para a compreensão da forma como o governo conduz sua política de endividamento. Quando a relação dívida/PIB evolui de forma não linear, é possível que ela venha a exibir momentos de comportamento explosivo sem que sua trajetória global o seja. Isso não é possível em um modelo linear, já que os comportamentos local e global são idênticos.

O alongamento sustentável da dívida requer ambiente favorável à demanda por títulos longos. E uma estrutura de dívida mais diversificada suaviza os efeitos do ciclo monetário.

10.5 Modelo Brasileiro

O arcabouço analítico empregado no estudo da composição ótima (*benchmark*) da DPF é baseado em simulações estocásticas derivadas das teorias de finanças e portfólio eficiente.

Contudo, antes de proceder à descrição do modelo propriamente dito, algumas ressalvas devem ser feitas quanto à aplicação direta dos instrumentos de análises financeiras tradicionais às políticas governamentais.

Em termos gerais, o governo pode ter objetivos mais complexos do que reduzir custos condicionado à manutenção de riscos em níveis prudentes.

Além disso, a evolução de seus fluxos de caixa e os indicadores de impactos orçamentários podem ter implicações sobre a escolha da estrutura ótima da dívida.

Há que se considerar ainda que, dada a natureza da dívida pública, as ações do governo têm forte influência sobre os preços dos títulos e, consequentemente, sobre o custo e risco de suas estratégias de financiamento.

Como resultado, estas peculiaridades podem levar os gestores da política econômica a definirem como *benchmark* uma composição de dívida diferente daquelas sobre a fronteira eficiente, obtida do ponto de vista estritamente financeiro.

O estudo da composição ótima (*benchmark*) para a dívida brasileira é baseado na aplicação de métodos de simulações estocásticas, com o propósito de se derivar uma fronteira eficiente de composições da dívida, que expresse potenciais *trade-offs* entre custos e riscos na gestão da DPF.

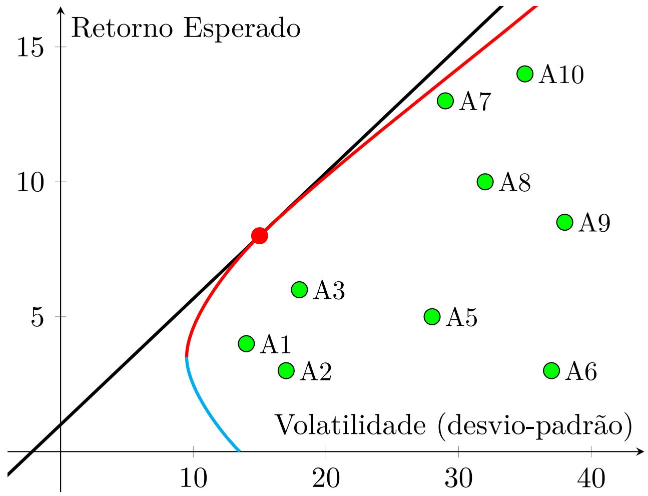
Neste sentido, define-se que uma composição é eficiente quando ela tem o menor risco para um determinado nível de custo ou, alternativamente, ela tem o menor custo para um determinado nível de risco.

O conjunto de todas as composições que atendem a essa condição define a fronteira eficiente, cabendo ao gestor da dívida escolher qual composição é desejável, uma vez que entre portfólios da fronteira não é possível obter ganhos simultâneos de redução de custo e risco.

Após efetuar as simulações, as métricas obtidas para cada composição avaliada são plotadas em um gráfico cujos eixos são o custo e o risco da DLSP/ PIB, de tal maneira que a fronteira eficiente é obtida como a curva composta pelos pontos que representam o menor custo para um determinado nível de risco.

Os portfólios da fronteira são eficientes porque não é possível alternar entre estes portfólios para obter ganhos de redução de custo e risco simultaneamente.

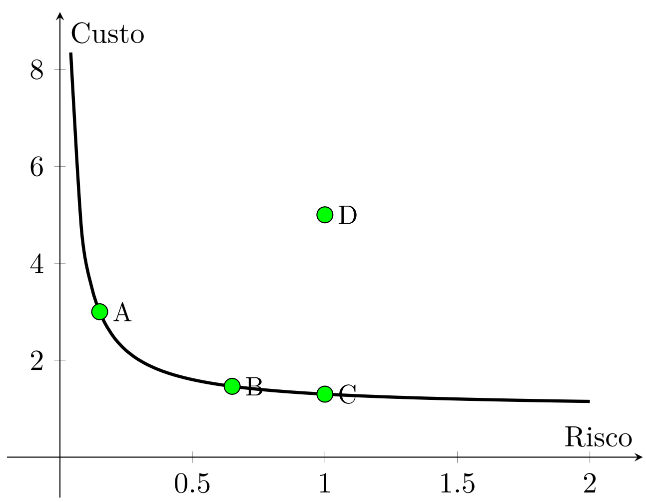
Finalmente, dado o apetite a risco do governo (que deveria refletir o da sociedade), é possível escolher um portfólio específico da fronteira que definirá o *benchmark*



da dívida.

A Figura 10.1 ilustra esse conceito.

Figura 10.1 – *Ilustração da Fronteira Eficiente*



Portfólios ao longo da fronteira (A, B e C) são eficientes porque o risco necessariamente aumenta quando se busca reduzir o custo da dívida alternando entre essas composições (de A para B; ou de B para C).

Composições acima e à direita da fronteira são ineficientes porque aumentam o risco, dado um nível de custo (D comparado com A), ou aumentam o custo para um dado nível de risco (D comparado com C), ou aumentam ambos, o custo e o risco, em comparação com uma composição eficiente (D comparado com B).

O conjunto de simulações depende da geração de cenários para as variáveis econômicas que determinam o custo de financiamento da dívida e a dinâmica da relação DLSP/PIB.

Para tanto, o modelo requer a especificação de um conjunto de equações usadas para descrever como essas variáveis evoluem ao longo do tempo (equações diferenciais estocásticas e econometria de séries de tempo).

Neste sentido, os processos básicos do modelo cobrem as seguintes variáveis:

- Taxa básica de juros (Selic)
- Estrutura a termo das taxas de juros
- Curva de juros para títulos prefixados
- Curva de juros para títulos remunerados pelo IPCA
- Curva de juros para títulos denominados em moeda externa
- Taxas de inflação (doméstica e externa)
- Taxa de câmbio (real e nominal)
- Produto Interno Bruto (PIB)

Basicamente, utilizam-se simulações de Monte Carlo para gerar milhares de cenários macroeconômicos.

O modelo CIR (Cox-Ingersoll-Ross) é utilizado para gerar os cenários de taxa de juros e o Movimento Browniano Geométrico para o PIB e a inflação.

Definidos os cenários, a seguinte identidade contábil é usada como ponto de partida para derivar a dinâmica da DLSP e sua relação com a DPF:

$$D_t = X_t + M_t - F_t + L_t^{\text{flut}} + L_t^{\pi} + L_t^{\text{câmbio}} \tag{10.13}$$

Nesta identidade, os ativos e as obrigações que compreendem a DLSP (D_t) são agrupados em quatro categorias:

- DPF (X_t),
- base monetária (M_t),
- reservas internacionais (F_t),
- outras obrigações líquidas do Setor Público:
  - remuneradas por taxas de juros flutuantes (L_t^{flut}),
  - indexadas à inflação (L_t^{π}),
  - com remuneração atrelada à variação cambial ($L_t^{\text{câmbio}}$).

O estoque de DPF no período t é igual ao seu estoque no período prévio mais seu custo de carregamento (c_t), menos o resultado fiscal primário (s_t), menos a variação da base monetária (ΔM_t). Assim, a DPF evolui de acordo com a seguinte equação:

$$X_t = X_{t-1}(1 + c_t) - s_t - \Delta M_t \tag{10.14}$$

Quanto às demais componentes da DLSP, supõe-se que:

- A base monetária é mantida constante como proporção do PIB ao longo do tempo.
- As reservas internacionais, a partir de um montante inicial, evoluem de acordo com sua remuneração:
- As outras obrigações líquidas do setor público evoluem da seguinte forma:

$$F_t = F_{t-1}(1 + r_t^{\text{reservas}})(1 + \Delta\text{câmbio}) \tag{10.15}$$

$$L_t^{\text{flut}} = L_{t-1}^{\text{flut}}(1 + \text{Selic}_t) \tag{10.16}$$

$$L_t^{IP} = L_{t-1}^{IP}(1 + c_t^{IP}) \tag{10.17}$$

$$L_t^{FX} = L_{t-1}^{FX}(1 + c_t^{FX}) \tag{10.18}$$

em que c^{IP} e c^{FX} são, respectivamente, o custo de carregamento dos títulos indexados à inflação e denominados em moeda externa; e r_t^{reservas} representa a taxa de retorno das reservas internacionais.

Após substituir as equações (10.14) a (10.18) na equação (10.13) e dividir a nova equação pelo PIB, algumas manipulações algébricas conduzem à seguinte relação para descrever a trajetória da razão DLSP/PIB ao longo do tempo:

$$d_t = x_{t-1}\frac{(1 + c_t)}{1 + \gamma_t} - s_t + \frac{m}{1 + \gamma_t} - f_{t-1}\frac{1 + c_t^{\text{reservas}}}{1 + \gamma_t} + \ell_{t-1}\frac{(1 + c_t^{\ell})}{1 + \gamma_t} \tag{10.19}$$

em que

- $c_t^{\text{reservas}} = (1 + r_t^{\text{reservas}})(1 + \Delta\text{câmbio}) - 1$
- $\ell_t = \ell_t^{\text{flutuante}} + \ell_t^{IP} + \ell_t^{FX}$
- $c_t^{\ell} = \dfrac{\text{Selic}_t \ell_t^{\text{flutuante}} + c_t^{IP}\ell_t^{IP} + c_t^{FX}\ell_t^{FX}}{\ell_t}$

Um ponto de destaque é que todos os modelos são calibrados para refletir expectativas de longo prazo e não níveis correntes.

O uso do modelo permite que os formuladores de políticas simulem o efeito de mudanças nas taxas de juros, impostos, gastos públicos e até mesmo na base monetária sobre a trajetória da dívida. Isso é crucial para avaliar a sustentabilidade fiscal e identificar as condições necessárias para manter a dívida pública em níveis controláveis sem recorrer a medidas drásticas, como a austeridade fiscal ou a emissão excessiva de moeda.

Se o crescimento do PIB não for suficiente para cobrir os custos da dívida (principalmente devido à alta taxa de juros), o Brasil pode enfrentar dificuldades em manter a sustentabilidade da dívida. Isso exige uma gestão fiscal responsável, com equilíbrio entre a arrecadação de impostos, os gastos públicos e o financiamento da dívida, para evitar que o endividamento cresça de forma insustentável.

Além disso, o modelo pode ser utilizado para avaliar a viabilidade de políticas monetárias que busquem controlar a inflação sem prejudicar o crescimento econômico, ou estratégias fiscais que busquem aumentar a arrecadação sem comprometer o crescimento sustentável.