Microeconomia

Dr. Victor Oliveira

Lista de Figuras

Sumário

Capítulo 1

Teoria do Consumidor

Contents

> É a teoria e a sua estrutura que se conformam às leis da lógica e da racionalidade. As regularidades predicáveis das respostas das pessoas a mudanças em seu ambiente não requerem que elas sejam "racionais" não mais que a reação da água ao declive requer a racionalidade de cada molécula de água.
>
> Armen Alchian e William Allen.

1.1 Introdução

Neste capítulo, exploraremos as características essenciais da moderna teoria do consumidor - uma base fundamental sobre a qual se constroem tantas estruturas teóricas em economia, e também é central no modo de pensar dos economistas. Um consumidor pode ser caracterizado por muitos fatores e aspectos como sexo, idade, estilo de vida, riqueza, parentesco, habilidade, inteligência, etc. Mas quais são os mais importantes para nós estudarmos o comportamento do consumidor ao fazer escolhas? Para apreender as características mais importantes no estudo do comportamento e das escolhas do consumidor na teoria moderna do consumidor, supõe-se que a característica principal de um consumidor consiste em três componentes essenciais: o conjunto de consumo, dotações iniciais e a relação de preferência. As características do consumidor, juntamente com o pressuposto de comportamento, são blocos de construção em qualquer modelo da teoria do consumidor. O conjunto de consumo representa o conjunto de todas as alternativas ou planos de consumo individualmente viáveis e, às vezes, também chamado de conjunto de opções. Uma dotação inicial representa a quantidade de vários bens que o consumidor inicialmente tem e pode consumir ou negociar com outros indivíduos. A relação de preferência especifica os gostos ou satisfações do consumidor para os diferentes objetos de escolha. A suposição de comportamento expressa o princípio orientador que o consumidor usa para fazer escolhas finais e identifica os objetos finais na escolha. É geralmente assumido que o consumidor procura identificar e selecionar uma alternativa disponível que seja mais preferida à luz dos seus gostos/interesses pessoais.

Vamos enunciar os postulados da Teoria dos Preços, como apresentados por Alchian e Allen em *University Economics: Elements of Inquiry.*

Postulado 1.1.1 (Cada pessoa deseja uma infinidade de bens)**.** *Um bem é qualquer entidade ou meta desejada. Se uma pessoa tiver preferência por não ter nenhuma meta, essa entidade é um bem. Seus bens podem diferir das outras pessoas. Apesar desta possível diferença de opinião, o termo bem significa não mais do que alguma pessoa - como ele julga a sua situação - prefere ter. Tudo isso pode ser compactamente sumarizado ao dizer que um bem proporciona utilidade a uma pessoa que o consome.*

Postulado 1.1.2 (Para cada pessoa, alguns bens são escassos)**.** *É útil distinguir dois tipos de bens: um bem "livre" e um bem "econômico". Bem econômico é o nome técnico para bem escasso. Um bem é escasso para alguém se, e somente se, essa pessoa prefere ter mais desse bem do que já tem. (...). Se um bem, não importa quão desejado, é tão abundante que a pessoa sequer queira mais dele, então esse bem é um bem livre (free good). Ambos os bens proporcionam utilidade para quem os tem, mas por definição, mais do bem abundante não adiciona utilidade. Se ocorrer de um bem ser abundante para todas as pessoas, então deixa de ser um bem econômico.*

Postulado 1.1.3 (Uma pessoa está disposta a sacrificar uma fração de bens para obter mais de outros bens)**.** *Para obter uma fração adicional de um bem, uma pessoa está disposta a sacrificar uma fração de um bem ou de um grupo de bens. Dito de outra forma, uma pessoa está disposta a sacrificar um pouco de um bem que ele já possui se ele puder obter um aumento suficiente nas quantidades dos outros bens desejados. Podemos supor que isso possa ser mensurado como uma razão entre quantidades pequenas, a taxa marginal de substituição. Logo, não é a abundância de um bem para alguém que gera excedentes de troca, mas essa abundância combinada com a escassez para outra pessoa. Portanto, a abundância, mesmo quando exista para alguma pessoa, não é suficiente para gerar valor. É preciso haver escassez, esta sim a condição necessária.*

Postulado 1.1.4 (A quantidade que uma pessoa está disposta a sacrificar depende do quanto possui do bem; quanto mais possui, menor o valor pessoal atribuído ao bem)**.** *O valor pessoal da substituição que uma pessoa está disposta a fazer não é inteiramente aleatório ou imprevisível. Embora isso dependa de muitas coisas - como experiência passada, educação, preferências - depende de maneira previsível das quantidades de bens que ele possui. Como pouco pode ser dito sobre como outros fatores afetam sua avaliação de substituição pessoal, nos concentraremos no efeito das quantias possuídas.*

Postulado 1.1.5 (Nem todas as pessoas possuem preferências idênticas)**.** *As pessoas não são semelhantes em todas as características. Isto significa que a mesma combinação de bens gera valorações diferentes.*

1.2 Conjunto de Consumo e Restrição Orçamentária

1.2.1 Conjunto de Consumo

Consideramos um consumidor confrontado com possíveis cestas de consumo no conjunto de consumo X. Geralmente, assumimos que X é o ortante não negativo em $\mathbb{R}^L$, como mostrado na Figura 1.1, mas conjuntos de consumo mais específicos podem ser usados.

Por exemplo, podemos permitir consumo de algum bem em um intervalo adequado como lazer, como mostrado na figura abaixo, ou podemos incluir apenas cestas que dariam ao consumidor pelo menos uma existência de subsistência ou que consiste apenas em unidades inteiras de consumos, como mostrado na Figura 1.2.

Assumimos que X é um conjunto fechado[1] e convexo[2], salvo indicação em contrário.

O conjunto X ser fechado significa que qualquer bem ou conjunto de bens está contido no conjunto de consumo. A convexidade de um conjunto de consumo significa que todo bem é divisível e pode ser consumido em unidades fracionárias.

[1] Um conjunto $X \subset \mathbb{R}$ é fechado se, e somente se, para toda sequência $\{x_k\}$ tal que $x_k \in S$ para todo k e $x_k \to x$, tem se que $x \in S$.

[2] Um conjunto $X \subset \mathbb{R}$ é dito convexo se qualquer combinação convexa de quaisquer dois pontos de X também está em X. Ou seja, se qualquer linha reta que liga dois pontos de X estiver completamente contida nele.

Figura 1.1 – Um conjunto de consumo que reflete o limite legal do número de horas de trabalho. À direita o conjunto de consumo no R^2_+

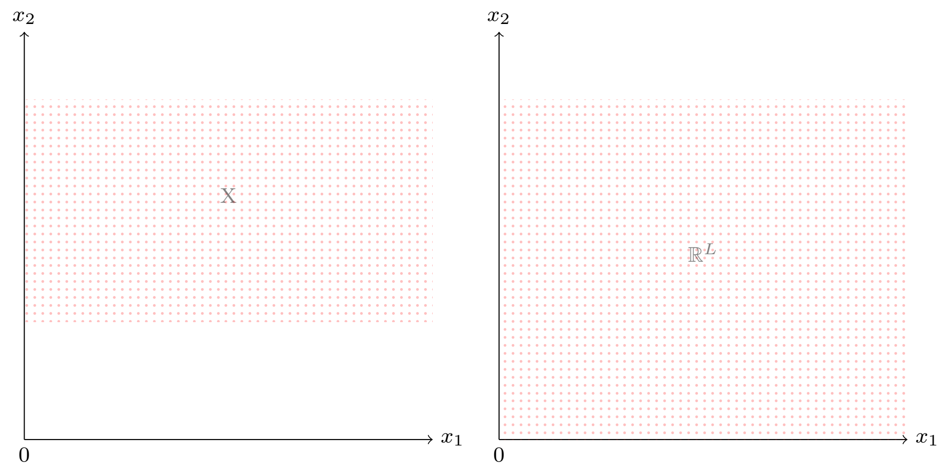
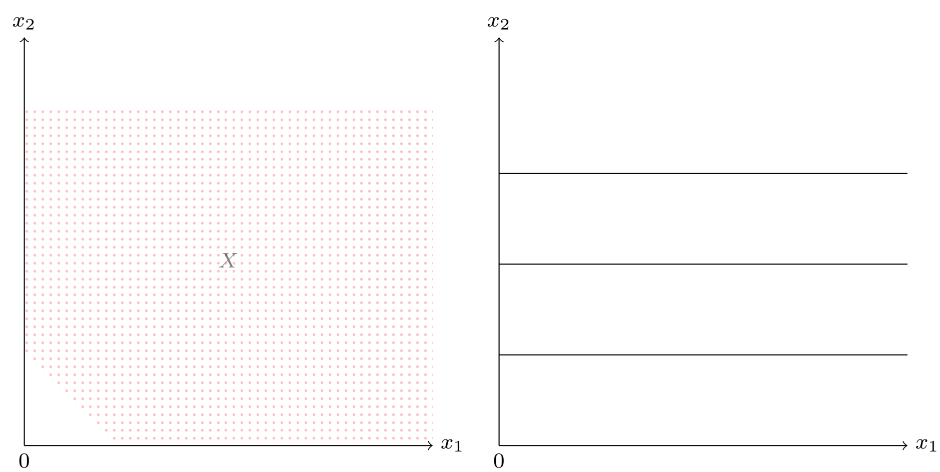


Figura 1.2 – À esquerda um conjunto de consumo que reflete as necessidades de sobrevivência. À direita um conjunto de consumo em que o bem 2 deve ser consumido em quantidades inteiras



1.2.2 Restrição Orçamentária

No problema básico da escolha do consumidor, nem todos as cestas de consumo são acessíveis em uma economia com recursos limitados, e um consumidor é limitado por sua riqueza. Em uma instituição de mercado, a riqueza pode ser determinada pelo valor de sua dotação inicial e/ou renda de participações acionárias de firmas. Presume-se que a renda ou a riqueza do consumidor seja fixa e que os preços das mercadorias não possam ser afetados pelo consumo do consumidor ao realizar a sua escolha. Seja m a quantia fixa de dinheiro disponível para um consumidor, e seja $\mathbf{p} = \begin{pmatrix} p_1 & \dots & p_L \end{pmatrix}$ o vetor de preços dos bens, $1, \dots, L$. Com $L = 2$, a restrição orçamental pode ser representada por uma reta, traduzindo a restrição orçamentária em uma equação do primeiro grau. Com $L = 3$ a representação geométrica é um plano. Mas, com L maior que três a restrição orçamentária torna-se o que é matematicamente conhecido por hiperplano ou plano multidimensional, o que não pode ser representado geometricamente. Foi Alfredo Marshall, no século XIX, que propôs uma solução para resolver o problema. Considera-se a escolha do consumidor entre um determinado bem, que denominamos x_1, e um conjunto de todos os outros bens, que denominamos por x_2, que é conhecido por bem compósito (composto). Por convenção considera-se que o preço unitário do bem composto é uma unidade monetária, o que permite pensar quanto ao bem composto como o rendimento do consumidor que resta depois de ter adquirido o bem x_1, ou seja, representa o rendimento gasto nos outros bens para além do bem x_1.

O conjunto de alternativas acessíveis é, portanto, apenas o conjunto de todas as cestas que satisfazem a restrição orçamentária do consumidor. O conjunto de cestas acessíveis, o conjunto orçamentário do consumidor, é dado por

$$B(p, m) = \{x \in X : px \leqq m\}, \tag{1.1}$$

em que **px** é o produto interno do vetor de preço e da cesta de consumo, ou seja, $px = \sum_{\ell=1}^{L} p_\ell x_\ell$, que é a soma das despesas das mercadorias a preços p. Observe que utilizamos o sinal $\leqq$ e não $=$ na restrição orçamentária, pois o indivíduo pode poupar. A razão, $\dfrac{p_\ell}{p_k}$, pode ser chamada de taxa econômica de substituição entre os bens i e k ou custo de oportunidade. Observe que multiplicar todos os preços e a renda por um número positivo não altera o conjunto orçamento.

Este é o conjunto orçamentário competitivo já que os preços não dependem da quantidade demandada. E isto é o que garante que a restrição orçamentária seja linear. Pode-se dizer que o conjunto orçamentário Walrasiano pressupõe implicitamente a existência de mercados eficientes e sem custos de transação. Quando essas hipóteses são relaxadas, surgem as restrições não lineares.

Com dois bens podemos escrever $p_1 x_1 + p_2 x_2 \leqq m$. Assim, a reta orçamentária é definida por

$$x_2 = \frac{m}{p_2} - \frac{p_1}{p_2} x_1. \tag{1.2}$$

Brevemente, consideremos os seguintes exemplos de restrições não lineares.

1. Numa economia de escambo, preços de compra e venda podem ser diferentes, pois há custos em encontrar pessoas que queiram comprar os bens que você quer vender, ou pessoas que queiram vender os bens que você quer comprar, isto é, existem custos de transação;

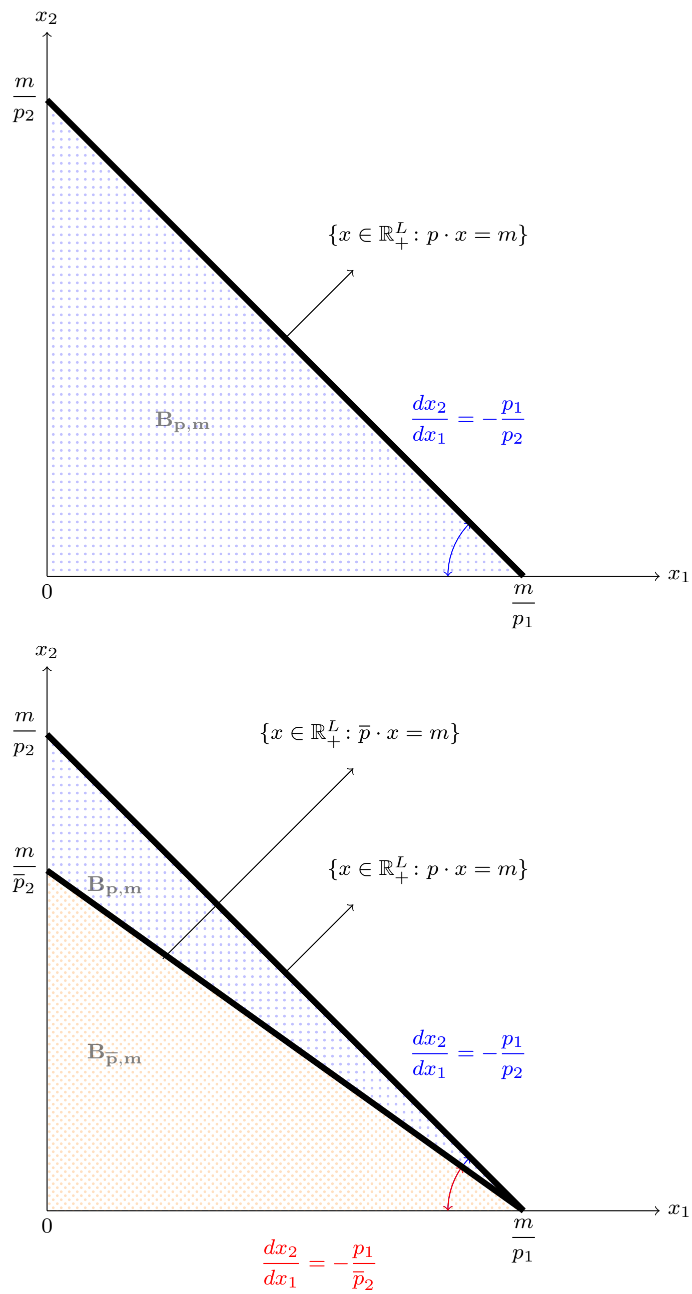
2. Um motivo para a existência de restrições não-lineares em economias monetizadas é a imposição de tarifas de duas partes, ou seja, mercados não são competitivos e existem custos de transação;

3. Problemas de escolha entre renda e lazer (i.e., oferta de trabalho) normalmente apresentam "quebras" na restrição orçamentária.

4. Escolha intertemporal quando o mercado de capitais é imperfeito (existem custos de transação).

5. Escolha social quando redistribuição afeta a estrutura de incentivos (mercados não competitivos e custos de transação).

Figura 1.3 – Restrição orçamentária



Assim, o orçamento definido reflete a capacidade objetiva do consumidor de comprar mercadorias e a escassez de recursos. Isso restringe significativamente as escolhas do consumidor. Para determinar as cestas de consumo ótimo, é necessário combinar a capacidade objetiva do consumidor de comprar várias mercadorias com o seu gosto subjetivo em várias cestas de consumo que são caracterizados pela noção de preferência ou utilidade.

Tabela 1.1 – Mudanças na restrição orçamentária

	$\frac{m}{p_1}$	$\frac{m}{p_2}$	$\frac{dx_1}{dx_2}$
Aumento em p_1	$\downarrow$	0	$\uparrow$
Aumento em p_2	0	$\downarrow$	$\downarrow$
Aumento em m	$\uparrow$	$\uparrow$	0
Redução em p_1	$\uparrow$	0	$\downarrow$
Redução em p_2	0	$\uparrow$	$\uparrow$
Redução em m	$\downarrow$	$\downarrow$	0

1.3 Preferências e Utilidade

1.3.1 Preferências

Preferências são caracterizadas de forma axiomática. Formalizam a ideia de que os consumidores podem escolher e que essas escolhas são consistentes em certo sentido. As preferências são representadas por uma relação binária $\succeq$, ou seja, uma regra que define subconjuntos específicos de $X \times X$.

Presume-se que o consumidor tenha preferências sobre as cestas de consumo em X, para que ele possa comparar e classificar vários bens disponíveis na economia. Quando escrevemos $x \succeq y$, queremos dizer "o consumidor pensa que a cesta x é tão boa quanto a cesta y". Queremos que as preferências ordenem o conjunto de cestas. A teoria econômica tem preocupação de analisar como os agentes escolhem as cestas. No entanto, só analisará um tipo específico de agente: aqueles que têm preferências racionais. O que significa ter preferências racionais? Um agente terá preferências racionais se atender a dois pressupostos.

1. Completude (completa): para todo x e y em X temos que ou $x \succeq y$, ou $y \succeq x$ ou ambos. Isso significa que o agente econômico deve sempre conseguir comparar duas cestas que lhes são apresentadas.

2. Transitividade: para todo x, y e z em X, se $x \succeq y$ e $y \succeq z$, então $x \succeq z$.

3. Reflexão: para todo x em X, $x \succeq x$ (é uma implicação da completeza desde que as cestas sejam definidas sem ambiguidade).

A relação binária definida no conjunto de consumo X é chamada uma relação de preferência racional se satisfizer os axiomas 1 e 2.

Vamos analisar apenas a transitividade (como equivalente da racionalidade) e seu real significado para a Economia. Em primeiro lugar, se há duas alternativas somente, a transitividade não tem sentido. Além disso, a completeza não garante transitividade quando há 3 alternativas ou mais.

O importante, porém, é que, com transitividade, o indivíduo é capaz de ordenar as suas alternativas vislumbradas desde a mais preferida até a menos preferida e fazê-lo de modo consistente. Se $x \succ y$ e $y \succ x$, isso é inconsistência. Se essa expressão fosse uma variável aleatória e se os indivíduos as expressassem frequentemente de acordo com essa aleatoriedade, então os indivíduos estariam frequentemente mudando de opinião quanto a duas alternativas, ou seja, seriam inconsistentes. Alguém poderia contra-argumentar que isso nem sempre é anormal, como quando a cada fim de semana o indivíduo muda de opinião sobre se vai ao cinema ou ao clube. Mas isso é incorreto, pois a escolha da sequência de passeios, caso se considere uma sequência intertemporal de ações, é feita no momento inicial, não a cada semana novamente. E se é feita a cada semana novamente (pensando-se apenas "naquela semana") e ela muda, então não cabe o problema da inconsistência, porque não há intertemporalidade. A inconsistência seria algo como isto: numa semana você diz "decidi que em todos os fins de semana daqui pra frente eu só irei ao cinema", mas na semana seguinte você diz "mudei de ideia, decidi que em todos os fins de semana daqui pra frente eu só irei ao clube". Aí na terceira semana você volta à primeira decisão e fica nisso o resto da vida. Isso é inconsistência. Fenômenos desse tipo podem ocorrer, mas se configuram em fatos esporádicos e correspondem a mudanças estruturais. Não é algo que ocorre todo dia com cada um, o tempo todo e em todas as instâncias relevantes e irrelevantes da vida.

A transitividade implica que um indivíduo consegue ordenar as alternativas: $x \succ y \succ z$. Ele escolhe x, mas ao escolher x, ele tem consciência das alternativas que foram sacrificadas. É graças a isso que ele pode atribuir à alternativa x a sua valoração marginal subjetiva, que é o valor da melhor alternativa sacrificada, a saber, y. E se, por alguma restrição institucional (lei, cultura, norma social etc.), ele fosse impedido de escolher x, então ele escolheria y e poderia atribuir a essa alternativa a sua valoração marginal, que é o valor da alternativa z que foi sacrificada. Racionalidade não implica conhecimento perfeito sobre potenciais decisões.

A relação binária $\succ$ representa $x \succ y$, isto é, x é estritamente preferível à y (ou é melhor do que). É definida da seguinte maneira:

$$x \succ y \Longleftrightarrow x \succeq y \quad \text{e} \quad y \not\succeq x. \tag{1.3}$$

A relação binária $\sim$ representa $x \sim y$, isto é, x é indiferente à y. É definida da seguinte maneira:

$$x \sim y \Longleftrightarrow x \succeq y \quad \text{e} \quad y \succeq x. \tag{1.4}$$

Tome qualquer cesta $x_0 \in X$. Definimos, então, os seguintes conjuntos:

$$\begin{aligned}
\succeq (x_0) &\equiv \{x|x \in X, x \succeq x_0\} && \text{cestas “pelo menos tão boas quanto } x_0\text{”}\\
\preceq (x_0) &\equiv \{x|x \in X, x \preceq x_0\} && \text{cestas “não melhores do que } x_0\text{”}\\
\succ (x_0) &\equiv \{x|x \in X, x \succ x_0\} && \text{cestas “melhores do que } x_0\text{”}\\
\prec (x_0) &\equiv \{x|x \in X, x \prec x_0\} && \text{cestas “piores do que } x_0\text{”}\\
\sim (x_0) &\equiv \{x|x \in X, x \sim x_0\} && \text{cestas “indiferentes a } x_0\text{”}
\end{aligned}$$

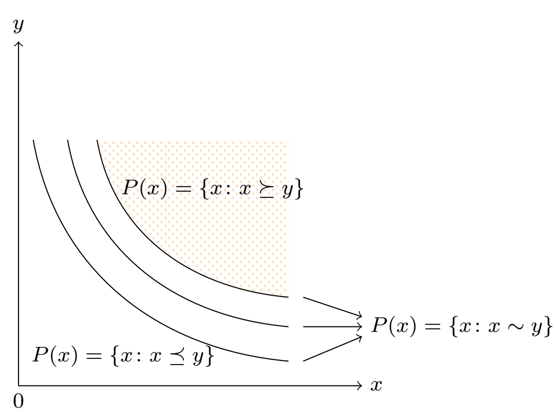
Os conjuntos $\succ (x_0)$, $\prec (x_0)$ e $\sim (x_0)$ particionam o conjunto X. Ou seja:

1. $\succ (x_0) \cap \prec (x_0) = \emptyset$
2. $\succ (x_0) \cap \sim (x_0) = \emptyset$
3. $\prec (x_0) \cap \sim (x_0) = \emptyset$
4. $\succ (x_0) \cup \prec (x_0) \cup \sim (x_0) = X$

Dada uma ordenação (relação) de preferências, geralmente mostramos isso graficamente, como mostra a figura abaixo. O conjunto de todos as cestas de consumo que são indiferentes umas as outras é chamado de curva de indiferença. Para um caso de dois bens, a inclinação de uma curva de indiferença em um ponto mede a taxa marginal de substituição entre os bens x e y. Para um caso L-dimensional, a taxa marginal de substituição entre dois bens é a inclinação de uma superfície de indiferença, medida em uma direção particular.

Para uma determinada cesta de consumo y, seja $P(x) = \{x \in X : x \succeq y\}$ o conjunto de todas as cestas na curva de indiferença acima de y, chamado de conjunto de contorno superior definido em y. Seja $P_s(x) = \{x \in X : x \succ y\}$ o conjunto de todas as cestas acima da curva de indiferença em y, chamado de conjunto de contorno estritamente superior definido em y. Seja $L(x) = \{x \in X : x \preceq y\}$ o conjunto de todos as cestas sobre ou abaixo da curva de indiferença através de y e é chamado de conjunto de contorno inferior definido em y. Seja $L_s(x) = \{x \in X : x \prec y\}$ o conjunto de todos as cestas sobre ou abaixo da curva de indiferença através de y e é chamado de conjunto de contorno estritamente inferior definido em y.

Figura 1.4 – Preferências em duas dimensões



Axiomas adicionais garantem que as preferências sejam bem comportadas.

- Continuidade: para todos os x em X, os conjuntos de contorno superior e inferior $P(x)$ e $L(x)$ são fechados. Segue que os conjuntos de contorno estritamente superior e inferior, $P_s(x)$ e $L_s(x)$, são conjuntos abertos.

  Essa suposição é necessária para descartar certo comportamento descontínuo. Ele diz que se (x_i) é uma sequência de cestas de consumo que são todas pelo menos tão boas quanto uma cesta y, e se essa sequência converge para alguma cesta x^*, então x^* é pelo menos tão boa quanto y. A consequência mais importante da continuidade é esta: se y é estritamente preferido a z e se x é uma cesta que é próxima o suficiente de y, então x deve ser estritamente preferido a z.

Exemplo 1.3.1. *As preferências lexicográficas (ou ordenações lexicográficas) ocorrem quando um consumidor prefere, em absoluto, um bem a outro qualquer. Neste caso um conjunto de bens contendo um bem lexicográfico é preferível a qualquer outro, sendo, por conseguinte, a substituibilidade nula.*

As preferências lexicográficas descrevem preferências comparativas em que um agente econômico prefere qualquer quantidade de um bem (X) a qualquer quantidade de outro (Y). Especificamente, se forem oferecidos vários pacotes de mercadorias, o agente escolherá o pacote que oferecer mais X, não importando o quanto exista de Y. Somente quando houver um empate entre as cestas em relação ao número de unidades de X, o agente começará a comparar o número de unidades de Y entre as cestas.

Por exemplo, se para uma determinada cesta (X, Y, Z) *um agente ordena suas preferências de acordo com a regra $X \gg Y \gg Z$, então as cestas* $(5,3,3), (5,1,6), (3,5,3)$ *seriam ordenadas, da mais para a menos preferida:*

$$5,3,3$$
$$5,1,6$$
$$3,5,3$$

Embora a primeira opção contenha menos bens totais do que a segunda opção, ela é preferida porque tem mais Y. Observe que o número de X é o mesmo e, portanto, o agente está comparando Y. Mesmo que a terceira opção tenha o mesmo total de mercadorias que a primeira opção, a primeira opção ainda é preferida porque tem mais X. Mesmo que a terceira opção tenha muito mais Y do que a segunda opção, a segunda opção ainda é preferida porque tem mais X.

Formalizando, de fato, $\forall n \in N$, $\left(\frac{1}{n}, 0\right) \succeq (0,1)$, porém $\lim_{n\to\infty} \frac{1}{n} = (0,0) \prec (0,1)$.

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

Definição 1.3.1 (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.

Definição 1.3.2 (Monotonicidade)**.** *Se $x > y$, então $x \succ y$.*

Ou seja, é preferível mais a menos quando um bem é desejável.

Definição 1.3.3 (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 seguinte:

Definição 1.3.4 (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 $y \succ x$.*

Definição 1.3.5 (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.

Nós agora vamos ver vários tipos de propriedades de convexidade usadas na teoria do consumidor.

Definição 1.3.6 (Convexidade estrita)**.** *Dados $x', x \in X$ tal que $x' \succ x$, então segue que $tx + (1-t)x' \succ x$ com $0 < t < 1$.*

Definição 1.3.7 (Convexidade)**.** *Dados $x', x \in X$ tal que $x' \succ x$, então segue que $tx + (1-t)x' \succ x$ com $0 \leqq t < 1$.*

Definição 1.3.8 (Convexidade fraca)**.** *Dados $x', x \in X$ tal que $x' \succeq x$, então segue que $tx + (1-t)x' \succeq x$ com $0 \leqq t \leqq 1$.*

A convexidade das preferências implica que as pessoas desejam diversificar seus consumos (o consumidor prefere médias a extremos) e, assim, a convexidade pode ser vista como a expressão formal da medida básica para a diversificação. Observe que as preferências convexas podem ter curvas de indiferença que exibem"pontos planos", enquanto as preferências estritamente convexas têm curvas de indiferença estritamente rotundas. A convexidade estrita de $\succ_i$ implica a hipótese neoclássica de taxa marginal de substituição decrescente entre quaisquer dois bens, como mostrado na figura abaixo. Neste exemplo, vemos que $\left.\frac{dx_2}{dx_1}\right|_x > \left.\frac{dx_2}{dx_1}\right|_{x'}$.

Figura 1.6 – Curva de indiferença estritamente convexa

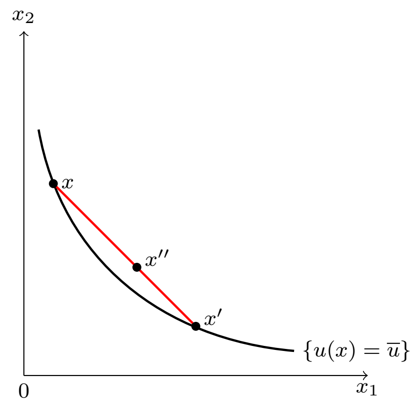


Figura 1.5 – Curvas de indiferença convexas, mas não estritamente convexas

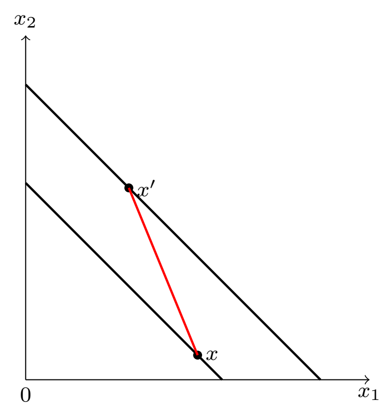


Figura 1.7 – Curvas de indiferença espessas são fracamente convexas, mas não convexas

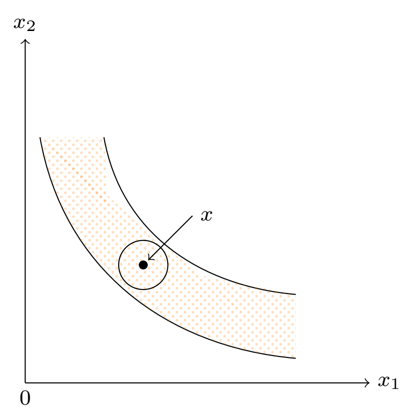
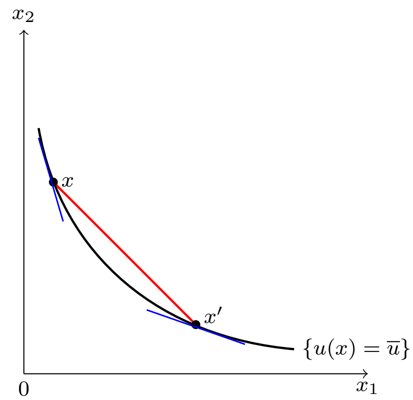


Figura 1.8 – A TAXA MARGINAL DE SUBSTITUIÇÃO É DECRESCENTE QUANDO O CONSUMO DO BEM 1 AUMENTA



Agora acrescentamos mais um elemento a nossa teoria da escolha: a hipótese comportamental: consumidores "racionais" escolhem a melhor (de acordo com suas ordenações de preferências) cesta x^* factível (i.e., dentro do conjunto orçamentário B):

$$x^* \in B \text{ tal que } x^* \succeq x \text{ para todo } x \in B. \tag{1.5}$$

Chamaremos o problema acima de "o problema do consumidor". A primeira pergunta relevante é: o problema do consumidor tem solução quando $B(p, m) \equiv \{x \in X : px \leqq m\}$? Sim, quando as preferências são contínuas. A solução é única? Por convexidade estrita, sim.

1.3.2 A Função Utilidade

Às vezes é mais fácil trabalhar diretamente com a relação de preferência e seus conjuntos associados. Mas outras vezes, especialmente quando você quer usar métodos de cálculo, é mais fácil trabalhar com preferências que podem ser representadas por uma função de utilidade; isto é, uma função $u : X \to \mathbb{R}$ tal que $x \succeq y$ se e somente se $u(x) \geqq u(y)$.

Isto é, a função u atribui o número x que é pelo menos tão grande quanto o número y se e somente se x é pelo menos tão bom quanto y. O interessante das funções de utilidade é que, se você conhece a função de utilidade que representa as preferências do consumidor, pode analisar essas preferências derivando as propriedades da função de utilidade. E como matemática é basicamente projetada para derivar propriedades de funções, isso pode nos ajudar a dizer muito sobre preferências.

Observe que os números atribuídos às curvas de indiferença na definição da função de utilidade são arbitrários. Qualquer atribuição de números funciona, desde que a ordem dos números atribuídos a várias cestas não seja perturbada. Assim, se multiplicássemos todos os números por 2, ou adicionássemos 6 a eles, ou tirássemos a raiz, os números atribuídos às curvas de indiferença após a transformação ainda representariam as mesmas preferências. Como a característica crucial de uma função de utilidade é a ordem dos números atribuídos a várias cestas, mas não as próprias cestas, dizemos que a utilidade é um conceito ordinal.

Nem todas as preferências podem ser representadas por funções de utilidade, mas pode ser mostrado que qualquer ordenação de preferência contínua pode ser representada por uma função de utilidade contínua.

Teorema 1.3.1 (Existência de uma função de utilidade)**.** *Suponha que as preferências sejam completas, reflexivas, transitivas, contínuas e fortemente monótonas. Então existe uma função de utilidade contínua* $u\colon \mathbb{R}^L_+ \to \mathbb{R}$ *que representa essas preferências.*

O papel da função de utilidade é registrar a ordem de preferência subjacente. Os valores numéricos atuais de u não têm significado: somente o sinal da diferença no valor entre dois pontos é significativo. Assim, uma função de utilidade é frequentemente uma maneira muito conveniente de descrever preferências, mas não deve ser dada qualquer interpretação psicológica. A única característica relevante de uma função de utilidade é seu caráter ordinal. Especificamente, podemos mostrar que uma função de utilidade é única de uma transformação arbitrária e estritamente crescente.

Teorema 1.3.2 (Invariância da função de utilidade a transformações monotônicas)**.** *Se* $u(x)$ *representa alguma preferência* $\succeq$ *e* $f\colon \mathbb{R} \to \mathbb{R}$ *é estritamente monotônica, então* $f(u(x))$ *representará exatamente as mesmas preferências.*

Corolário 1.3.1. *Disso decorre que se existe pelo menos uma função utilidade que representa as preferências, existem infinitas, pois funções utilidade são invariantes em relação a transformações monotônicas.*

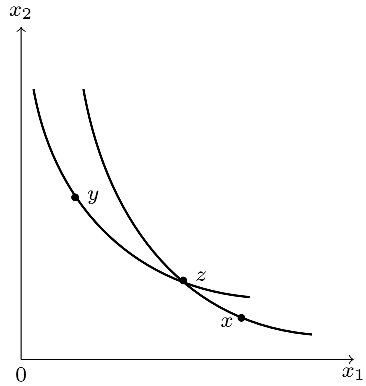
Este teorema de invariância é útil em muitos aspectos. Por exemplo, como será mostrado, podemos usá-lo para simplificar o cálculo da função de demanda a partir da maximização da função utilidade.

Se as preferências de uma consumidor são racionais então, para quaisquer duas cestas de bens x e y:

1. Se $x \sim y$, então $CI_x = CI_y$
2. Caso contrário, $CI_x \cap CI_y = \emptyset$

Isso significa que, se duas curvas de indiferença têm um ponto em comum, elas terão todos os pontos em comum. Portanto, duas curvas de indiferença distintas não se cruzam.

Figura 1.9 – Curvas de indiferença não podem se cruzar



Podemos representar uma função de utilidade por meio das suas curvas de nível, isto é, cortes no plano $z = k$. No caso da função $U(x_1, x_2) = x_1 x_2$ fazemos

$$x_1 x_2 = k \quad \longrightarrow \quad x_2 = \frac{k}{x_1}. \tag{1.6}$$

Este resultado mostra que o ortante não-negativo é um mapa denso de funções de utilidade. A seguir, damos alguns exemplos de funções de utilidade.

Exemplo 1.3.2 (Função de utilidade Cobb-Douglas)**.** *Uma função de utilidade que é usada frequentemente para fins ilustrativos e empíricos é a função de utilidade Cobb-Douglas,*

$$u(x_1, \ldots, x_L) = x_1^{a_1} x_2^{a_2} \ldots x_L^{a_L}, \tag{1.7}$$

com $a_\ell > 0, \ell = 1, \ldots, L$. Esta função de utilidade representa uma ordenação de preferência que é contínua, estritamente monotônica e estritamente convexa em $\mathbb{R}^L_{++}$.

Figura 1.10 – Função de utilidade Cobb-Douglas: $U = x_1^{0.8} x_2^{0.8}$

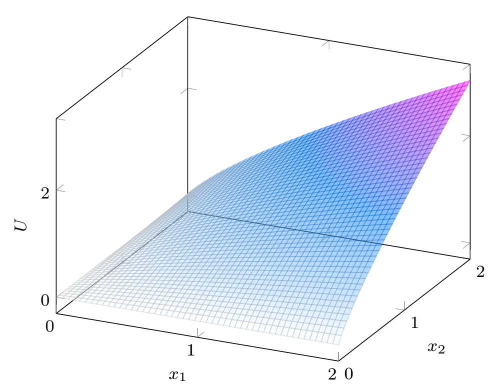
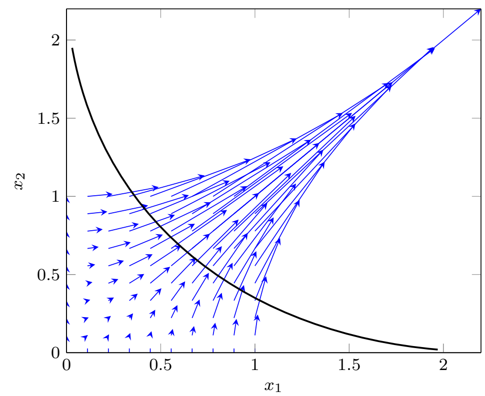


Figura 1.11 – Curva de indiferença da função de utilidade Cobb-Douglas e seu gradiente



Exemplo 1.3.3 (Função de utilidade linear)**.** *Uma função de utilidade que descreve bens substitutos perfeitos entre si é a função de utilidade linear*

$$u(x_1, \ldots, x_L) = a_1 x_1 + a_2 x_2 + \ldots + a_L x_L, \tag{1.8}$$

com $a_\ell \geqq 0, \ell = 1, \ldots, L$ *e* $a_\ell > 0$ *para no mínimo um* ℓ*.*

Esta função de utilidade representa uma ordenação de preferência que é contínua, monotônica e convexa em $\mathbb{R}^L_{++}$*.*

Figura 1.12 – Função de utilidade linear: $U = 5x_1 + 6x_2$

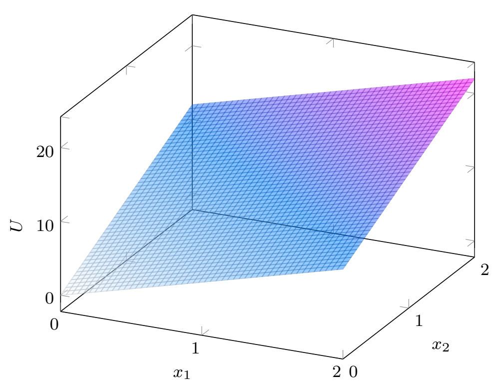
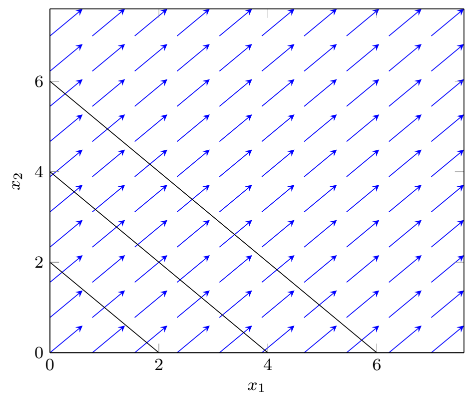


Figura 1.13 – Curva de indiferença da função de utilidade linear e seu gradiente



Exemplo 1.3.4 (Função de utilidade Leontief)**.** *Uma função que descreve bens complementares perfeitos é uma função de utilidade Leontief:*

$$u(x_1, \ldots, x_L) = A \min\{a_1 x_1, a_2 x_2, \ldots, a_L x_L\}, \tag{1.9}$$

com $A > 0$, $a_\ell \geqq 0, \ell = 1, \ldots, L$ e $a_\ell > 0$ para no mínimo dois ℓ. Esta função de utilidade representa uma ordenação de preferência em que os bens devem ser consumidos conjuntamente para aumentar a utilidade. É contínua, monotônica e convexa em $\mathbb{R}^L_+$.

Figura 1.14 – Função de utilidade Leontief: $U = \min\{5x_1, 6x_2\}$

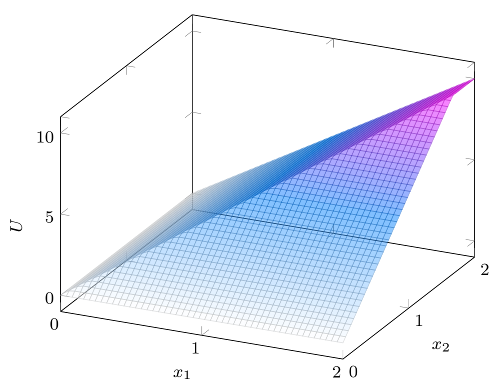
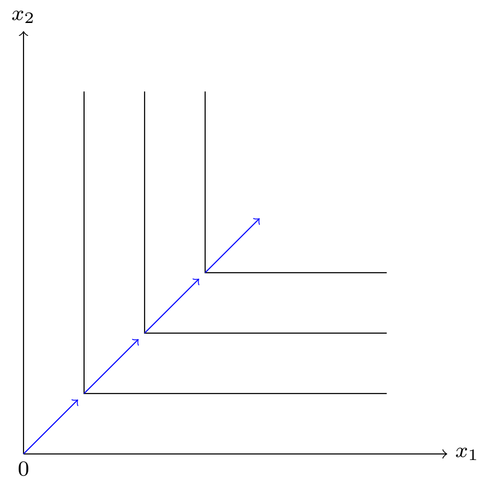


Figura 1.15 – Curva de indiferença da função de utilidade Leontief e seu gradiente



Exemplo 1.3.5 (Função de utilidade CES)**.** *Uma função de utilidade que é usada frequentemente para fins ilustrativos e empíricos é a função de utilidade com elasticidade de substituição constante,*

$$\begin{aligned} u(x_1, \ldots, x_L) &= \gamma \left[a_1 x_1^{\left(\frac{\sigma-1}{\sigma}\right)} + a_2 x_2^{\left(\frac{\sigma-1}{\sigma}\right)} + \ldots + a_L x_L^{\left(\frac{\sigma-1}{\sigma}\right)} \right]^{\frac{\sigma}{\sigma-1}} \\ &= \gamma \left[a_1 x_1^{-\rho} + a_2 x_2^{-\rho} + \ldots + a_L x_L^{-\rho} \right]^{-\frac{1}{\rho}}, \end{aligned} \tag{1.10}$$

com $\sum_L a_L = 1$*,* $\sigma \neq 0$*,* $\sigma < 1$ *e* $\rho > -1$*.*

Esta função de utilidade representa uma ordenação de preferência que é contínua, estritamente monotônica e estritamente convexa em $\mathbb{R}^L_{++}$*.*

Figura 1.16 – Função de utilidade CES: $U = 0.8 \left(0.4x_1^{-2} + 0.6x_2^{-2}\right)^{-0.5}$

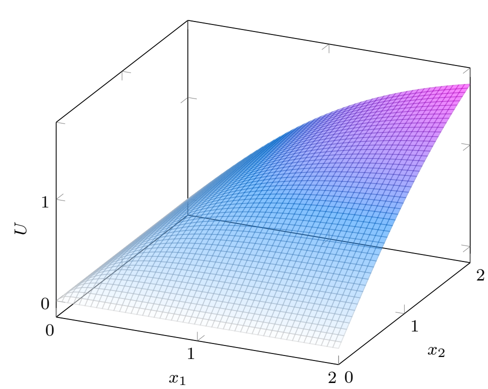
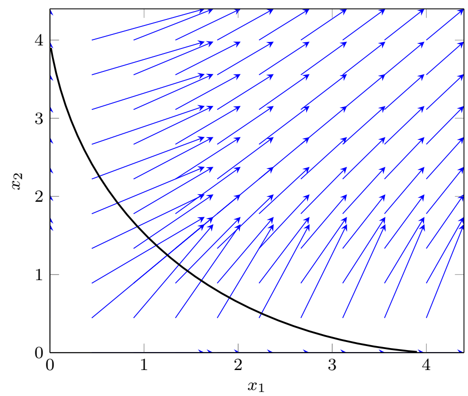


Figura 1.17 – Curva de indiferença da função de utilidade CES e seu gradiente



Exemplo 1.3.6 (Função de utilidade com preferências aditivas (côncava))**.** *Outra função de utilidade que é usada frequentemente para fins ilustrativos e empíricos é a função de utilidade com preferências aditivas*

$$u(x_1, \ldots, x_L) = v_1(x_1) + v_2(x_2) + \ldots + v_L(x_L). \tag{1.11}$$

Esta função de utilidade representa uma ordenação de preferência que é contínua, não é monotônica e estritamente côncava em $\mathbb{R}^L$*.*

Figura 1.18 – Função de utilidade com preferências aditivas: $U = x_1^2 + x_2^2$

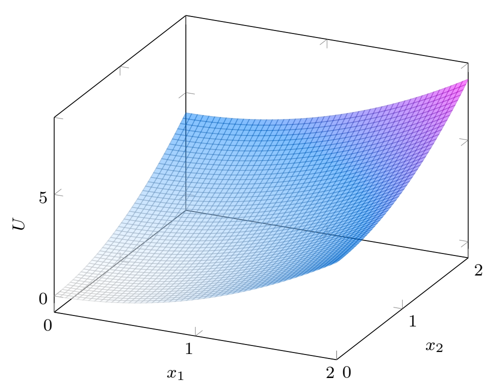
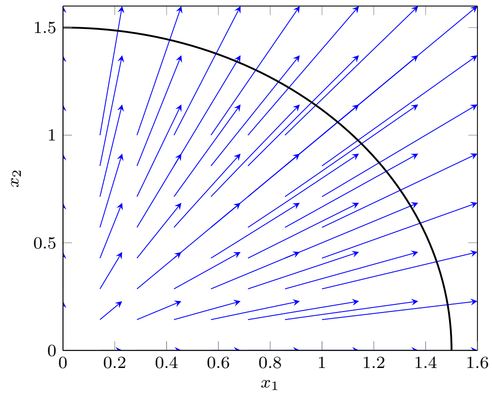


Figura 1.19 – Curva de indiferença da função de utilidade com preferências aditivas e seu gradiente



Exemplo 1.3.7 (Função de utilidade bliss point)**.** *Outra função de utilidade que é usada frequentemente para fins ilustrativos e empíricos é a função de utilidade com preferências aditivas*

$$u(x_1, \ldots, x_L) = a_1(x_1 - c)^2 + a_2(x_2 - c)^2 + \ldots + a_L(x_L - c)^2, \tag{1.12}$$

com $a_\ell < 0, \ell = 1, \ldots, L$. *Esta função de utilidade representa uma ordenação de preferência que é contínua, não é monotônica e é convexa.*

Estamos assumindo que os dois bens estudados são sempre considerados pelo indivíduo como sendo bens, e não maus - que o indivíduo sempre preferirá mais de cada um. Você pode estar se perguntando o que acontece se um ou ambos os produtos forem ruins ou se tornarem ruins em algum momento. Suponha, por exemplo, que o indivíduo possa ter "muito" de um ou ambos os bens - isto é, fica satisfeito com um bem e, a partir daí, o bem se torna ruim. Como isso afeta nossa representação gráfica das preferências? Como vemos abaixo, depende muito se o indivíduo realmente consumiu os bens, mesmo quando os considera maus. Devemos primeiro presumir que o indivíduo é de fato forçado a consumir os males - então discutiremos como nossa análise muda se esse não for o caso. Consideramos aqui apenas uma possibilidade - em que o indivíduo gosta de ambos os produtos até o nível c, mas se consumiu mais de c unidades de um bem, então esse bem se torna ruim. Então suas curvas de indiferença se tornam concêntricas.

Figura 1.20 – Função de utilidade bliss point: $U = -2(x_1 - 1)^2 - 2(x_2 - 1)^2$

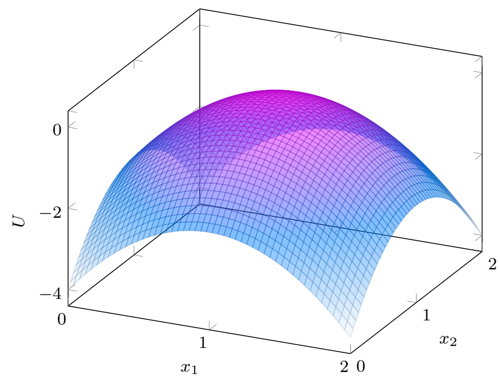
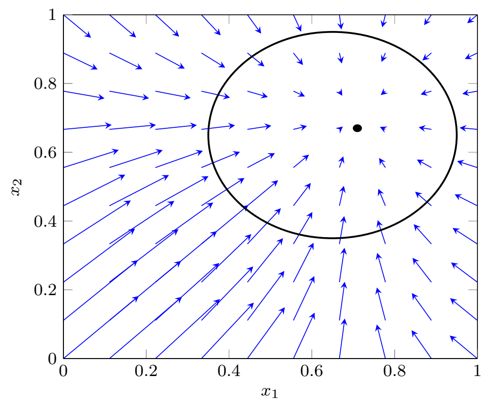


Figura 1.21 – Curva de indiferença da função de utilidade com preferências bliss point e seu gradiente



Exemplo 1.3.8 (Função de utilidade quase linear)**.** *Um agente tem preferências quase lineares se elas podem ser representadas por uma função da forma*

$$u(y, x_1, \ldots, x_L) = y + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \ldots + \alpha_\ell \ln x_L. \tag{1.13}$$

em que $\alpha_\ell > 0, \ell = 1, \ldots, L$ *e* $\sum_{\ell=1}^{L} \alpha_\ell = 1$. *Esta função de utilidade representa uma ordenação de preferência que é contínua, monotônica e é convexa se e somente se* $v(x_\ell)$ *é uma função côncava.*

Figura 1.22 – Função de utilidade quase-linear: $U = y + \ln x_1$

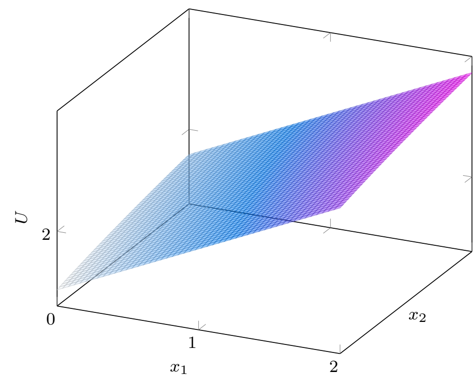
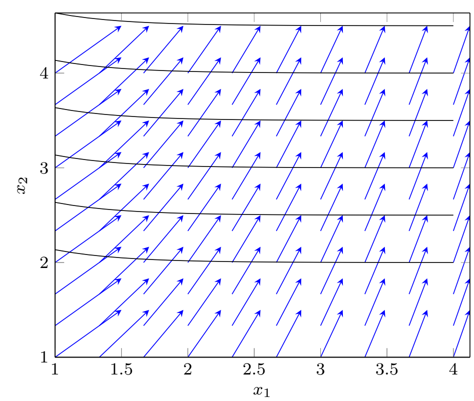
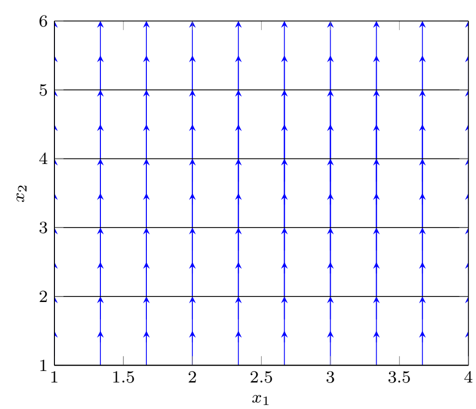


Figura 1.23 – Curva de indiferença da função de utilidade quase linear e seu gradiente



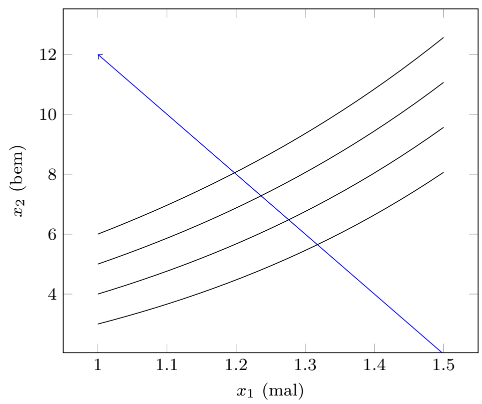
Exemplo 1.3.9 (Bens neutros). *Um bem é dito um neutro para um consumidor caso, este seja indiferente entre quaisquer duas cestas de bens que contenham quantidades diferentes desse bem e quantidades iguais de todos os outros bens.*

Figura 1.24 – Bem 1 é um neutro



Exemplo 1.3.10 (Males). *Um consumidor considera um serviço ou objeto físico, i, um mal caso, para quaisquer duas cestas de bens x, $y \in X$, se as duas cestas de bens contém as mesmas quantidade de todos os bens exceto o bem i, então ela prefere a cesta de bens com a menor quantidade do bem i. Mais formalmente, o bem i é um mal, quando e apenas quando, para quaisquer cestas x, $y \in X$ se $x_j = y_j$ para todo $j \neq i$, $j \in \{1, 2, \ldots, L\}$, $x \succeq y$ se, e somente se, $x_i \leq y_i$.*

Figura 1.25 – Curva de indiferença de um bem e de um mal



Podemos notar que a função de utilidade com elasticidade de substituição constante pode ser vista como um caso geral. Veja que:

$$\begin{aligned}
u &= \gamma \left[a_1 x_1^{-\rho} + a_2 x_2^{-\rho} \ldots + a_L x_L^{-\rho}\right]^{-\frac{1}{\rho}} \\
\left(\frac{u}{\gamma}\right)^{-\rho} &= a_1 x_1^{-\rho} + a_2 x_2^{-\rho} \ldots + a_L x_L^{-\rho} \\
-\rho \left(\frac{u}{\gamma}\right)^{-\rho-1} du &= -\rho a_1 x_1^{-\rho-1} dx_1 - \rho a_2 x_2^{-\rho-1} dx_2 - \ldots - \rho a_L x_L^{-\rho-1} dx_L \\
du &= a_1 \left(\frac{u}{\gamma x_1}\right)^{1+\rho} dx_1 + a_2 \left(\frac{u}{\gamma x_2}\right)^{1+\rho} dx_2 + \ldots + a_L \left(\frac{u}{\gamma x_L}\right)^{1+\rho} dx_L.
\end{aligned}$$

1. Se $\rho \to 0$

$$\begin{aligned}
\lim_{\rho \to 0} du &= a_1 \left(\frac{u}{\gamma x_1}\right)^{1+\rho} dx_1 + a_2 \left(\frac{u}{\gamma x_2}\right)^{1+\rho} dx_2 + \ldots + a_L \left(\frac{u}{\gamma x_L}\right)^{1+\rho} dx_L \\
&= a_1 \left(\frac{u}{\gamma x_1}\right) dx_1 + a_2 \left(\frac{u}{\gamma x_2}\right) dx_2 + \ldots + a_L \left(\frac{u}{\gamma x_L}\right) dx_L \\
\int \frac{du}{u} &= \int a_1 \left(\frac{1}{\gamma x_1}\right) dx_1 + \int a_2 \left(\frac{1}{\gamma x_2}\right) dx_2 + \ldots + \int a_L \left(\frac{1}{\gamma x_L}\right) dx_L \\
\int \frac{du}{u} &= a_1 \frac{1}{\gamma} \int \left(\frac{1}{x_1}\right) dx_1 + a_2 \frac{1}{\gamma} \int \left(\frac{1}{x_2}\right) dx_2 + \ldots + a_L \frac{1}{\gamma} \int \left(\frac{1}{x_L}\right) dx_L \\
\ln u &= a_1 \frac{1}{\gamma} \ln x_1 + a_2 \frac{1}{\gamma} \ln x_2 + \ldots + a_L \frac{1}{\gamma} \ln x_L
\end{aligned}$$

$$u = x_1^{a_1} x_2^{a_2} \dots x_L^{a_L}.$$

2. Se $\rho \to -1$:

$$\begin{aligned}
\lim_{\rho \to -1} du &= a_1 \left(\frac{u}{\gamma x_1}\right)^{1+\rho} dx_1 + a_2 \left(\frac{u}{\gamma x_2}\right)^{1+\rho} dx_2 + \dots + a_L \left(\frac{u}{\gamma x_L}\right)^{1+\rho} dx_L \\
&= a_1 dx_1 + a_2 dx_2 + \dots + a_L dx_L \\
\int du &= \int a_1 dx_1 + \int a_2 dx_2 + \dots + \int a_L dx_L \\
&= a_1 x_1 + a_2 x_2 + \dots + a_L x_L.
\end{aligned}$$

3. Se $\rho \to \infty$:

$$\lim_{\rho \to \infty} du = \min\{a_1 x_1, a_2 x_2, \dots, a_L x_L\}.$$

Também podemos usar a função de utilidade para encontrar a taxa marginal de substituição entre mercadorias. Seja $u(x_1, \dots, x_L)$ uma função de utilidade. Suponha que aumentemos a quantidade de bem i. Como o consumidor tem que mudar o consumo do bem j para manter a utilidade constante? Seja dx_i e dx_j os diferenciais de x_i e x_j. Por suposição, a mudança na utilidade deve ser zero, então

$$\frac{\partial u(x)}{\partial x_i} dx_i + \frac{\partial u(x)}{\partial x_j} dx_j = 0 \quad \Rightarrow \quad \frac{dx_j}{dx_i} = -\frac{\frac{\partial u(x)}{\partial x_i}}{\frac{\partial u(x)}{\partial x_j}} \quad \Rightarrow \quad \frac{UM_{x_i}}{UM_{x_j}}, \tag{1.14}$$

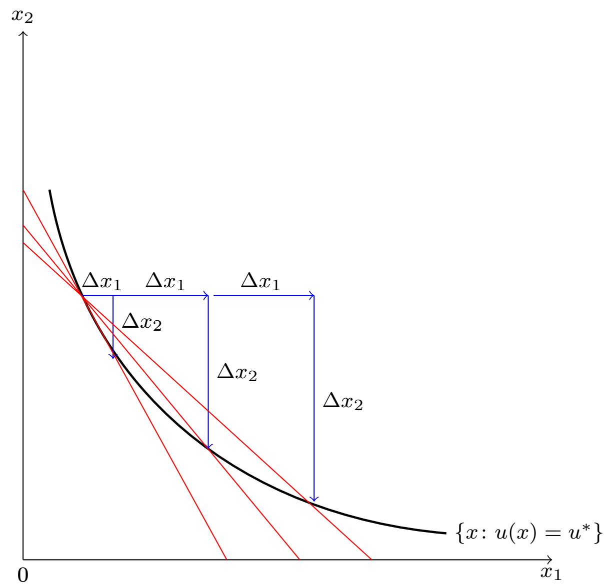
que dá a taxa marginal de substituição entre bens i e j é definida como a razão entre a utilidade marginal de x_i e a utilidade marginal de x_j.

Corolário 1.3.2. *A taxa marginal de substituição não depende da função de utilidade escolhida para representar as preferências subjacentes. Para provar isso, deixe $v(u)$ ser uma transformação monotônica da utilidade. A taxa marginal de substituição para esta função de utilidade é*

$$\frac{dx_j}{dx_i} = -\frac{\frac{\partial v}{\partial u(x)} \frac{\partial u(x)}{\partial x_i}}{\frac{\partial v}{\partial u(x)} \frac{\partial u(x)}{\partial x_j}} = -\frac{\frac{\partial u(x)}{\partial x_i}}{\frac{\partial u(x)}{\partial x_j}}. \tag{1.15}$$

A utilidade marginal é uma propriedade cardinal da função de utilidade. A utilidade marginal é medida em unidades de utilidade por unidades do bem 1. Se a unidade em que o bem i é medido é pequena, então a utilidade marginal pode ser interpretada como de quanto cresce o valor da função de utilidade quando o consumo desse bem é aumentado em uma unidade.

Figura 1.26 – Interpretação gráfica da taxa marginal de substituição



Se as preferências são fortemente monotônicas, então, para qualquer cesta de bens, todos os bens possuem utilidade marginal positiva (> 0). Se as preferências são fracamente monotônicas, então, para qualquer cesta de bens, todos os bens possuem utilidade marginal não negativa (≤ 0). O sinal da utilidade marginal é uma propriedade ordinal, pois não pode ser alterado por transformações monotônicas da função de utilidade.

Se o bem 1 é medido em unidades pequenas, a taxa marginal de substituição pode ser interpretada como, aproximadamente, a quantidade máxima do bem 1 da qual a consumidora está disposta a abrir mão para ter uma unidade adicional do bem 1. Alternativamente, ele pode ser interpretada como a quantidade mínima do bem 2 que é preciso dar à consumidora para que ela aceite reduzir seu consumo do bem 1 de uma unidade.

As propriedades importantes de uma ordenação de preferência podem ser facilmente verificadas examinando a função de utilidade. As propriedades são resumidas na seguinte proposição.

Proposição 1.3.1. *Seja $\succeq$ representada por uma função de utilidade $u\colon X \to \mathbb{R}$. Então:*

- *Uma ordenação é estritamente monotônica se e somente se u é estritamente monotônica.*
- *Uma ordenação é contínua se e somente se u é contínua.*
- *Uma ordenação é fracamente convexa se e somente se u é quase-côncava.*
- *Uma ordenação é fortemente convexa se e somente se u é estritamente quase-côncava.*

Teorema 1.3.3. *Seja uma função $u(x) \in \mathcal{C}^2$. Seja a seguinte matriz hessiana orlada:*

$$H = \begin{pmatrix} 0 & u_1 & u_2 & \cdots & u_L \\ u_1 & u_{11} & u_{12} & \cdots & u_{1L} \\ u_2 & u_{21} & u_{22} & \cdots & u_{2L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_L & u_{L1} & u_{L2} & \cdots & u_{LL} \end{pmatrix}. \tag{1.16}$$

Para uma função $u(x)$ ser quase-côncava no ortante não-negativo, é necessário que:

$$|H_L| = \begin{cases} \leq 0 & \textit{se } L \textit{ é par} \\ \geq 0 & \textit{se } L \textit{ é ímpar} \end{cases} \tag{1.17}$$

e é suficiente que

$$|H_L| = \begin{cases} < 0 & \textit{se } L \textit{ é par} \\ > 0 & \textit{se } L \textit{ é ímpar} \end{cases} \tag{1.18}$$

Exemplo 1.3.11. *Seja a função de utilidade $u(x_1, x_2) = x_1^\alpha x_2^\beta$. É fácil verificar que $u(x_1, x_2)$ é estritamente quase-côncava. Veja que*

$$|H| = \begin{vmatrix} 0 & \alpha x_1^{\alpha-1} y^\beta & \beta x^\alpha x_2^{\beta-1} \\ \alpha x_1^{\alpha-1} x_2^\beta & \alpha(\alpha-1) x_1^{\alpha-2} x_2^\beta & \alpha\beta x_1^{\alpha-1} x_2^{\beta-1} \\ \beta x_1^\alpha x_2^{\beta-1} & \alpha\beta x_1^{\alpha-1} x_2^{\beta-1} & \beta(\beta-1) x_1^\alpha x_2^{\beta-2} \end{vmatrix}. \tag{1.19}$$

Observe que $H_2 = -\alpha x_1^{2\alpha-1} x_2^{2\beta} < 0$ e $H_3 = \alpha\beta(\alpha+\beta) x_1^{3\alpha-2} x_2^{3\beta-2} > 0$.

Podemos resolver este mesmo problema usando uma transformação monotônica $v(x_1, x_2) = \ln(u(x_1, x_2)) = \alpha \ln x_1 + \beta \ln x_2$.

$$|H| = \begin{vmatrix} 0 & \frac{\alpha}{x_1} & \frac{\beta}{x_2} \\ \frac{\alpha}{x_1} & -\frac{\alpha}{x_1^2} & 0 \\ \frac{\beta}{x_2} & 0 & -\frac{\beta}{x_2^2} \end{vmatrix}. \tag{1.20}$$

Observe que $H_2 = -\dfrac{\alpha^2}{x_1^2} < 0$ e $H_3 = \dfrac{\alpha\beta}{x_1^2 x_2^2}(\alpha+\beta) > 0$.

A quase-concavidade é uma condição mais fraca que a concavidade. Concavidade é uma suposição sobre como os números atribuídos às curvas de indiferença mudam conforme você se move para fora a partir da origem. Ele diz que o aumento na utilidade associado a um aumento na cesta diminui à medida que você se afasta da origem. Como tal, é um conceito cardinal. A quase-concavidade é um conceito ordinal. Ele fala apenas sobre a forma das curvas de indiferença, não os números atribuídos a elas. Pode-se mostrar que a concavidade implica quase-concavidade, mas uma função pode ser quase-concavidade sem ser côncava (você pode desenhar uma em duas dimensões). Acontece que, para os resultados sobre a maximização da utilidade, que vamos desenvolver mais tarde, tudo o que realmente precisamos é a quase-concavidade. Uma vez que a concavidade impõe restrições cardinais à utilidade (que é um conceito ordinal) e é mais forte do que precisamos para os nossos resultados de maximização, permanecemos com a suposição mais fraca da quase-concavidade.

As funções quase-côncavas preservam a quase concavidade sob qualquer transformação monotônica positiva.

As três propriedades são equivalentes:

Convexidade das preferências $\Longleftrightarrow$ Conjunto de contorno superior é convexo $\Longleftrightarrow$ $u(\cdot)$ é quase-côncava

1.3.3 Propriedades das Relações de Preferência

Vamos apresentar algumas propriedades das relações de preferências e suas implicações.

1. Homogeneidade: uma função de utilidade é homogênea de grau k se variar as quantidades de todos os bens por um fator comum $\alpha > 0$ produz um aumento no nível da utilidade por α^k. Isto é,

$$u(\alpha x_1, \ldots, \alpha x_L) = \alpha^k u(x_1, \ldots, x_L), \tag{1.21}$$

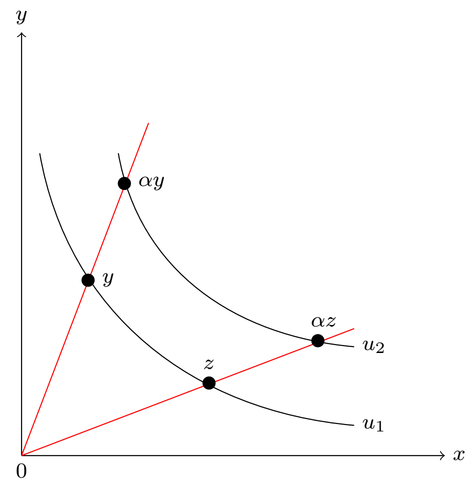
em que $\alpha > 0$. Dessa propriedade derivam outras três propriedades:

(a) A derivada de primeira ordem de uma função $u(\alpha x_1, \ldots, \alpha x_L)$ que é homogênea de grau k é homogênea de grau $k-1$, ou seja,

$$\begin{gathered}\frac{\partial u(\alpha x_1, \ldots, \alpha x_L)}{\partial x_i}\alpha = \alpha^k \frac{\partial u(x_1, \ldots, x_L)}{\partial x_i} \\ \frac{\partial u(\alpha x_1, \ldots, \alpha x_L)}{\partial x_i} = \alpha^{k-1} \frac{\partial u(x_1, \ldots, x_L)}{\partial x_i}.\end{gathered} \tag{1.22}$$

(b) As curvas de indiferença de funções homogêneas são expansões radiais umas das outras. Ou seja, se duas cestas y e z estão na mesma curva de indiferença, isto é, $u(y) = u(z)$, as cestas αy e αz também estão na mesma curva de indiferença, i.e., $u(\alpha y) = u(\alpha z)$.

Figura 1.27 – PREFERÊNCIAS HOMOGÊNEAS



(c) A taxa marginal de substituição de uma função homogênea é constante para todos os pontos ao longo de cada raio da origem. Ou seja, a inclinação da curva de indiferença no ponto y corresponde à inclinação de uma versão aumentada do ponto y, αy com $\alpha > 1$. A taxa marginal de substituição em $x = (x_i, x_j)$ é:

$$TMS_{i,j} = -\frac{\dfrac{\partial u(x_1, \ldots, x_L)}{\partial x_i}}{\dfrac{\partial u(x_1, \ldots, x_L)}{\partial x_j}}. \tag{1.23}$$

A taxa marginal de substituição em $x = (\alpha x_i, \alpha x_j)$ é:

$$\begin{aligned} TMS_{i,j} &= -\frac{\dfrac{\partial u(\alpha x_1, \ldots, \alpha x_L)}{\partial x_i}}{\dfrac{\partial u(\alpha x_1, \ldots, \alpha x_L)}{\partial x_j}} \\ &= -\frac{\alpha^{k-1}\dfrac{\partial u(x_1, \ldots, x_L)}{\partial x_i}}{\alpha^{k-1}\dfrac{\partial u(x_1, \ldots, x_L)}{\partial x_j}} \\ &= -\frac{\dfrac{\partial u(x_1, \ldots, x_L)}{\partial x_i}}{\dfrac{\partial u(x_1, \ldots, x_L)}{\partial x_j}}. \end{aligned} \tag{1.24}$$

Assim, a TMS não é afetada ao longo de todos os pontos atravessados por um raio a partir origem.

2. Homotética: uma função de utilidade $u(x)$ é homotética se for uma transformação monotônica

de uma função homogênea. Isto é,

$$u(x) = g(v(x)), \tag{1.25}$$

em que $g\colon \mathbb{R} \to \mathbb{R}$ é uma função estritamente crescente, e $v\colon \mathbb{R}^n \to \mathbb{R}$ é homogênea de grau k.

(a) Se $u(x)$ é homotética, e duas cestas y e z estão na mesma curva de indiferença, i.e., $u(y) = u(z)$, as cestas αy e αz também estão na mesma curva de indiferença, isto é, $u(\alpha y) = u(\alpha z)$ para todos $\alpha > 0$. Em particular,

$$u(\alpha y) = g(v(\alpha y)) = g(\alpha^k v(y)), \tag{1.26}$$

$$u(\alpha z) = g(v(\alpha z)) = g(\alpha^k v(z)), \tag{1.27}$$

(b) A taxa marginal de substituição de uma função homotética é homogênea de grau zero:

$$\begin{aligned} TMS_{i,j}u(\alpha x_1,\ldots,\alpha x_L) &= -\frac{\frac{\partial u(\alpha x_1,\ldots,\alpha x_L)}{\partial x_i}}{\frac{\partial u(\alpha x_1,\ldots,\alpha x_L)}{\partial x_j}} \\ &= -\frac{\frac{\partial g}{\partial v}\frac{\partial v(\alpha x_1,\ldots,\alpha x_L)}{\partial x_i}}{\frac{\partial g}{\partial v}\frac{\partial v(\alpha x_1,\ldots,\alpha x_L)}{\partial x_j}} \\ &= -\frac{\alpha^{k-1}\frac{\partial v(x_1,\ldots,x_L)}{\partial x_i}}{\alpha^{k-1}\frac{\partial v(x_1,\ldots,x_L)}{\partial x_j}} \\ &= -\frac{\frac{\partial v(x_1,\ldots,x_L)}{\partial x_i}}{\frac{\partial v(x_1,\ldots,x_L)}{\partial x_j}}. \end{aligned} \tag{1.28}$$

Ou seja,

$$TMS_{i,j}u(\alpha x_1,\ldots,\alpha x_L) = TMS_{i,j}u(x_1,\ldots,x_L). \tag{1.29}$$

(c) Uma relação de preferência em $x = \mathbb{R}^L$ é homotética se todos os conjuntos de indiferença estiverem relacionados a expansões proporcionais ao longo de raios. Isto é, se o consumidor é indiferente entre as cestas x e y, ou seja, $x \sim y$, ele também deve ser indiferente entre um escalonamento comum nessas duas cestas, ou seja, $\alpha x \sim \alpha y$, para todo $\alpha \geq 0$.

Para um dado raio desde a origem, a inclinação das curvas de indiferença (isto é, a TMS) que o raio atravessa coincide. A relação entre os dois bens $\frac{x_i}{x_j}$ permanece constante ao longo de todos os pontos no raio. Intuitivamente, a taxa na qual um consumidor está disposto a substituir um bem por outro depende apenas da taxa na qual ele consome os dois bens, ou seja, $\frac{x_i}{x_j}$, mas não depende do nível de utilidade que ele obtém.

3. Funções homogêneas são homotéticas, isto é, $u(x_1, \ldots, x_L) = g(v(u(x_1, \ldots, x_L))$. O inverso não é necessariamente válido, isto é, nem toda função homotética é homogênea.

Exemplo 1.3.12. *Seja a função homogênea de grau 1* $v(x_1, \ldots, x_L) = x_1 x_2 \ldots x_L$. *Aplique a transformação monotônica* $g(y) = y + a$, *em que* $a > 0$ *para obter a função homotética* $u(x_1, x_2, \ldots, x_L) = x_1 x_2 \ldots x_L + a$. *Esta função não é homogênea, já que aumentar todos os argumentos por* α *produz* $u(\alpha x_1, \alpha x_2, \ldots, \alpha x_L) = \alpha^L v(x_1, x_2, \ldots, x_L) + a$.

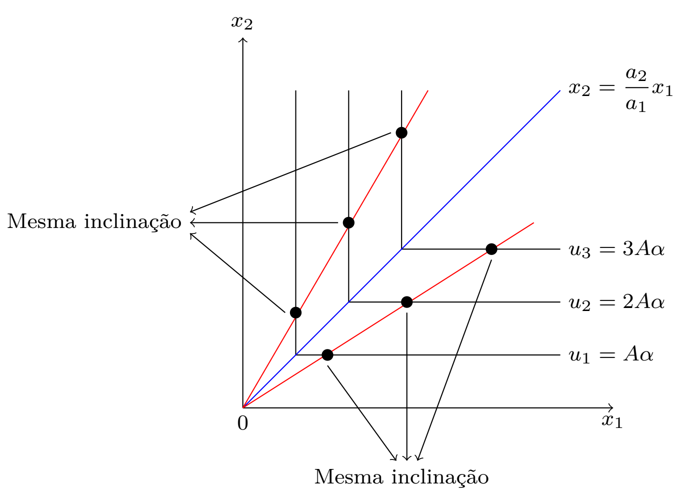
Vejamos exemplos de funções homotéticas.

Exemplo 1.3.13. *Seja a função de utilidade linear* $u(x_1, \ldots, x_L) = a_1 x_1 + a_2 x_2 + \ldots + a_L x_L$, *com* $a_\ell > 0, \ell = 1, \ldots, L$. *A taxa marginal de substituição entre os bens* i *e* j *é dada por* $TMS_{i,j} u(x_1, \ldots, x_L) = \frac{a_i}{a_j}$ *e* $TMS_{i,j} u(\alpha x_1, \ldots, \alpha x_L) = \frac{\alpha x_i}{\alpha x_j} = \frac{a_i}{a_j}$. *Logo, a função de utilidade linear é homotética.*

Exemplo 1.3.14. *Seja a função de Leontief* $u(x_1, \ldots, x_L) = A \min\{a_1 x_1, a_2 x_2, \ldots, a_L x_L\}$, *com* $A > 0$. *Nós não podemos definir a taxa marginal de substituição ao longo de todos os pontos das curvas de indiferenças. No entanto, a inclinação das curvas de indiferença coincide com os pontos em que essas curvas são cruzadas a partir de um raio da origem.*

Se conhecemos uma única curva de indiferença gerada por uma preferência homotética, somos capazes de descrever todas as curvas de indiferença geradas por essa preferência, pois toda as curvas de indiferença serão versões aumentadas ou diminuídas umas das outras.

Figura 1.28 – Complementares perfeitos e homoteticidade



	Fórmula	Intuição
Cobb-Douglas		
	$U = \prod_{\ell=1}^{L} x_\ell^{a_\ell}$	O formato multiplicativo implica que todos os bens devem ser consumidos conjunto (mesmo que em uma fração irrisória de algum bem)
	$0 < a_\ell < 1$	As preferências são incorporadas usando a
	$u = \ln U = \sum_{\ell=1}^{L} a_\ell \ln x_\ell$	U é a utilidade e u é uma transformação
	Monotônica: $\frac{\partial u}{\partial x_\ell} = \frac{a_\ell}{x_\ell} > 0$	A utilidade marginal de cada bem é positiva
	Concavidade: $\frac{\partial^2 u}{\partial x_\ell^2} = -\frac{a_\ell}{x_\ell^2} < 0$	A utilidade marginal de cada bem é decrescente
	Fortemente aditiva: $\frac{\partial^2 u}{\partial x_i x_j} = 0$	A utilidade marginal do bem i é independente da utilidade marginal do bem j
	Homotética: $\theta u = \sum_{\ell=1}^{L} a_\ell \ln(\theta x_\ell)$	A utilidade aumenta em uma proporção escalar se cada bem é multiplicado por um escalar

	Fórmula	Intuição
Linear		
	$u = \sum_{\ell=1}^{L} a_\ell x_\ell$	O formato de somatório implica que os bens são substitutos perfeitos
	$a_\ell \geq 0$ e no mínimo um $a_\ell > 0$	As preferências são incorporadas usando a
	Monotônica: $\dfrac{\partial u}{\partial x_\ell} = a_\ell > 0$	A utilidade marginal de cada bem é positiva e independente da quantidade consumida
	Concavidade: $\dfrac{\partial^2 u}{\partial x_\ell^2} = 0$	Essa função de utilidade não é estritamente quase-côncava
	Fortemente aditiva: $\dfrac{\partial^2 u}{\partial x_i x_j} = 0$	A utilidade marginal do bem i é independente da utilidade marginal do bem j
	Homotética: $\theta u = \sum_{\ell=1}^{L} a_\ell(\theta x_\ell)$	A utilidade aumenta em uma proporção escalar se cada bem é multiplicado por um escalar

	Fórmula	Intuição
Leontief		
	$u = \min\{a_1 x_1, \ldots, a_L x_L\}$	O formato da função implica que os bens são complementares perfeitos
	$a_\ell \geq 0$ e no mínimo dois $a_\ell > 0$	
	Monotônica: $\frac{\partial u}{\partial x_\ell}$ não funciona. A função não é diferenciável em $x_i = \frac{a_j}{a_i} x_j$	A utilidade marginal de cada bem é positiva: $\min\{x_1+, x_2+\} > \min\{x_1, x_2\}$
	Concavidade: $\frac{\partial^2 u}{\partial x_\ell^2} = 0$	Essa função de utilidade não é estritamente quase-côncava
	Fortemente aditiva: $\frac{\partial^2 u}{\partial x_i x_j} = 0$	A utilidade marginal do bem i é independente da utilidade marginal do bem j
	Homotética: $\theta u = \min\{\theta a_1 x_1, \ldots, \theta a_L x_L\}$	A utilidade aumenta em uma proporção escalar se cada bem é multiplicado por um escalar

	Fórmula	Intuição
Quase-Linear		
	$u = y + \sum_{\ell=1}^{L} a_\ell \ln x_\ell$	O formato é o caso de uma função de utilidade em que apenas a demanda pelo bem y depende da renda
	$a_\ell > 0$ e $\sum_{\ell=1}^{L} a_\ell = 1$	As preferências são incorporadas usando a
	Monotônica: $\dfrac{\partial u}{\partial x_\ell} = \dfrac{a_\ell}{x_\ell} > 0$ e $\dfrac{\partial u}{\partial y} = 1 > 0$	A utilidade marginal de cada bem x_ℓ é positiva e a utilidade marginal do bem y é constante
	Concavidade: $\dfrac{\partial^2 u}{\partial x_\ell^2} = -\dfrac{a_\ell}{x_\ell^2} < 0$	A utilidade marginal de cada bem x_ℓ é decrescente e a do bem y é constante
	Fortemente aditiva: $\dfrac{\partial^2 u}{\partial x_i x_j} = 0$	A utilidade marginal do bem i é independente da utilidade marginal do bem j
	Homotética: $\theta u \neq (\theta y) + \sum_{\ell=1}^{L} a_\ell \ln(\theta x_\ell)$	A utilidade não aumenta em uma proporção escalar se cada bem é multiplicado por um escalar

Fórmula	Intuição
CES	
$u = \left(\sum_{\ell=1}^{L} a_\ell x_\ell^{-\rho}\right)^{-\frac{1}{\rho}}$	O formato é o caso mais geral de uma função de utilidade com elasticidade de substituição igual a σ
$\rho = \frac{1-\sigma}{\sigma} < 1$	As preferências são incorporadas usando a e ρ
Monotônica: $\frac{\partial u}{\partial x_\ell} = a_\ell \left(\frac{u}{x_\ell}\right)^{(1+\rho)} > 0$	A utilidade marginal de cada bem é positiva
Concavidade: $\frac{\partial^2 u}{\partial x_\ell^2} = \frac{1+\rho}{u} UMg_\ell \frac{UMg_\ell x_\ell - u}{x_\ell} < 0$	A utilidade marginal de cada bem é decrescente
Fortemente aditiva: $\frac{\partial^2 u}{\partial x_i x_j} = -a_i a_j \rho(1+\rho)(x_i x_j)^{-(1+\rho)} u^\rho$	A utilidade marginal do bem i não é independente da utilidade marginal do bem j
Homotética: $\theta u = \left(\sum_{\ell=1}^{L} a_\ell (\theta x_\ell)^{-\rho}\right)^{-\frac{1}{\rho}}$	A utilidade aumenta em uma proporção escalar se cada bem é multiplicado por um escalar

	Fórmula	Intuição
Stone-Geary		
	$U = \prod_{\ell=1}^{L} (x_\ell - \gamma_\ell)^{a_\ell}$	Introduz um nível de subsistência na função de utilidade Cobb-Douglas
	$u = \ln U = \sum_{\ell=1}^{L} a_\ell \ln(x_\ell - \gamma_\ell)$	Se $\gamma_\ell = 0$, temos a Cobb-Douglas
	$0 < a_\ell < 1$	As preferências são incorporadas usando a
	Monotônica: $\frac{\partial u}{\partial x_\ell} = \frac{a_\ell}{x_\ell - \gamma_\ell} > 0$	A utilidade marginal de cada bem é positiva
	Concavidade: $\frac{\partial^2 u}{\partial x_\ell^2} = -\frac{a_\ell}{(x_\ell - \gamma_\ell)^2} < 0$	A utilidade marginal de cada bem é decrescente
	Fortemente aditiva: $\frac{\partial^2 u}{\partial x_i x_j} = 0$	A utilidade marginal do bem i é independente da utilidade marginal do bem j
	Homotética: $\theta u \neq \sum_{\ell=1}^{L} a_\ell \ln(\theta x_\ell - \gamma_\ell)$	A utilidade não aumenta em uma proporção escalar se cada bem é multiplicado por um escalar

1.4 Maximização da Utilidade e Escolha Ótima

1.4.1 Escolha Ótima do Consumidor

Uma hipótese fundamental sobre o comportamento individual na economia moderna em geral e a teoria do consumidor em particular é que um agente racional sempre escolherá uma cesta preferida do conjunto de alternativas acessíveis. Derivaremos funções de demanda considerando um modelo de comportamento de maximização de utilidade associado a uma descrição de restrições econômicas subjacentes.

Nesta seção, vamos pensar sobre as ferramentas que precisamos para resolver o problema de otimização e, portanto, completar nosso modelo de comportamento do consumidor. Para começar, vamos pensar sobre o problema intuitivamente, e então mostraremos como podemos usar as ferramentas de cálculo para encontrar a solução.

No problema básico da maximização de preferências, o conjunto de alternativas acessíveis é apenas o conjunto de todas as cestas que satisfazem a restrição orçamentária do consumidor que discutimos anteriormente. Ou seja, o problema da maximização de preferência pode ser escrito como:

$$\begin{aligned} & \max u(x) \\ \text{sujeito a} \quad & px \leqq m \\ & x \in X. \end{aligned} \tag{1.30}$$

Haverá uma solução para esse problema se a função de utilidade for contínua e se o conjunto de restrições for fechado e limitado. O conjunto de restrições está certamente fechado. Se $p_\ell > 0$ para $\ell = 1, \ldots, L$ e $m > 0$, não é difícil mostrar que o conjunto de restrições será limitado. Se algum preço for zero, o consumidor pode querer uma quantidade infinita do correspondente bem .

Proposição 1.4.1. *Sob a hipótese de não-saciedade local, uma cesta x^* que maximiza a utilidade deve atender à restrição orçamentária com igualdade.*

Esta proposição nos permite reescrever o problema do consumidor como

$$\begin{aligned} & \max u(x) \\ \text{sujeito a} \quad & px = m. \end{aligned} \tag{1.31}$$

O valor de x que resolve esse problema é a cesta demandada pelo consumidor: ela expressa quanto de cada bem o consumidor deseja em um determinado nível de preços e renda. Em geral, o ótimo não é único. Denote por $x(p, m)$ o conjunto de todos as cestas maximizadoras de utilidade, chamado de correspondência de demanda do consumidor. Quando há uma cesta de demanda única para cada (p, m), $x(p, m)$ se torna uma função e isso é chamado de função de demanda do consumidor. Veremos a partir da seguinte proposição que a convexidade estrita das preferências garantirá a unicidade da cesta ótima.

Proposição 1.4.2 (Unicidade da cesta demandada)**.** *Se as preferências são estritamente convexas, então para cada $p > 0$ existe uma cesta única x que maximiza u no conjunto orçamentário do consumidor, $B(p, m)$.*

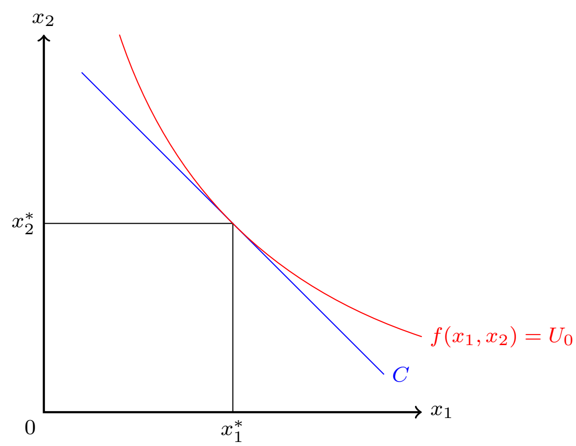
Como a multiplicação de todos os preços e a renda por algum número positivo não altera o orçamento, não pode mudar a resposta para o problema de maximização da utilidade.

Proposição 1.4.3 (Homogeneidade da função demanda)**.** *A função de demanda do consumidor $x(p, m)$ é homogênea de grau 0 em $(p, m) > 0$, ou seja, $x(\alpha p, \alpha m) = x(p, m)$. Isso significa que se os preços e a renda aumentarem em uma mesma proporção, a demanda Marshalliana não mudará.*

1.4.2 Condições de Primeira Ordem: Otimização com Igualdade

Começamos com o problema de maximização condicionada mais simples, o de maximizar uma função $f(x_1, x_2)$ de duas variáveis sujeita a uma única restrição de igualdade: $h(x_1, x_2) = c$. Em Economia, um problema semelhante seria o de maximizar a função utilidade.

Figura 1.29 – No máximo condicionado x^*, a curva de nível de f de valor mais alto é tangente ao conjunto restrição C



A curva de nível de f ser tangente ao conjunto-restrição C no máximo condicionado x^* significa que, em x^*, a inclinação do conjunto de nível de f é igual à inclinação da curva de restrição C, ou seja,

$$\frac{\dfrac{\partial f}{\partial x_1}(x^*)}{\dfrac{\partial f}{\partial x_2}(x^*)} = \frac{\dfrac{\partial h}{\partial x_1}(x^*)}{\dfrac{\partial h}{\partial x_2}(x^*)}, \tag{1.32}$$

ou

$$\frac{\dfrac{\partial f}{\partial x_1}(x^*)}{\dfrac{\partial h}{\partial x_1}(x^*)} = \frac{\dfrac{\partial f}{\partial x_2}(x^*)}{\dfrac{\partial h}{\partial x_2}(x^*)}. \tag{1.33}$$

Para evitarmos denominadores nulos, seja μ o valor comum dos dois quocientes em (1.33):

$$\frac{\dfrac{\partial f}{\partial x_1}(x^*)}{\dfrac{\partial h}{\partial x_1}(x^*)} = \frac{\dfrac{\partial f}{\partial x_2}(x^*)}{\dfrac{\partial h}{\partial x_2}(x^*)} = \mu, \tag{1.34}$$

ou

$$\frac{\partial f}{\partial x_1}(x^*) - \mu\frac{\partial h}{\partial x_1}(x^*) = 0, \tag{1.35}$$

$$\frac{\partial f}{\partial x_2}(x^*) - \mu\frac{\partial h}{\partial x_2}(x^*) = 0. \tag{1.36}$$

Juntando a equação de restrição com as duas equações acima, obtemos um sistema de três equações a três incógnitas:

$$\frac{\partial f}{\partial x_1}(x) - \mu\frac{\partial h}{\partial x_1}(x) = 0, \tag{1.37}$$

$$\frac{\partial f}{\partial x_2}(x) - \mu\frac{\partial h}{\partial x_2}(x) = 0, \tag{1.38}$$

$$h(x_1, x_2) - c = 0. \tag{1.39}$$

Podemos reescrever o sistema acima por meio da função Lagrangeana:

$$L(x_1, x_2, \mu) \equiv f(x_1, x_2) - \mu(h(x_1, x_2) - c). \tag{1.40}$$

Para encontrar os pontos críticos do Lagrangeano L, calculamos $\dfrac{\partial L}{\partial x_1}$, $\dfrac{\partial L}{\partial x_2}$ e $\dfrac{\partial L}{\partial \mu}$ e igualamos cada uma a zero.

Esta nova variável μ que multiplica a restrição é denominada multiplicador de Lagrange. Com este processo, reduzimos um problema de otimização com restrições em duas variáveis a um problema de três variáveis sem restrições. A penalidade para esta redução é a inclusão da variável μ.

Esta redução não teria funcionado se $\dfrac{\partial h}{\partial x_1}$ e $\dfrac{\partial h}{\partial x_2}$ fossem zero no máximo x^*. Por isso, precisamos criar a hipótese de que a parcial $\dfrac{\partial h}{\partial x_1}$, ou a parcial $\dfrac{\partial h}{\partial x_2}$, ou ambas, são não-nulas no máximo condicionado. Como esta é uma imposição no conjunto-restrição, é denominada qualificação de restrição. Se a restrição é linear, esta qualificação é satisfeita automaticamente.

Teorema 1.4.1 (Otimização com Igualdade)**.** *Sejam f e h funções $\mathcal{C}^1$ de duas variáveis. Suponha que $x^* = (x_1^*, x_2^*)$ é uma solução do problema*

$$\begin{aligned} \textit{maximizar} \quad & f(x_1, x_2) \\ \textit{sujeito a} \quad & h(x_1, x_2) = c. \end{aligned}$$

Então, existe um número real μ^* tal que (x_1^*, x_2^*, μ^*) é um ponto crítico da função Lagrangeana

$$L(x_1, x_2, \mu) \equiv f(x_1, x_2) - \mu \left[h(x_1, x_2) - c \right]. \tag{1.41}$$

Em outras palavras, em (x_1^*, x_2^*, μ^*) temos:

$$\frac{\partial L}{\partial x_1} = 0, \quad \frac{\partial L}{\partial x_2} = 0, \quad \frac{\partial L}{\partial \mu} = 0. \tag{1.42}$$

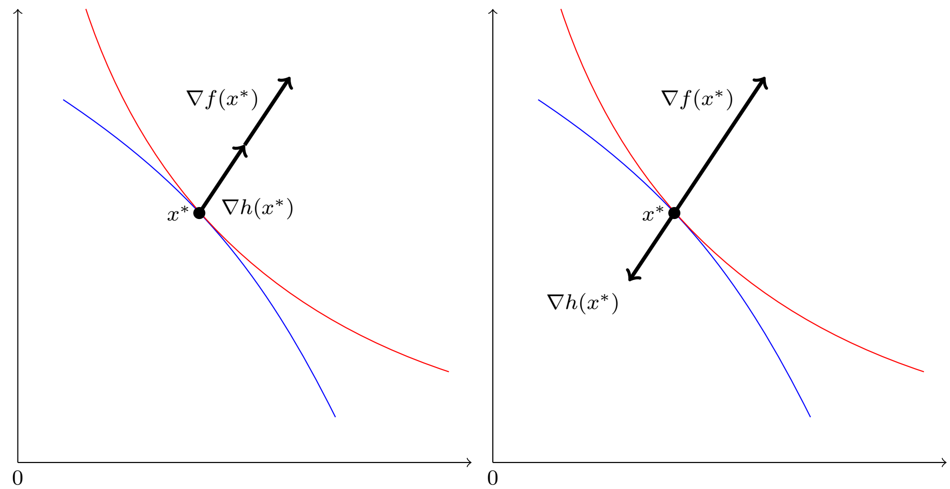
Geometricamente, podemos ver esse teorema por meio dos vetores gradientes:

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} \quad \text{e} \quad \nabla h(x) = \begin{pmatrix} \dfrac{\partial h}{\partial x_1} \\ \dfrac{\partial h}{\partial x_2} \end{pmatrix}, \tag{1.43}$$

considerados como vetores-deslocamento ou flechas no ponto x, são perpendiculares aos conjuntos de nível de f e de h, respectivamente. Como os conjuntos de nível de f e de h têm a mesma inclinação em x^*, os vetores gradiente $\nabla f(x^*)$ e $\nabla h(x^*)$ devem estar alinhados em x^*, ou seja, têm a mesma direção. Os gradientes são múltiplos escalares um do outro. Se escrevermos esse multiplicador como μ^*, obteremos $\nabla f(x^*) = \mu^* \nabla h(x^*)$, ou seja,

$$\begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} = \mu^* \begin{pmatrix} \dfrac{\partial h}{\partial x_1} \\ \dfrac{\partial h}{\partial x_2} \end{pmatrix}. \tag{1.44}$$

Figura 1.30 – $\nabla f(x^*)$ e $\nabla h(x^*)$ estão alinhados no max ou no min condiconado x^*



Exemplo 1.4.1. *Seja o seguinte problema de maximização*

$$\begin{aligned} \textit{maximizar} \quad & f(x_1, x_2) = x_1 x_2 \\ \textit{sujeito a} \quad & h(x_1, x_2) \equiv x_1 + 4x_2 = 16. \end{aligned}$$

Como o gradiente de h é $(1, 4)$, h não tem pontos críticos e a qualificação de restrição está satisfeita. Montando o Lagrangeano

$$L(x_1, x_2, \mu) = x_1 x_2 - \mu(x_1 + 4x_2 - 16), \tag{1.45}$$

e igualando as derivadas parciais a zero

$$\frac{\partial L}{\partial x_1} = x_2 - \mu = 0, \tag{1.46}$$
$$\frac{\partial L}{\partial x_2} = x_1 - 4\mu = 0, \tag{1.47}$$
$$\frac{\partial L}{\partial \mu} = -(x_1 + 4x_2 - 16) = 0. \tag{1.48}$$

Das duas primeiras equações acima decorre que

$$\mu = x_2 = \frac{1}{4} x_1, \tag{1.49}$$

e, portanto,

$$x_1 = 4x_2. \tag{1.50}$$

Substituindo (1.50) *na terceira equação de* (1.46)*, obtemos* $x_2 = 2$*. Disso concluímos que* $x_1 = 8$ *e* $\mu = 2$*.*

Em seguida consideramos o problema de maximizar uma função $f(x_1, \ldots, x_n)$ de n variáveis condicionada por m restrições de igualdade. Em outras palavras,

$$\begin{aligned} \text{maximizar ou minimizar} \quad & f(x_1, \ldots, x_n) \\ \text{sujeito a} \quad & C_h = \{x = (x_1, \ldots, x_n) | h_1(x) = a_1, \ldots, h_m(x) = a_m\}. \end{aligned}$$

Precisamos generalizar para m funções a qualificação de restrição que utilizamos para uma função de duas variáveis:

$$\left(\frac{\partial h}{\partial x_1}(x^*), \frac{\partial h}{\partial x_2}(x^*) \right) \neq (0, 0), \tag{1.51}$$

como

$$\left(\frac{\partial h}{\partial x_1}(x^*), \frac{\partial h}{\partial x_2}(x^*), \ldots, \frac{\partial h}{\partial x_n}(x^*) \right) \neq (0, \ldots, 0). \tag{1.52}$$

Se estivermos tratando com m funções de restrição, $m > 1$, a generalização natural de (1.51) e (1.52) envolve a derivada jacobiana

$$Dh(x^*) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(x^*) & \ldots & \frac{\partial h_1}{\partial x_n}(x^*) \\ \frac{\partial h_2}{\partial x_1}(x^*) & \ldots & \frac{\partial h_2}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(x^*) & \ldots & \frac{\partial h_m}{\partial x_n}(x^*) \end{pmatrix}, \tag{1.53}$$

das funções de restrição. Em geral, um ponto x^* é denominado ponto crítico de h se o posto da matriz $Dh(x^*)$ é menor que m. Portanto, a generalização natural da qualificação de restrição é que o posto $Dh(x^*)$ seja m. Dizemos que $(h_1, \ldots, h_m)$ satisfaz a qualificação de restrição não-degenerada em x^* se o posto da matriz $Dh(x^*)$ é m. A QRND é uma condição de regularidade e implica que o conjunto-restrição tem um plano tangente $(n - m)$-dimensional bem-definido em todos os pontos.

Teorema 1.4.2. *Sejam* $f, h_1, \ldots, h_m$ *funções* $\mathcal{C}^1$ *de* n *variáveis. Considere o problema de maximi-*

zar (minimizar) $f(x)$ no conjunto-restrição

$$C_h \equiv \{x = (x_1, \ldots, x_n) \colon h_1(x) = a_1, \ldots, h_m(x) = a_m\}. \tag{1.54}$$

Suponha que $x^* \in C_h$ e que x^* é uma max ou min (local) de f em C_h. Suponha também que x^* satisfaz a QRND. Então existem $\mu_1^*, \ldots, \mu_m^*$ tais que $(x_1^*, \ldots, x_n^*, \mu_1^*, \ldots, \mu_m^*) \equiv (x^*, \mu^*)$ é um ponto crítico do Lagrangeano

$$L(x, \mu) \equiv f(x) - \mu_1 [h_1(x) - a_1] - \mu_2 [h_2(x) - a_2] - \ldots - \mu_m [h_m(x) - a_m]. \tag{1.55}$$

Em outras palavras,

$$\frac{\partial L}{\partial x_1}(x^*, \mu^*) = 0, \ldots, \frac{\partial L}{\partial x_n}(x^*, \mu^*) = 0, \tag{1.56}$$
$$\frac{\partial L}{\partial \mu_1}(x^*, \mu^*) = 0, \ldots, \frac{\partial L}{\partial \mu_m}(x^*, \mu^*) = 0. \tag{1.57}$$

Exemplo 1.4.2. *Maximize $f(x, y, z) = yz + xz$ sujeito a $y^2 + z^2 = 1$ e $xz = 3$.*
Monte o Lagrangeano,

$$L = yz + xz - \mu_1(y^2 + z^2 - 1) - \mu_2(xz - 3). \tag{1.58}$$

As CPO's são:

$$\frac{\partial L}{\partial x} = z - \mu_2 z = 0, \tag{1.59}$$
$$\frac{\partial L}{\partial y} = z - 2y\mu_1 = 0, \tag{1.60}$$
$$\frac{\partial L}{\partial z} = y + x - 2z\mu_1 - x\mu_2 = 0, \tag{1.61}$$
$$\frac{\partial L}{\partial \mu_1} = 1 - y^2 - z^2 = 0, \tag{1.62}$$
$$\frac{\partial L}{\partial \mu_2} = 3 - xz = 0. \tag{1.63}$$

De (1.59), $\mu_2 = 1$. Isolando μ_1 em (1.60) e (1.61), temos que

$$\mu_1 = \frac{z}{2y}, \tag{1.64}$$
$$\mu_1 = \frac{y}{2z}. \tag{1.65}$$

Igualando (1.64) e (1.65):

$$\frac{z}{2y} = \frac{y}{2z} \;\Rightarrow\; y = z. \tag{1.66}$$

Substituindo (1.66) em (1.62), encontramos que

$$\begin{aligned} 1 - y^2 - z^2 &= 0 \\ 1 - y^2 - y^2 &= 0 \\ 1 &= 2y^2 \\ y &= \pm\frac{\sqrt{2}}{2}. \end{aligned} \tag{1.67}$$

Assim, temos os seguintes pontos: $\left(-3\sqrt{2}, -\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, \frac{1}{2}, 1\right)$ *e* $\left(3\sqrt{2}, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \frac{1}{2}, 1\right)$.

1.4.3 Condições de Segunda Ordem

Agora que sabemos como determinar os pontos críticos em um problema de otimização com restrições, precisamos determinar sob quais circunstâncias esses pontos serão de máximo e de mínimo local da função f. Vamos ver dois teoremas que dão condições suficientes de segunda ordem para garantir que um ponto seja a solução de um problema de otimização com restrição.

Teorema 1.4.3. *Sejam f, $g_1, \ldots, g_m$ e $h_1, \ldots, h_k$ funções $\mathcal{C}^2$ no $\mathbb{R}^n$. Considere o problema de maximizar f sob o conjunto-restrição*

$$C_{g,h} \equiv \{x : g_1(x) \leq b_1, \ldots, g_m(x) \leq b_m, h_1(x) = c_1, \ldots, h_k(x) = c_k\}. \tag{1.68}$$

Monte o Lagrangeano

$$\begin{aligned} L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_k) = f(x) - \lambda_1(g_1(x) - b_1) - \ldots - \lambda_m(g_m(x) - b_m) + \\ - \mu_1(h_1(x) - c_1) - \ldots - \mu_k(h_k(x) - c_k). \end{aligned} \tag{1.69}$$

Suponha que existam multiplicadores $\lambda_1^* \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_k^*$, isto é,

$$\frac{\partial L}{\partial x_1} = 0, \ldots, \frac{\partial L}{\partial x_n} = 0 \quad em \quad (x^*, \lambda^*, \mu^*), \tag{1.70}$$

$$\lambda_1^* \geq 0, \ldots, \lambda_m^* \geq 0, \tag{1.71}$$

$$\lambda_1^*(g_1(x^*) - b_1) = 0, \ldots, \lambda_m^*(g_m(x^*) - b_m) = 0, \tag{1.72}$$

$$h_1(x^*) = c_1, \ldots, h_k(x^*) = c_k. \tag{1.73}$$

Suponha que $g_1, \ldots, g_e$ são ativas em x^* e que $g_{e+1}, \ldots, g_m$ são inativas. Escreva $(g_1, \ldots, g_e) =$

g_E. Suponha que o hessiano de L com respeito a x em (x^*, λ^*, μ^*) é negativo definido sob o conjunto-restrição

$$\{v : Dg_E(x^*)v = 0 \quad e \quad Dh(x^*)v = 0\}, \tag{1.74}$$

isto é,

$$v \neq 0, Dg_E(x^*)v = 0, Dh(x^*)v = 0 \;\Rightarrow\; v^\top \left(D_x^2 L(x^*, \lambda^*, \mu^*)v < 0\right). \tag{1.75}$$

Então, x^* é um máximo restrio local restrito de f sob o conjunto-restrição $C_{g,h}$.

O Hessiano orlado nesse caso é dado por

$$H = \left(\begin{array}{cccccccccc}
0 & \dots & 0 & 0 & \dots & 0 & | & \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & | & \frac{\partial g_e}{\partial x_1} & \dots & \frac{\partial g_e}{\partial x_n} \\
0 & \dots & 0 & 0 & \dots & 0 & | & \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_1}{\partial x_n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & | & \frac{\partial h_k}{\partial x_1} & \dots & \frac{\partial h_k}{\partial x_n} \\
\hline
\frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_e}{\partial x_1} & \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_k}{\partial x_1} & | & \frac{\partial^2 L}{\partial x_1^2} & \dots & \frac{\partial^2 L}{\partial x_n x_1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\
\frac{\partial g_1}{\partial x_n} & \dots & \frac{\partial g_e}{\partial x_n} & \frac{\partial h_1}{\partial x_n} & \dots & \frac{\partial h_k}{\partial x_n} & | & \frac{\partial^2 L}{\partial x_1 x_n} & \dots & \frac{\partial^2 L}{\partial x_n^2}
\end{array}\right). \tag{1.76}$$

Se os últimos $n - (e + k)$ menores principais líderes alternam de sinal, sendo o sinal do determinante da maior matriz igual a $(-1)^n$, então o hessiano é negativo definido.

Teorema 1.4.4. *Sejam f, $g_1, \dots, g_m$ e $h_1, \dots, h_k$ funções $\mathcal{C}^2$ no $\mathbb{R}^n$. Considere o problema de minimizar f sob o conjunto-restrição*

$$C_{g,h} \equiv \{x : g_1(x) \geq b_1, \dots, g_m(x) \geq b_m, h_1(x) = c_1, \dots, h_k(x) = c_k\}. \tag{1.77}$$

Monte o Lagrangeano

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_k) = f(x) - \lambda_1(g_1(x) - b_1) - \dots - \lambda_m(g_m(x) - b_m) +$$

$$- \mu_1(h_1(x) - c_1) - \ldots - \mu_k(h_k(x) - c_k). \tag{1.78}$$

Suponha que existam multiplicadores $\lambda_1^* \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_k^*$, isto é,

$$\frac{\partial L}{\partial x_1} = 0, \ldots, \frac{\partial L}{\partial x_n} = 0 \quad em \quad (x^*, \lambda^*, \mu^*), \tag{1.79}$$

$$\lambda_1^* \geq 0, \ldots, \lambda_m^* \geq 0, \tag{1.80}$$

$$\lambda_1^*(g_1(x^*) - b_1) = 0, \ldots, \lambda_m^*(g_m(x^*) - b_m) = 0, \tag{1.81}$$

$$h_1(x^*) = c_1, \ldots, h_k(x^*) = c_k. \tag{1.82}$$

Suponha que $g_1, \ldots, g_e$ são ativas em x^* e que $g_{e+1}, \ldots, g_m$ são inativas. Escreva $(g_1, \ldots, g_e) = g_E$. Suponha que o hessiano de L com respeito a x em (x^*, λ^*, μ^*) é positivo definido sob o conjunto-restrição

$$\{v : Dg_E(x^*)v = 0 \quad e \quad Dh(x^*)v = 0\}, \tag{1.83}$$

isto é,

$$v \neq 0, Dg_E(x^*)v = 0, Dh(x^*)v = 0 \ \Rightarrow \ v^\top \left(D_x^2 L(x^*, \lambda^*, \mu^*)v > 0\right). \tag{1.84}$$

Então, x^* é um mínimo estrio local restrito de f sob o conjunto-restrição $C_{g,h}$.

O Hessiano orlado nesse caso é dado por

$$H = \left(\begin{array}{cccccccccc}
0 & \ldots & 0 & 0 & \ldots & 0 & | & \frac{\partial g_1}{\partial x_1} & \ldots & \frac{\partial g_1}{\partial x_n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 0 & | & \frac{\partial g_e}{\partial x_1} & \ldots & \frac{\partial g_e}{\partial x_n} \\
0 & \ldots & 0 & 0 & \ldots & 0 & | & \frac{\partial h_1}{\partial x_1} & \ldots & \frac{\partial h_1}{\partial x_n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & \ldots & 0 & | & \frac{\partial h_k}{\partial x_1} & \ldots & \frac{\partial h_k}{\partial x_n} \\
\hline
\frac{\partial g_1}{\partial x_1} & \ldots & \frac{\partial g_e}{\partial x_1} & \frac{\partial h_1}{\partial x_1} & \ldots & \frac{\partial h_k}{\partial x_1} & | & \frac{\partial^2 L}{\partial x_1^2} & \ldots & \frac{\partial^2 L}{\partial x_n x_1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots \\
\frac{\partial g_1}{\partial x_n} & \ldots & \frac{\partial g_e}{\partial x_n} & \frac{\partial h_1}{\partial x_n} & \ldots & \frac{\partial h_k}{\partial x_n} & | & \frac{\partial^2 L}{\partial x_1 x_n} & \ldots & \frac{\partial^2 L}{\partial x_n^2}
\end{array}\right). \tag{1.85}$$

Se os últimos $n-(e+k)$ menores principais líderes têm todos o mesmo sinal de $(-1)^{e+k}$, então o hessiano é positivo definido.

Exemplo 1.4.3. *Maximize a função $f(x,y)=x^2y$ no conjunto-restrição $h(x,y)=2x^2+y^2=3$ e prove que a solução se trata de um máximo.*

Montando o Lagrangeano, temos:

$$L=x^2y-\mu(2x^2+y^2-3). \tag{1.86}$$

As CPO's são:

$$\frac{\partial L}{\partial x}=2xy-4x\mu=0, \tag{1.87}$$

$$\frac{\partial L}{\partial y}=x^2-2y\mu=0, \tag{1.88}$$

$$2x^2+y^2-3=0. \tag{1.89}$$

Igualando as duas primeiras condições de primeira ordem para μ, obtemos $x=y$. Portanto, $x=\pm 1$ e $y=\pm 1$

Assim, temos quatro candidatos a máximo:

$$(1,1), \tag{1.90}$$

$$(-1,-1) \tag{1.91}$$

$$(-1,1), \tag{1.92}$$

$$(1,-1). \tag{1.93}$$

Agora vamos utilizar as condições de segunda ordem. Montando o hessiano, temos:

$$H=\begin{pmatrix}0 & 4x & 2y\\ 4x & 2y & 2x\\ 2y & 2x & 0\end{pmatrix}=8y(4x^2-y^2). \tag{1.94}$$

Como $n=2$ e $k=1$ basta conferir o sinal de $n-k=1$ determinante, o determinante de H. Se $\det H$ tiver o mesmo sinal de $(-1)^n=1$, o ponto é de máximo local; se $\det H$ tiver o mesmo sinal de $(-1)^k=-1$, o ponto é de mínimo local.

- *Se $(x^*,y^*)=(1,1)$, então $H=24$. Máximo local.*
- *Se $(x^*,y^*)=(-1,-1)$, então $H=-24$. Mínimo local.*
- *Se $(x^*,y^*)=(-1,1)$, então $H=24$. Máximo local.*

- *Se* $(x^*, y^*) = (1, -1)$, *então* $H = -24$. *Mínimo local.*

1.4.4 Condições de Primeira Ordem: Problema do Consumidor

Podemos caracterizar o comportamento de otimização por cálculo, desde que a função de utilidade seja diferenciável. Vamos analisar este problema de maximização restrita usando o método dos multiplicadores de Lagrange. O Lagrangeano para o problema de maximização da utilidade pode ser escrito como

$$L = u(x) - \lambda(px - m), \tag{1.95}$$

em que λ é o multiplicador de Lagrange. Suponha que as preferências seja localmente não-saciadas. Diferenciando o Lagrangeano em relação a x_i, obtemos as condições de primeira ordem para a solução interior:

$$\frac{\partial u(x)}{\partial x_i} - \lambda p_i = 0,\ \forall i = 1, \ldots, L \tag{1.96}$$

$$px = m. \tag{1.97}$$

Usando notação vetorial, temos:

$$Du(x) = \lambda p, \tag{1.98}$$

em que $Du(x) = \left(\frac{\partial u(x)}{\partial x_1}, \ldots, \frac{\partial u(x)}{\partial x_L}\right)$ é o vetor gradiente de $u(x)$ com relação a cada um de seus argumentos.

Para interpretar estas condições, podemos dividir a i-éisma condição de primeira ordem pela j-ésima condição de primeira ordem para eliminar o multiplicador Lagrange. Isso nos dá

$$\frac{\dfrac{\partial u(x^*)}{\partial x_i}}{\dfrac{\partial u(x^*)}{\partial x_j}} = \frac{p_i}{p_j}, \quad i, j = 1, \ldots, L. \tag{1.99}$$

A fração à esquerda é a taxa marginal de substituição entre os bens i e j, e a fração à direita é a taxa econômica de substituição entre os bens i e j. Maximização implica que essas duas taxas de substituição devem ser iguais. Suponha que eles não fossem; por exemplo, suponha

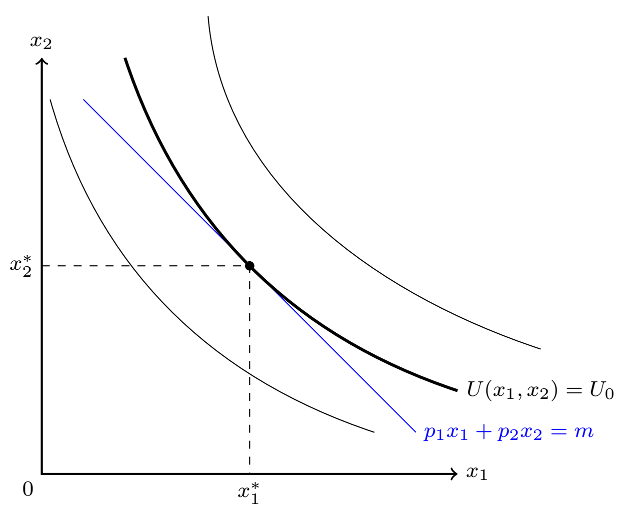
$$\frac{\dfrac{\partial u(x^*)}{\partial x_i}}{\dfrac{\partial u(x^*)}{\partial x_j}} = \frac{1}{1} \neq \frac{2}{1} = \frac{p_i}{p_j}. \tag{1.100}$$

Então, se o consumidor desiste de uma unidade do bem i e compra uma unidade do bem j, ele permanecerá na mesma curva e terá um dólar a mais para gastar. Assim, a utilidade total pode ser aumentada, contradizendo a maximização.

A Figura 1.31 ilustra o argumento geometricamente. A linha orçamentária do consumidor é dada por $\{x\colon p_1x_1 + p_2x_2 = m\}$. Isso pode ser escrito como o gráfico de uma função implícita: $x_2 = \frac{m}{p_2} - \frac{p_1}{p_2}x_1$. Assim, a linha orçamentária tem inclinação $-\frac{p_1}{p_2}$ e interceptação vertical em $\frac{m}{p_2}$. O consumidor quer encontrar o ponto nesta linha de orçamento que atinge a maior utilidade. Isto deve claramente satisfazer a condição de tangência de que a inclinação da curva de indiferença seja igual à inclinação da linha orçamentária, de modo que a taxa marginal de substituição entre x_1 e x_2 seja igual à taxa econômica de substituição entre x_1 e x_2.

Essa condição de tangência também faz sentido: no ponto de tangência, sabemos que a inclinação da linha orçamentária é igual à inclinação da curva de indiferença. O primeiro desses declives é a taxa na qual o consumidor pode negociar o bem i pelo bem j no mercado, enquanto o segundo mede a taxa a qual o consumidor negocia o bem i pelo bem j mantendo o nível de utilidade constante. Se houvesse uma diferença entre estas duas razões, então o consumidor poderia passar para uma curva de indiferença diferente, porque os preços de mercado iriam sobre ou subvalorizar o bem i em relação ao bem j em comparação com as preferências do consumidor. O consumidor poderia, portanto, trocar o bem i pelo bem j (ou vice-versa) e assim passar para uma curva de indiferença mais alta. Isso ocorre apenas quando as duas taxas são iguais, o que significa que o mercado “valoriza” a troca do bem i pelo bem j na mesma taxa que o consumidor o faz, e assim o consumidor não pode “passar” para uma curva de indiferença mais alta.

Figura 1.31 – Maximização da utilidade



As condições de cálculo derivadas acima só fazem sentido quando a variável de escolha pode ser variada em uma vizinhança aberta da escolha ótima e a restrição orçamentária é vinculativa. Em muitos problemas econômicos, as variáveis são naturalmente não-negativas. Se algumas variáveis tiverem um valor de zero na escolha ótima, as condições de cálculo descritas acima podem ser inadequadas. As modificações necessárias das condições para lidar com soluções de limite não são difíceis de indicar. As condições relevantes de primeira ordem são dadas por meio das chamadas condições de Kuhn-Tucker:

$$\frac{\partial u(x)}{\partial x_i} - \lambda p_i \leqq 0, \text{ com igualdade se } x_i > 0 \ \forall i = 1, \ldots, L \tag{1.101}$$

$$px \leqq m, \text{ com igualdade se } \lambda > 0 \tag{1.102}$$

Assim, a utilidade marginal do aumento de x_i deve ser menor ou igual a λp_i, caso contrário o consumidor aumentaria x_i. Se $x_i = 0$, a utilidade marginal do aumento x_i pode ser menor que λp_i, ou seja, o consumidor gostaria de diminuir x_i. Mas como x_i já é zero, isso é impossível. Finalmente, se $x_i > 0$ para que a restrição de não-negatividade não seja vinculativa, teremos as condições usuais para uma solução interna.

Sabemos que a solução para o problema de otimização será uma solução de canto (ou seja, $x_1 = 0$ ou $x_2 = 0$), ou será um ponto de tangência. Assim, uma maneira segura de encontrar a cesta ideal (ótima) é listar todos esses pontos (ou seja, todos os pontos de tangência e todas as soluções de canto) e determinar qual é a melhor cesta. A menos que você tenha uma razão para, a priori, descartar algumas soluções possíveis para o problema de otimização, essa é a abordagem que você deve seguir para resolver o problema.

1.4.5 Condições de Segunda Ordem: Problema do Consumidor

As condições de primeira ordem acima são apenas condições necessárias para um ótimo local. No entanto, para o problema específico em questão, essas condições necessárias de primeira ordem são de fato suficientes para um ótimo global quando a função de utilidade é quase-côncava. Temos então a seguinte proposição.

Proposição 1.4.4. *Suponha que $u(x)$ seja diferenciável e quase-côncava em $\mathbb{R}^L_+$ e $(p, m) > 0$. Se (x, λ) satisfizer as condições de primeira ordem, então x resolve o problema de maximização da utilidade do consumidor aos preços p e renda m.*

Esta é uma das razões pelas quais os economistas amam tanto as preferências convexas! Embora note que ainda há um pequeno problema - mesmo com uma função convexa, pode não haver pontos de tangência, caso em que a solução para o problema será uma solução de canto.

1.4.6 Exemplos

Exemplo 1.4.4. *Suponha que as preferências de um consumidor sejam representadas pela função de utilidade Cobb-Douglas: $u(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$, que é estritamente quase-côncava no $\mathbb{R}^2_{++}$. Como* *qualquer transformação monotônica dessa função representa as mesmas preferências, também podemos escrever* $u(x_1, x_2) = \alpha \ln x_1 + (1-\alpha)\ln x_2$.

O problema de maximização pode ser escrito como

$$\begin{aligned} &\max \alpha \ln x_1 + (1-\alpha)\ln x_2 \\ \textit{sujeito a} \quad & p_1 x_1 + p_2 x_2 = m \end{aligned} \tag{1.103}$$

As CPO's são:

$$\frac{\alpha}{x_1} - \lambda p_1 = 0 \tag{1.104}$$

$$\frac{1-\alpha}{x_2} - \lambda p_2 = 0 \tag{1.105}$$

$$p_1 x_1 + p_2 x_2 = m \tag{1.106}$$

Igualando as duas primeiras condições de primeira ordem, temos:

$$\frac{\alpha}{p_1 x_1} = \frac{1-\alpha}{p_2 x_2} \tag{1.107}$$

Resolvendo, obtemos a demanda Marshalliana pelo bem 1:

$$x_1(p_1, p_2, m) = \frac{\alpha m}{p_1} \tag{1.108}$$

Substitua esse resultado dentro da restrição orçamentária para obter:

$$x_2(p_1, p_2, m) = \frac{(1-\alpha)m}{p_2} \tag{1.109}$$

Para ver que se trata de um máximo, podemos utilizar duas estratégias:

- *Verificar se a função de utilidade é quase-côncava. Para tanto, seja a seguinte matriz hessiana:*

$$H = \begin{vmatrix} 0 & \dfrac{\alpha}{x_1} & \dfrac{(1-\alpha)}{x_2} \\ \dfrac{\alpha}{x_1} & -\dfrac{\alpha}{x_1^2} & 0 \\ \dfrac{(1-\alpha)}{x_2} & 0 & -\dfrac{(1-\alpha)}{x_2^2} \end{vmatrix} \tag{1.110}$$

Como $|H_2| = -\left(\dfrac{\alpha}{x_1}\right)^2 < 0$ *e* $|H_3| = \dfrac{\alpha(1-\alpha)}{(x_1 x_2)^2} > 0$, *a função de utilidade é quase-côncava e,* *portanto, a candidata a solução é ponto de máximo de* $u(x_1, x_2)$.

- *Verificar se o hessiano orlado é negativo definido. Temos* $n = 2$ *e* $k = 1$. *Basta verificarmos se o determinante de* $|H|$ *temo o mesmo sinal que* $(-1)^n$:

$$H = \begin{vmatrix} 0 & p_1 & p_2 \\ p_1 & -\dfrac{\alpha}{x_1^2} & 0 \\ p_2 & 0 & -\dfrac{(1-\alpha)}{x_2^2} \end{vmatrix} \tag{1.111}$$

Como $|H| = \alpha \left(\dfrac{p_2}{x_1}\right)^2 + (1-\alpha)\left(\dfrac{p_1}{x_2}\right)^2 > 0$, *então o ponto é de máximo local sob o conjunto-restrição acima.*

Exemplo 1.4.5. *Suponha que as preferências sejas representadas pela função de utilidade Leotief:* $u(x_1, x_2) = \min\{a_1x_1, a_2x_2\}$. *Como a função de utilidade Leontief não é diferenciável, o máximo deve ser encontrado por um argumento direto. Suponha que* $p > 0$. *A solução ótima deve estar no ponto de dobra da curva de indiferença. Isto é*

$$a_1x_1 = a_2x_2 \tag{1.112}$$

Substituindo $x_1 = \dfrac{a_2}{a_1}x_2$ *dentro da restrição orçamentária, obtemos:*

$$p_1\frac{a_2}{a_1}x_1 + p_2x_2 = m \tag{1.113}$$

Portanto, as funções de demanda Marshalliana são:

$$x_1(p_1, p_2, m) = \frac{a_2m}{a_2p_1 + a_1p_2} \tag{1.114}$$

$$x_2(p_1, p_2, m) = \frac{a_1m}{a_2p_1 + a_1p_2} \tag{1.115}$$

Exemplo 1.4.6. *Agora, suponha que as preferências sejas representadas pela função de utilidade linear*

$$u(x_1, x_2) = a_1x_1 + a_2x_2 \tag{1.116}$$

Como a taxa marginal de substituição é $\dfrac{a_1}{a_2}$ *e a taxa econômica de substituição é* $\dfrac{p_1}{p_2}$ *são ambas constantes, elas não podem ser em geral iguais. Portanto, a condição de primeira ordem não pode ser mantida com igualdade quando* $\dfrac{a_1}{a_2} \neq \dfrac{p_1}{p_2}$. *Nesse caso, a resposta ao problema de* *maximização da utilidade normalmente envolve uma solução de limite: somente um dos dois bens será consumido. Vale a pena apresentar uma solução mais formal, pois serve como um bom exemplo do teorema de Kuhn-Tucker em ação. O teorema de Kuhn-Tucker é a ferramenta apropriada para usar aqui, já que quase nunca teremos uma solução interior. A função Lagrange é*

$$L(x_1, x_2, \lambda) = a_1x_1 + a_2x_2 + \lambda(m - p_1x_1 - p_2x_2), \tag{1.117}$$

e, portanto,

$$\frac{\partial L}{\partial x_1} = a_1 - \lambda p_1 \tag{1.118}$$
$$\frac{\partial L}{\partial x_2} = a_2 - \lambda p_2 \tag{1.119}$$
$$\frac{\partial L}{\partial \lambda} = m - p_1x_1 - p_2x_2 \tag{1.120}$$

Há quatro casos a serem considerados:

1. $x_1 > 0$ *e* $x_2 > 0$*: então temos* $\frac{\partial L}{\partial x_1} = 0$ *e* $\frac{\partial L}{\partial x_2} = 0$*. Portanto,* $\frac{a_1}{a_2} = \frac{p_1}{p_2}$*. Uma vez que* $\lambda = \frac{a_1}{p_1} > 0$*, temos que* $p_1x_1 + p_2x_2 = m$ *e qualquer* x_1 *e* x_2 *que satisfizer a restrição será um ótimo.*

2. $x_1 > 0$ *e* $x_2 = 0$*: então temos* $\frac{\partial L}{\partial x_1} = 0$ *e* $\frac{\partial L}{\partial x_2} \leqq 0$*. Portanto,* $\frac{a_1}{a_2} \geqq \frac{p_1}{p_2}$*. Uma vez que* $\lambda = \frac{a_1}{p_1} > 0$*, temos que* $p_1x_1 + p_2x_2 = m$ *e portanto* $x_1 = \frac{m}{p_1}$*.*

3. $x_1 = 0$ *e* $x_2 > 0$*: então temos* $\frac{\partial L}{\partial x_1} \leqq 0$ *e* $\frac{\partial L}{\partial x_2} = 0$*. Portanto,* $\frac{a_1}{a_2} \leqq \frac{p_1}{p_2}$*. Uma vez que* $\lambda = \frac{a_2}{p_2} > 0$*, temos que* $p_1x_1 + p_2x_2 = m$ *e portanto* $x_2 = \frac{m}{p_2}$*.*

4. $x_1 = 0$ *e* $x_2 = 0$*: então temos* $\frac{\partial L}{\partial x_1} \leqq 0$ *e* $\frac{\partial L}{\partial x_2} \leqq 0$*. Temos que* $p_1x_1 + p_2x_2 = m$ *e portanto* $m = 0$ *uma vez que* $x_1 = 0$ *e* $x_2 = 0$*.*

Em resumo, temos as seguintes demandas mashallianas:

$$(x_1(p_1, p_2, m), x_2(p_1, p_2, m)) = \begin{cases} \left(\frac{m}{p_1}, 0\right), & se\ \frac{a_1}{a_2} > \frac{p_1}{p_2} \\ \left(0, \frac{m}{p_2}\right), & se\ \frac{a_1}{a_2} < \frac{p_1}{p_2} \\ \left(x_1, \frac{m}{p_1} - \frac{p_2}{p_1}x_2\right), & se\ \frac{a_1}{a_2} = \frac{p_1}{p_2} \end{cases} \tag{1.121}$$

em que $x \in \left[0, \dfrac{m}{p_1}\right]$.

Como a função de utilidade é linear, ela é quase-côncava.

Exemplo 1.4.7. *Suponha que as preferências de um consumidor sejam representadas pela função de utilidade CES:* $u(x_1, x_2) = \gamma \left[a_1 x_1^{-\rho} + a_2 x_2^{-\rho}\right]^{-\frac{1}{\rho}}$*, que é estritamente quase-côncava no* $\mathbb{R}^2_{++}$.

O problema de maximização pode ser escrito como

$$\begin{aligned} &\max \gamma \left[a_1 x_1^{-\rho} + a_2 x_2^{-\rho}\right]^{-\frac{1}{\rho}} \\ &\textit{sujeito a} \quad p_1 x_1 + p_2 x_2 = m \end{aligned} \tag{1.122}$$

As CPO's são:

$$\gamma a_1 x_1^{-(1+\rho)} \left[a_1 x_1^{-\rho} + a_2 x_2^{-\rho}\right]^{-\frac{(1+\rho)}{\rho}} - \lambda p_1 = 0 \tag{1.123}$$

$$\gamma a_2 x_2^{-(1+\rho)} \left[a_1 x_1^{-\rho} + a_2 x_2^{-\rho}\right]^{-\frac{(1+\rho)}{\rho}} - \lambda p_2 = 0 \tag{1.124}$$

$$p_1 x_1 + p_2 x_2 = m \tag{1.125}$$

Igualando as duas primeiras condições de primeira ordem, temos:

$$\frac{a_1 x_1^{-(1+\rho)}}{p_1} = \frac{a_2 x_2^{-(1+\rho)}}{p_2} \tag{1.126}$$

Resolvendo, obtemos as seguintes demandas Marshallianas:

$$x_1(p_1, p_2, m) = \frac{(a_2 p_1)^{-\left(\frac{1}{(1+\rho)}\right)}}{a_2^{-\left(\frac{1}{(1+\rho)}\right)} p_1^{\left(\frac{\rho}{(1+\rho)}\right)} + a_1^{-\left(\frac{1}{(1+\rho)}\right)} p_2^{\left(\frac{\rho}{(1+\rho)}\right)}} m \tag{1.127}$$

$$x_2(p_1, p_2, m) = \frac{(a_1 p_2)^{-\left(\frac{1}{(1+\rho)}\right)}}{a_2^{-\left(\frac{1}{(1+\rho)}\right)} p_1^{\left(\frac{\rho}{(1+\rho)}\right)} + a_1^{-\left(\frac{1}{(1+\rho)}\right)} p_2^{\left(\frac{\rho}{(1+\rho)}\right)}} m \tag{1.128}$$

Exemplo 1.4.8. *Suponha que as preferências de um consumidor sejam representadas pela função de utilidade quase-linear:* $u(x_1, x_2) = x_1 + \ln x_2$.

O problema de maximização pode ser escrito como

$$\max x_1 + \ln x_2$$
$$\textit{sujeito a} \quad p_1x_1 + p_2x_2 = m \tag{1.129}$$

As CPO's são:

$$1 - \lambda p_1 = 0 \tag{1.130}$$
$$\frac{1}{x_2} - \lambda p_2 = 0 \tag{1.131}$$
$$p_1x_1 + p_2x_2 = m \tag{1.132}$$

Igualando as duas primeiras condições de primeira ordem, temos:

$$x_2 = \frac{p_1}{p_2} \tag{1.133}$$

Resolvendo, obtemos as seguintes demandas Marshallianas:

$$x_1(p_1, p_2, m) = \frac{m - p_1}{p_1} \tag{1.134}$$
$$x_2(p_1, p_2, m) = \frac{p_1}{p_2} \tag{1.135}$$

1.4.7 Uma Interpretação do Multiplicador de Lagrange

Lembre-se do problema de otimização por meio da função Lagrangeana:

$$Z = L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda[m - g(x_1, x_2)].$$

Encontre os pontos críticos da função Lagrangeana $L(x_1, x_2, \lambda)$ calculando $\frac{\partial L}{\partial x_1}$, $\frac{\partial L}{\partial x_2}$ e $\frac{\partial L}{\partial \lambda}$ e definindo cada expressão igual a 0 para resolver para o ótimo $(x_1^*, x_2^*, \lambda^*)$:

$$\frac{\partial L}{\partial x_1} = 0, \tag{1.136}$$
$$\frac{\partial L}{\partial x_2} = 0, \tag{1.137}$$
$$\frac{\partial L}{\partial \lambda} = 0. \tag{1.138}$$

Note que como λ apenas multiplica a restrição na definição de L, a equação $\frac{\partial L}{\partial \lambda} = 0$ é equivalente à restrição $m - g(x_1, x_2) = 0$.

Note que introduzindo o multiplicador de Lagrange λ no problema restrito, transformamos um problema restrito de duas variáveis no problema irrestrito de três variáveis de encontrar os pontos críticos de uma função $L(x_1, x_2, \lambda)$.

O multiplicador de Lagrange λ mede a sensibilidade de Z^* (que é o valor de Z no ótimo) a alguma mudança na restrição. Em outras palavras, nos dá uma nova medida de valor dos recursos escassos (o efeito de um aumento em m indicaria como a solução ótima é afetada por um relaxamento da restrição). Se pudermos expressar os valores ótimos de $(x_1^*, x_2^*, \lambda^*)$ como funções implícitas de m, teríamos:

$$x_1^* = x_1^*(m), \tag{1.139}$$
$$x_2^* = x_2^*(m), \tag{1.140}$$
$$\lambda^* = \lambda^*(m). \tag{1.141}$$

Vamos assumir que todas essas expressões têm derivdas contínuas, isto é, são funções de classe $\mathcal{C}^1$. Disso, temos as seguintes identidades:

$$f_{x_1}(x_1^*, x_2^*) - \lambda^* g_{x_1}(x_1^*, x_2^*) = 0, \tag{1.142}$$
$$f_{x_2}(x_1^*, x_2^*) - \lambda^* g_{x_2}(x_1^*, x_2^*) = 0, \tag{1.143}$$
$$m - (x_1^*, x_2^*) = 0. \tag{1.144}$$

Com isso, podemos escrever a função Lagrangeana na solução ótima como:

$$Z^* = L(x_1^*, x_2^*, \lambda) = f(x_1^*, x_2^*) + \lambda^*[m - g(x_1^*, x_2^*)]. \tag{1.145}$$

Dissemos que o valor de λ media a sensibilidade da função Lagrangeana a uma mudança em m quando dos valores ótimos. Vamos ver isso.

$$\begin{aligned}
\frac{dZ^*}{dm} &= f_{x_1}\frac{dx_1^*}{dm} + f_{x_2}\frac{dx_2^*}{dm} + \frac{d\lambda^*}{dm}[m - g(x_1^*, x_2^*)] + \lambda^*\left[1 - \left(g_{x_1}\frac{dx_1^*}{dm} + g_{x_2}\frac{dx_2^*}{dm}\right)\right] \\
&= f_{x_1}\frac{dx_1^*}{dm} + f_{x_2}\frac{dx_2^*}{dm} + \lambda^*\left[1 - \left(g_{x_1}\frac{dx_1^*}{dm} + g_{x_2}\frac{dx_2^*}{dm}\right)\right], \quad (m - (x_1^*, x_2^*) = 0) \\
&= \left(f_{x_1} - \lambda^* g_{x_1}\right)\frac{dx_1^*}{dm} + \left(f_{x_2} - \lambda^* g_{x_2}\right)\frac{dx_2^*}{dm} + \lambda^* \\
&= \lambda^*,
\end{aligned} \tag{1.146}$$

em que f_{x_1} e f_{x_2} medem a contribuição marginal (ou benefício) de cada x a função a ser maximizada, ou a mudança aproximada em f decorrente de uma mudança unitária em x. De forma semelhante, g_{x_1} e g_{x_2} dão o “custo marginal” de se utilizar cada x, ou a mudança aproximada em g decorrente de uma mudança unitária em x. Lembre-se que você pode escrever

$$\lambda^* = \frac{f_{x_1}(x_1^*, x_2^*)}{g_{x_1}(x_1^*, x_2^*)}, \tag{1.147}$$

$$\lambda^* = \frac{f_{x_2}(x_1^*, x_2^*)}{g_{x_2}(x_1^*, x_2^*)}, \tag{1.148}$$

ou seja, λ^* nos dá uma razão benefício-custo de se utilizar x_i:

$$\lambda^* = \frac{\text{benefício marginal de } x_i}{\text{custo marginal de } x_i}. \tag{1.149}$$

Como normalmente m significa restrições impostas (orçamento, custo, limitação de produção), o valor do multiplicador indica o chamado custo de oportunidade (dessa restrição). Se pudéssemos reduzir a restrição (ou seja, aumentar m), então o custo extra é $-\lambda$. Claramente para o tomador de decisão econômica, tal informação sobre os custos de oportunidade é de considerável importância.

Note que esta expressão só é válida quando $x_1 = x_1^*$ e $x_2 = x_2^*$. Se x_1 e x_2 não estivessem em seus valores ótimos, então a derivada total de L em relação a m também incluiria termos parciais cruzados adicionais. Essas derivadas parciais cruzadas são zero em $x_1 = x_1^*$ e $x_2 = x_2^*$. A variável λ também é conhecida como preço-sombra. Observe que esse preço-sombra não é definido exclusivamente, pois corresponde à utilidade marginal da renda em "utils", que é um valor ordinal.

Note que, embora Z^* seja de fato um tipo padrão de extremo com referência às variáveis de escolha, não é assim com referência ao multiplicador de Lagrange. A equação mostra que, diferentemente de (x_1^*, x_2^*), se λ^* é substituído por qualquer outro valor de λ, nenhum efeito será produzido em Z^*, já que $m - g(x_1^*, x_2^*) = 0$. Assim, o papel desempenhado por λ^* na solução ótima difere do papel desempenhado por x_1^* e x_2^*. Embora seja seguro tratar λ como outra variável de escolha na discussão das condições de primeira ordem, devemos tratar λ^* de maneira diferente na discussão das condições de segunda ordem. As condições de segunda ordem podem ser novamente declarados em termos do diferencial total d^2Z.

A restrição $g(x_1, x_2) = m$ implica que $dg = g_{x_1} d_{x_1} + g_{x_2} d_{x_2} = 0$. d_{x_1} e d_{x_2} não são mais arbitrários: podemos tomar d_{x_1} como uma mudança arbitrária, mas então d_{x_2} depende de d_{x_1}, isto é, $d_{x_2} = -\left(\frac{g_{x_1}}{g_{x_2}}\right) d_{x_2}$. Observe que, como g_{x_1} e g_{x_2} dependem de x_1 e x_2, d_{x_2} também depende de x_1 e x_2. Assim,

$$\begin{aligned}
d^2Z = d(dZ) &= \frac{\partial(dZ)}{\partial x_1} d_{x_1} + \frac{\partial(dZ)}{\partial x_2} d_{x_2} \\
&= \left(\frac{\partial(f_{x_1} dx_1 + f_{x_2} dx_2)}{\partial x_1}\right) d_{x_1} + \left(\frac{\partial(f_{x_1} dx_1 + f_{x_2} dx_2)}{\partial x_2}\right) d_{x_2} \\
&= \left[f_{x_1 x_1} dx_1 + \left(f_{x_1 x_2} dx_2 + f_{x_2} \frac{\partial dx_2}{\partial x_1}\right)\right] dx_1 + \left[f_{x_1 x_2} dx_1 + \left(f_{x_2 x_2} dx_2 + f_{x_2} \frac{\partial dx_2}{\partial x_2}\right)\right] dx_2
\end{aligned}$$

$$= f_{x_1x_1}dx_1^2 + 2f_{x_1x_2}dx_1dx_2 + f_{x_2x_2}dx_2^2 + \left(f_{x_2}\frac{\partial dx_2}{\partial x_1}dx_1 + f_{x_2}\frac{\partial dx_2}{\partial x_2}dx_2\right)$$
$$= f_{x_1x_1}dx_1^2 + 2f_{x_1x_2}dx_1dx_2 + f_{x_2x_2}dx_2^2 + f_{x_2}d^2x_2. \tag{1.150}$$

O último termo desqualifica d^2Z como uma forma quadrática. Mas d^2Z pode ser transformado em uma forma quadrática em virtude da restrição $g(x_1, x_2) = m$:

$$dg = 0 \;\Rightarrow\; d(dg) = g_{x_1x_1}dx_1^2 + 2g_{x_1x_2}dx_1dx_2 + g_{x_2x_2}dx_2^2 + g_{x_2}d^2x_2 = 0. \tag{1.151}$$

Resolvendo a última equação para d^2x_2 e substituindo o resultado na expressão de d^2Z, obtemos:

$$d^2Z = \left(f_{x_1x_1} - \frac{f_{x_2}}{g_{x_2}}g_{x_1x_1}\right)dx_1^2 + 2\left(f_{x_1x_2} - \frac{f_{x_2}}{g_{x_2}}g_{x_1x_2}\right)dx_1dx_2 + \left(f_{x_2x_2} - \frac{f_{x_2}}{g_{x_2}}g_{x_2x_2}\right)dx_2^2$$
$$= \left(f_{x_1x_1} - \lambda g_{x_1x_1}\right)dx_1^2 + 2\left(f_{x_1x_2} - \lambda g_{x_1x_2}\right)dx_1dx_2 + \left(f_{x_2x_2} - \lambda g_{x_2x_2}\right)dx_2^2. \tag{1.152}$$

Para um problema extremo restrito, as condições necessárias e suficientes ainda são determinadas pela diferencial total d^2Z para dx_1 e dx_2 satisfazendo $dg = g_{x_1}dx_1 + g_{x_2}dx_2 = 0$.

Teorema 1.4.5 (Condição Suficiente)**.** *Para um máximo temos que ter d^2Z negativo definido sujeito a $dg = 0$. Para um mínimo temos que ter d^2Z positivo definido sujeito a $dg = 0$.*

Teorema 1.4.6 (Condição Necessária)**.** *Para um máximo temos que ter d^2Z negativo definido sujeito a $dg = 0$. Para um mínimo temos que ter d^2Z positivo definido sujeito a $dg = 0$.*

Resumo 1.4.1. *Como vimos, um consumidor tem um conjunto bem definido de desejos, ou "preferências", que podem ser representados por uma função de utilidade numérica. Além disso, supomos que o consumidor escolhe otimamente, no sentido de que eles escolhem a opção com a maior utilidade entre as disponíveis para elas. Isso significa que um consumidor está resolvendo um problema de otimização. Essa é uma classe importante de problemas que surgem várias vezes ao longo da economia (por exemplo, supõe-se também que as empresas resolvam problemas de otimização). Portanto, vale a pena descrever como são esses problemas em termos gerais.*

Um problema de otimização tem três componentes principais:

- *Os objetos de escolha: o que é isso que está sendo escolhido? No caso do nosso consumidor, serão as diferentes cestas de bens que eles podem comprar.*
- *A função objetivo: o que o consumidor está tentando maximizar? No caso do consumidor, esta é a função de utilidade.*
- *Restrições: quais restrições existem nas escolhas que podem ser feitas? No caso do consumidor, este é o conjunto de bens que eles podem pagar.*

Isto pode ser reescrito como:

$$\max_{\textit{cesta de consumo}} \quad \textit{preferências (ou utilidade) sujeito a uma restrição orçamentária}$$

ou

$$\max_{x,y} U(x,y) - \lambda[g(x,y) - m]$$

Uma das habilidades mais importantes para aprender em economia é ser capaz de formular problemas de otimização. Aqui estão alguns exemplos:

Exemplo 1.4.9. *Um estudante, Pedro, está tentando decidir quais cursos que ele vai fazer. Ele quer obter a maior média possível de notas, mas também quer fazer cursos com concentração em Economia. Ele escolhe os cursos oferecidos por uma universidade, a fim de maximizar a média de notas, satisfazendo as exigências da concentração em Economia.*

Exemplo 1.4.10. *O governo brasileiro está inscrito para reduzir as emissões de* CO_2 *em 25%. No entanto, eles querem fazer isso de uma maneira que minimize os danos à economia. O ideal (ótimo) é escolher uma política fiscal (impostos e subsídios) para maximizar a produção econômica sujeita à redução das emissões de gases de efeito estufa em 25%.*

1.5 Utilidade Indireta e Dispêndio

A função de utilidade ordinária, $u(x)$, é definida sobre o conjunto de consumo X e, portanto, como a função de utilidade direta. Dados os preços p e a renda m, o consumidor escolhe uma cesta $x(p,m)$ que maximiza a utilidade. O nível de utilidade alcançado quando $x(p,m)$ é escolhido, portanto, será o nível mais alto permitido pela restrição orçamentária do consumidor para p e m, e pode ser denotado por

$$\begin{aligned} v(p,m) &= \max u(x) \\ \text{sujeito a} \quad px &= m. \end{aligned} \tag{1.153}$$

A função $v(p,m)$ que nos dá a utilidade máxima alcançável a preços e renda é chamada de função de utilidade indireta e, portanto, é uma função composta de $u(\cdot)$ e $x(p,m)$, ou seja,

$$v(p,m) = u(x(p,m)). \tag{1.154}$$

Essa função é denominada "indireta" porque os consumidores geralmente consideram suas preferências em termos do que consomem em vez do preço (como é usado na função). A função de utilidade indireta é de particular importância na teoria microeconômica, pois agrega valor ao desenvolvimento contínuo da teoria da escolha do consumidor e da teoria microeconômica aplicada. Relacionado à função de utilidade indireta está a função dispêndio, que fornece a quantia mínima de dinheiro ou renda que um indivíduo deve gastar para atingir algum nível de utilidade pré-definido. Na microeconomia, a função de utilidade indireta de um consumidor ilustra tanto as preferências do consumidor quanto as condições de mercado prevalecentes e o ambiente econômico.

As propriedades da função de utilidade indireta estão resumidas na seguinte proposição.

Proposição 1.5.1. *Se $u(x)$ é contínua e monotônica no $\mathbb{R}^L_+$ e $(p, m) > 0$, a função de utilidade indireta tem as seguintes propriedades:*

1. *$v(p, m)$ é não-crescente em p e não-decrescente em m.*
2. *$v(p, m)$ é homogênea de grau zero em (p, m).*
3. *$v(p, m)$ é quase-convexa em p.*
4. *$v(p, m)$ é contínua para $p \gg 0$ e $m > 0$.*

Exemplo 1.5.1. *Seja a seguinte função de utilidade (ordenamento de preferências) Cobb-Douglas:*

$$u(x_1, x_2) = x_1^{\alpha_1} x_2^{\alpha_2}, \quad \alpha_1 > 0, \alpha_2 > 0 \tag{1.155}$$

Agora já sabemos que as demandas Marshallianas por x_1 e x_2 são dadas por

$$x_1^*(p_1, p_2, m) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{m}{p_1} \tag{1.156}$$

$$x_2^*(p_1, p_2, m) = \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{m}{p_2} \tag{1.157}$$

Substituindo as demandas ótimas dentro da função de utilidade, obtemos a função de utilidade indireta:

$$\begin{aligned} v(p, m) = u(x(p, m)) &= \left(\frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{m}{p_1} \right)^{\alpha_1} \left(\frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{m}{p_2} \right)^{\alpha_2} \\ &= \left(\frac{m}{\alpha_1 + \alpha_2} \right)^{\alpha_1 + \alpha_2} \left(\frac{\alpha_1}{p_1} \right)^{\alpha_1} \left(\frac{\alpha_2}{p_2} \right)^{\alpha_2} \end{aligned} \tag{1.158}$$

Vamos ver que $v(p, m)$ é não-crescente e p e não-decrescente em m:

$$\frac{\partial v(p, m)}{\partial p_1} = -\left(\frac{m}{\alpha_1 + \alpha_2} \right)^{\alpha_1 + \alpha_2} \left(\frac{\alpha_1}{p_1} \right)^{\alpha_1 + 1} \left(\frac{\alpha_2}{p_2} \right)^{\alpha_2} = -v(p, m) \left(\frac{\alpha_1}{p_1} \right) < 0 \tag{1.159}$$

$$\frac{\partial v(p, m)}{\partial p_2} = -\left(\frac{m}{\alpha_1 + \alpha_2} \right)^{\alpha_1 + \alpha_2} \left(\frac{\alpha_1}{p_1} \right)^{\alpha_1} \left(\frac{\alpha_2}{p_2} \right)^{\alpha_2 + 1} = -v(p, m) \left(\frac{\alpha_2}{p_2} \right) < 0 \tag{1.160}$$

$$\frac{\partial v(p, m)}{\partial m} = \left(\frac{m}{\alpha_1 + \alpha_2} \right)^{\alpha_1 + \alpha_2 - 1} \left(\frac{\alpha_1}{p_1} \right)^{\alpha_1} \left(\frac{\alpha_2}{p_2} \right)^{\alpha_2} = v(p, m) \left(\frac{\alpha_1 + \alpha_2}{m} \right) > 0 \tag{1.161}$$

Vamos provar que $v(p, m)$ é homogênea de grau zero em (p, m):

$$
\begin{aligned}
v(\lambda p, \lambda m) &= \left(\frac{\lambda m}{\alpha_1 + \alpha_2}\right)^{\alpha_1+\alpha_2} \left(\frac{\alpha_1}{\lambda p_1}\right)^{\alpha_1} \left(\frac{\alpha_2}{\lambda p_2}\right)^{\alpha_2} \\
&= \left(\frac{\lambda^{\alpha_1+\alpha_2}}{\lambda^{\alpha_1+\alpha_2}}\right) \left(\frac{m}{\alpha_1 + \alpha_2}\right)^{\alpha_1+\alpha_2} \left(\frac{\alpha_1}{p_1}\right)^{\alpha_1} \left(\frac{\alpha_2}{p_2}\right)^{\alpha_2} \\
&= \lambda^0 \left(\frac{m}{\alpha_1 + \alpha_2}\right)^{\alpha_1+\alpha_2} \left(\frac{\alpha_1}{p_1}\right)^{\alpha_1} \left(\frac{\alpha_2}{p_2}\right)^{\alpha_2} \\
&= \lambda^0 v(p, m)
\end{aligned} \tag{1.162}
$$

Lembre-se que se uma função $\mathcal{C}^2$ num subconjunto convexo W de $\mathbb{R}^2$ é monótona ($F'_x > 0, F'_y > 0$) e se o determinante

$$
\begin{vmatrix}
0 & F'_x & F'_y \\
F'_x & F''_{xx} & F''_{xy} \\
F'_y & F''_{yx} & F''_{yy}
\end{vmatrix}
$$

é negativo para cada x, y, então F é quase-convexa em W.

Calculando o determinante para essa função de utilidade indireta, temos que:

$$
\begin{vmatrix}
0 & -v(p,m)\left(\dfrac{\alpha_1}{p_1}\right) & -v(p,m)\left(\dfrac{\alpha_2}{p_2}\right) \\
-v(p,m)\left(\dfrac{\alpha_1}{p_1}\right) & v(p,m)\left(\dfrac{\alpha_1(\alpha_1+1)}{p_1^2}\right) & v(p,m)\left(\dfrac{\alpha_1\alpha_2}{p_1p_2}\right) \\
-v(p,m)\left(\dfrac{\alpha_2}{p_2}\right) & v(p,m)\left(\dfrac{\alpha_1\alpha_2}{p_1p_2}\right) & v(p,m)\left(\dfrac{\alpha_2(\alpha_2+1)}{p_2^2}\right)
\end{vmatrix} \tag{1.163}
$$

que é igual a $-\alpha_1\alpha_2(\alpha_1+\alpha_2)\dfrac{(v(p,m))^3}{(p_1p_2)^2} < 0.$

Notamos que se as preferências satisfizerem a hipótese de não-saciedade local, então $v(p, m)$ será estritamente crescente em m. Podemos então inverter a função e resolver m em função do nível de utilidade; isto é, dado qualquer nível de utilidade, você pode encontrar a quantidade mínima de renda necessária para alcançar a utilidade u a preços p. A função que relaciona renda e utilidade dessa maneira - o inverso da função de utilidade indireta - é conhecida como a função dispêndio e é denotada por $e(p, u)$. Formalmente, a função dispêndio é dada pelo seguinte problema:

$$
\begin{aligned}
e(p,u) &= \min px \\
\text{sujeito a} \quad & u(x) \geqq \overline{u}.
\end{aligned} \tag{1.164}
$$

A função dispêndio dá o custo mínimo de atingir um nível fixo de utilidade. A solução do problema acima é função de (p, u) e é denotada por $h(p, u)$ e chamada de função de demanda de Hicks. A função de demanda Hicksiana nos diz qual a cesta de consumo atinge um determinado nível de utilidade e minimiza o gasto total.

Uma função de demanda Hicksiana é às vezes chamada de função de demanda compensada. Essa terminologia vem de ver a função demanda como sendo construída por preços e renda variáveis, de modo a manter o consumidor em um nível fixo de utilidade. Assim, as mudanças de renda são organizadas para "compensar" as mudanças de preço. Assim, $h(p, u)$ mantém o nível de u fixo à medida que os preços variam, em contraste com $x(p, m)$, a qual mantém a renda fixa enquanto varia o nível de utilidade. As funções de demanda Hicksianas não são diretamente observáveis, pois dependem da utilidade, que não é diretamente observável. As funções de demanda expressas em função de preços e renda são observáveis; quando queremos enfatizar a diferença entre a função de demanda de Hicks e a função de demanda usual, nos referiremos a esta última como a função de demanda Marshalliana, $x(p, m)$. A função de demanda Marshalliana é apenas a função ordinária de demanda de mercado que estivemos discutindo o tempo todo.

Proposição 1.5.2. *Se $u(x)$ é contínua e monotônica no $\mathbb{R}^L_+$ e $(p, m) > 0$, a função dispêndio tem as seguintes propriedades:*

1. *$e(p, u)$ é não-decrescente em p.*
2. *$e(p, u)$ é homogênea de grau 1 em p.*
3. *$e(p, u)$ é côncava em p.*
4. *$e(p, u)$ é contínua em p, para $p \gg 0$.*
5. *Para todo $p > 0$, $e(p, u)$ é estritamente crescente em u.*
6. *Lema de Shephard: se $h(p, u)$ é a cesta que minimiza o dispêndio e é necessária para atingir o nível de utilidade u aos preços p, então $h_i(p, u) = \dfrac{\partial e(p, u)}{\partial p_i}$, assumindo que $p_i > 0$ e que a derivada existe.*

Nós agora damos algumas propriedades básicas das funções de demanda Hicksiana.

Proposição 1.5.3 (A matriz de substituição é negativa semidefinida)**.** *A matriz dos termos de substituição, $\left(\dfrac{\partial h_j(p, u)}{\partial p_i} = \dfrac{\partial^2 e(p, u)}{\partial p_i p_j}\right)$, é negativa semidefinida porque a função dispêndio é côncava. Portanto, a matriz é simétrica e os termos diagonais são não-positivos.*

A demanda Hicksiana é negativamente inclinada:

$$\frac{\partial h_i(p, u)}{\partial p_i} = \frac{\partial^2 e(p, u)}{\partial p_i^2} \leqq 0. \tag{1.165}$$

Exemplo 1.5.2. *Seja a seguinte função de utilidade (ordenamento de preferências) Cobb-Douglas:*

$$u(x_1, x_2) = x_1^{\alpha_1} x_2^{\alpha_2}, \quad \alpha_1 > 0, \alpha_2 > 0 \tag{1.166}$$

Agora já sabemos que as demandas Marshallianas por x_1 e x_2 são dadas por

$$x_1^*(p_1, p_2, m) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{m}{p_1} \tag{1.167}$$

$$x_2^*(p_1, p_2, m) = \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{m}{p_2} \tag{1.168}$$

Substituindo as demandas ótimas dentro da função de utilidade, obtemos a função de utilidade indireta:

$$\begin{aligned} v(p, m) = u(x(p, m)) &= \left(\frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{m}{p_1}\right)^{\alpha_1} \left(\frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{m}{p_2}\right)^{\alpha_2} \\ &= \left(\frac{m}{\alpha_1 + \alpha_2}\right)^{\alpha_1 + \alpha_2} \left(\frac{\alpha_1}{p_1}\right)^{\alpha_1} \left(\frac{\alpha_2}{p_2}\right)^{\alpha_2} \end{aligned} \tag{1.169}$$

Invertendo a função acima, obtemos a função dispêndio:

$$e(p, u) = (\alpha_1 + \alpha_2) \left[u \left(\frac{p_1}{\alpha_1}\right)^{\alpha_1} \left(\frac{p_2}{\alpha_2}\right)^{\alpha_2}\right]^{\frac{1}{(\alpha_1 + \alpha_2)}} \tag{1.170}$$

A função dispêndio obtida a partir de preferências Cobb-Douglas é não-decrescente em p:

$$\frac{\partial e(p, u)}{\partial p_1} = \left[u \left(\frac{\alpha_1}{\alpha_2} \frac{p_2}{p_1}\right)^{\alpha_2}\right]^{\frac{1}{\alpha_1 + \alpha_2}} > 0 \tag{1.171}$$

$$\frac{\partial e(p, u)}{\partial p_2} = \left[u \left(\frac{\alpha_2}{\alpha_1} \frac{p_1}{p_2}\right)^{\alpha_1}\right]^{\frac{1}{\alpha_1 + \alpha_2}} > 0 \tag{1.172}$$

A função dispêndio é homogênea de grau 1 em p:

$$\begin{aligned} e(\lambda p, u) &= (\alpha_1 + \alpha_2) \left[u \left(\frac{\lambda p_1}{\alpha_1}\right)^{\alpha_1} \left(\frac{\lambda p_2}{\alpha_2}\right)^{\alpha_2}\right]^{\frac{1}{(\alpha_1 + \alpha_2)}} \\ &= \left(\lambda^{\alpha_1 + \alpha_2}\right)^{\frac{1}{(\alpha_1 + \alpha_2)}} (\alpha_1 + \alpha_2) \left[u \left(\frac{p_1}{\alpha_1}\right)^{\alpha_1} \left(\frac{p_2}{\alpha_2}\right)^{\alpha_2}\right]^{\frac{1}{(\alpha_1 + \alpha_2)}} \\ &= \lambda^1 (\alpha_1 + \alpha_2) \left[u \left(\frac{p_1}{\alpha_1}\right)^{\alpha_1} \left(\frac{p_2}{\alpha_2}\right)^{\alpha_2}\right]^{\frac{1}{(\alpha_1 + \alpha_2)}} \\ &= \lambda^1 e(p, u) \end{aligned} \tag{1.173}$$

Também sabemos que $e(p, u)$ é côncava em p. Para provar isso, seja $e(p, u)$ uma função $\mathcal{C}^2$ num conjunto aberto convexo W de $\mathbb{R}^n$. Então, $e(p, u)$ é uma função côncava em W se, e somente *se, a matriz hessiana $D^2e(p,u)$ é não-positiva para cada p em W. No nosso exemplo, temos:*

$$D^2e(p,u) = \begin{vmatrix} -\left(\dfrac{\alpha_2}{p_1(\alpha_1+\alpha_2)}\right)h_1(p,u) & \left(\dfrac{\alpha_2}{p_2(\alpha_1+\alpha_2)}\right)h_1(p,u) \\ \left(\dfrac{\alpha_1}{p_1(\alpha_1+\alpha_2)}\right)h_2(p,u) & -\left(\dfrac{\alpha_1}{p_2(\alpha_1+\alpha_2)}\right)h_2(p,u) \end{vmatrix} \tag{1.174}$$

Vemos que

$$-\left(\frac{\alpha_2}{p_1(\alpha_1+\alpha_2)}\right)h_1(p,u) < 0 \tag{1.175}$$

$$-\left(\frac{\alpha_1}{p_2(\alpha_1+\alpha_2)}\right)h_2(p,u) < 0 \tag{1.176}$$

$$\left(\frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2p_1p_2}p_1p_2\right)h_1(p,u)h_2(p,u) - \left(\frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2p_1p_2}p_1p_2\right)h_1(p,u)h_2(p,u) = 0 \tag{1.177}$$

Assim, os menores principais de ordem ímpar são ≤ 0 e os de ordem par são ≥ 0. Portanto, a função dispêndio é côncava em p.

Para todo $p > 0$, $e(p,u)$ é estritamente crescente em u:

$$\frac{\partial e(p,u)}{\partial u} = u^{\frac{1-(\alpha_1+\alpha_2)}{\alpha_1+\alpha_2}}\left[\left(\frac{p_1}{\alpha_1}\right)^{\alpha_1}\left(\frac{p_2}{\alpha_2}\right)^{\alpha_2}\right] > 0 \tag{1.178}$$

Do lema de Shephard obtemos as demandas Hicksianas:

$$h_1(p,u) = \frac{\partial e(p,u)}{\partial p_1} = \left[u\left(\frac{\alpha_1}{\alpha_2}\frac{p_2}{p_1}\right)^{\alpha_2}\right]^{\frac{1}{\alpha_1+\alpha_2}} \tag{1.179}$$

$$h_2(p,u) = \frac{\partial e(p,u)}{\partial p_2} = \left[u\left(\frac{\alpha_2}{\alpha_1}\frac{p_1}{p_2}\right)^{\alpha_1}\right]^{\frac{1}{\alpha_1+\alpha_2}} \tag{1.180}$$

1.6 Identidades

Existem algumas identidades importantes que unem a função dispêndio, a função de utilidade indireta, a função de demanda Marshalliana e a função de demanda Hicksiana.

Proposição 1.6.1 (Equivalência da Maximização da Utilidade e da Minimização do Dispêndio)**.** *Suponha que a função de utilidade u seja contínua e localmente não saciada, e suponha que $m > 0$. Se as soluções dos dois problemas existem, então os dois problemas acima têm a mesma solução x^*.*

Proposição 1.6.2. *Suponha que a função de utilidade u seja contínua e localmente não saciada, e suponha que $m > 0$. Nós temos:*

- *$e(p, v(p,m)) \equiv m$: o gasto mínimo necessário para atingir a utilidade $v(p,m)$ é m.*
- *$v(p, e(p,u)) \equiv u$: a utilidade máxima da renda $e(p,u)$, $v(p,m)$, é u.*
- *$x_i(p,m) \equiv h_i(p, v(p,m))$: a demanda Marshalliana ao nível de renda m é a mesma que a demanda Hicksiana ao nível de utilidade $v(p,m)$.*
- *$h_i(p,u) \equiv x_i(p, e(p,u))$: a demanda Hicksiana ao nível de utilidade u é a mesma que a demanda Marshalliana ao nível de renda $e(p,u)$.*

Esta última identidade é talvez a mais importante, uma vez que une a função de demanda Marshalliana "observável" com a função de demanda Hicksiana "não observável". Assim, qualquer cesta demandada pode ser expressa como a solução para o problema de maximização de utilidade ou o problema de minimização de despesas.

Proposição 1.6.3 (Identidade de Roy)**.** *Se $x(p,m)$ é a função de demanda Marshalliana, então:*

$$x_i(p,m) = -\frac{\dfrac{\partial v(p,m)}{\partial p_i}}{\dfrac{\partial v(p,m)}{\partial m}}, \tag{1.181}$$

desde que o lado direito esteja bem definido e que $p_i > 0$ e $m > 0$.

Demonstração. Suponha que x^* produz uma utilidade máxima de u^* em (p^*, m^*). Nós sabemos de nossas identidades que

$$x(p^*, m^*) \equiv h(p^*, u^*). \tag{1.182}$$

De outra das identidades fundamentais, também sabemos que

$$u^* \equiv v(p, e(p,u)). \tag{1.183}$$

Como essa é uma identidade, podemos diferenciá-la em relação a p_i para obter:

$$0 = \frac{\partial v(p^*, m^*)}{\partial p_i} + \frac{\partial v(p^*, m^*)}{\partial m}\frac{\partial e(p^*, u^*)}{\partial p_i}. \tag{1.184}$$

Rearranjando, obtemos:

$$x_i(p^*, m^*) \equiv h_i(p^*, u^*) \equiv \frac{\partial e(p^*, u^*)}{\partial p_i} \equiv -\frac{\dfrac{\partial v(p^*, m^*)}{\partial p_i}}{\dfrac{\partial v(p^*, m^*)}{\partial m}}. \tag{1.185}$$

Como essa identidade é satisfeita para todos (p^*, m^*) e como $x^* = x(p^*, m^*)$, o resultado é provado. ■

Exemplo 1.6.1. *Seja a função de utilidade indireta Cobb-Douglas:*

$$v(p,m) = u(x(p,m)) = \left(\frac{\alpha_1}{\alpha_1+\alpha_2}\frac{m}{p_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{\alpha_1+\alpha_2}\frac{m}{p_2}\right)^{\alpha_2} \tag{1.186}$$

Temos que:

$$\frac{\partial v(p^*,m^*)}{\partial p_1} = \left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1+1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2} \tag{1.187}$$

$$\frac{\partial v(p^*,m^*)}{\partial p_2} = \left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2+1} \tag{1.188}$$

$$\frac{\partial v(p^*,m^*)}{\partial m} = \left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2-1}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2} \tag{1.189}$$

Logo,

$$x_1(p,m) = \frac{\left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1+1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2}}{\left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2-1}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2}} = \frac{\alpha_1}{\alpha_1+\alpha_2}\frac{m}{p_1} \tag{1.190}$$

$$x_2(p,m) = \frac{\left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2+1}}{\left(\frac{m}{\alpha_1+\alpha_2}\right)^{\alpha_1+\alpha_2-1}\left(\frac{\alpha_1}{p_1}\right)^{\alpha_1}\left(\frac{\alpha_2}{p_2}\right)^{\alpha_2}} = \frac{\alpha_2}{\alpha_1+\alpha_2}\frac{m}{p_2} \tag{1.191}$$

Exemplo 1.6.2. *Seja a função de utilidade Leontief:*

$$u(x_1,x_2) = \min\{a_1x_1, a_2x_2\} \tag{1.192}$$

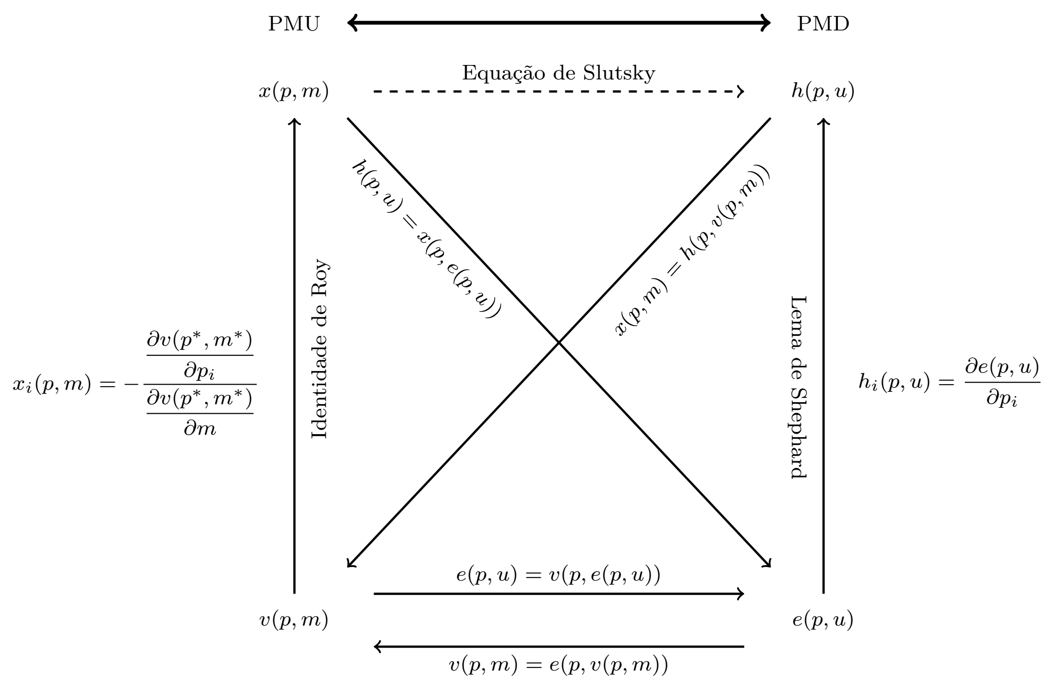
que não é diferenciável, mas tem funão de utilidade indireta:

$$v(p,m) = \frac{a_1a_2m}{a_1p_2+a_2p_1} \tag{1.193}$$

Aplicando a identidade de Roy, temos:

$$x_1(p,m) = -\frac{\frac{\partial v(p^*,m^*)}{\partial p_1}}{\frac{\partial v(p^*,m^*)}{\partial m}} = -\frac{\frac{-a_1a_2^2m}{(a_1p_2+a_2p_1)^2}}{\frac{a_1a_2}{(a_1p_2+a_2p_1)}} = \frac{a_2m}{a_1p_2+a_2p_1} \tag{1.194}$$

Figura 1.32 – Problemas duais



1.7 Propriedades da Demanda

Nesta seção, examinaremos a estatística comparativa do comportamento da demanda do consumidor: como a demanda do consumidor muda à medida que os preços e a renda mudam.

1.7.1 Mudanças na Renda

Seja a renda variável, mas as preferências e os preços (relativos) dos bens constantes. Devido a essa mudança, a linha orçamentária se desloca paralelamente para si mesma. O ótimo do consumidor é o ponto em que a nova linha orçamental é tangente à curva de indiferença mais alta possível.

É interessante observar como a demanda do consumidor muda à medida que mantemos os preços fixos e permitimos que a renda varie; o locus resultante de cestas que maximizam a utilidade é conhecido como o caminho de expansão da renda. A partir do caminho da expansão de renda, podemos derivar uma função que relaciona a renda à demanda de cada bem (a preços constantes). Essas funções são chamadas de curvas Engel. Há duas possibilidades: (i) à medida que a renda aumenta, o consumo ótimo de um bem aumenta. Esse bem é chamado de bem normal; (ii) à medida que a renda aumenta, o consumo ótimo de um bem diminui. Esse bem é chamado bem inferior. Para o problema da maximização do consumo de dois bens, quando a trajetória de expansão de renda (e, portanto, cada curva de Engel) é crescente, os dois bens são bens normais. Quando a trajetória de expansão de renda pode se dobrar para trás, há um e somente um bem que é inferior quando a função de utilidade é localmente não saciada; um aumento na renda significa que o consumidor realmente quer consumir menos do bem.

Figura 1.33 – Curva renda-consumo: efeito renda é igual para os dois bens

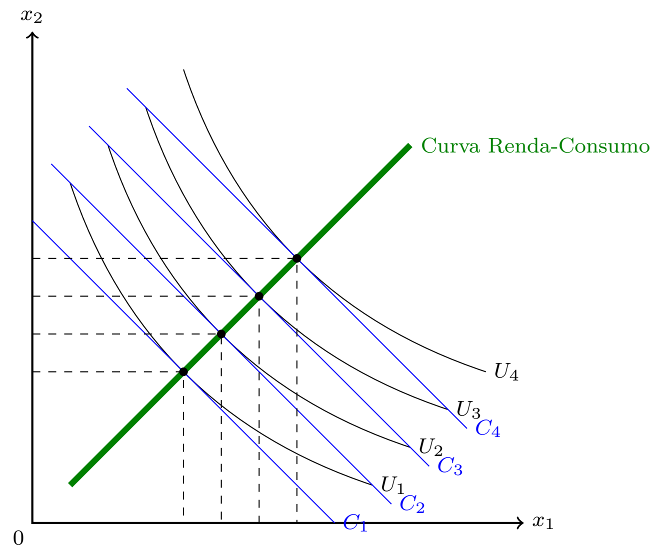


Figura 1.34 – Curva renda-consumo: x_1 é um bem necessário x_2 é um bem de luxo

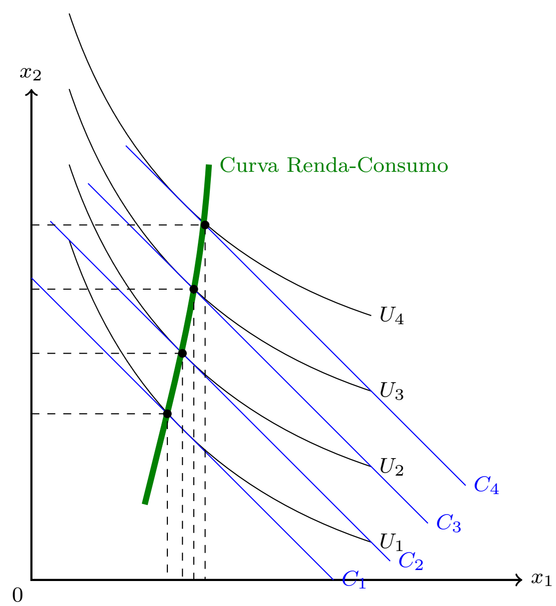
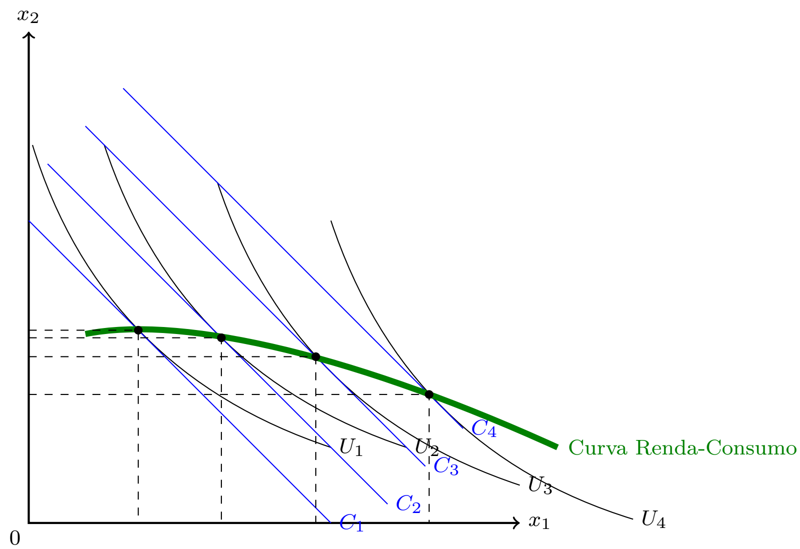


Figura 1.35 – Curva renda-consumo: x_1 é um bem normal x_2 é um bem inferior

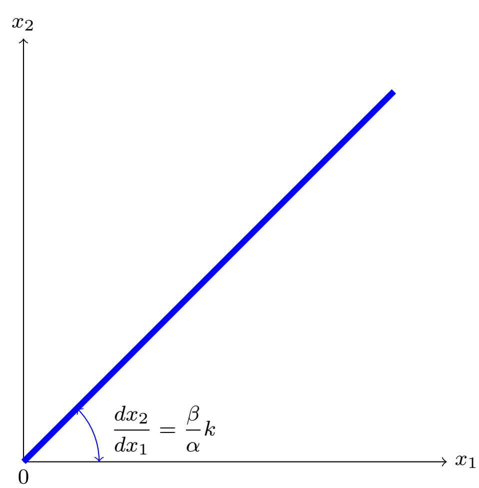


Como derivamos a trajetória de renda-consumo dos bens? Lembramos que $TMgS = \dfrac{p_1}{p_2}$ e escrevemos o resultado no plano $x_1 \times x_2$.

Exemplo 1.7.1. *Seja a função de utilidade Cobb-Douglas $u(x_1, x_2) = x_1^\alpha x_2^\beta$. Sabemos que*

$$\frac{\alpha x_1^{\alpha-1} x_2^\beta}{\beta x_1^\alpha x_2^{\beta-1}} = \frac{p_1}{p_2} \iff \frac{\alpha}{\beta}\frac{x_2}{x_1} = \frac{p_1}{p_2} \iff x_2 = \frac{\beta}{\alpha} k x_1$$

Figura 1.36 – Curva renda-consumo: Cobb-Douglas



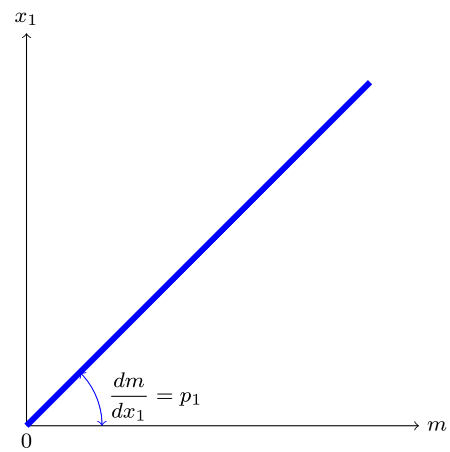
Observe que $\dfrac{dx_2}{dx_1} > 0$. Portanto, a curva de renda-consumo é crescente.

As formas funcionais que retratam a relação entre renda e consumo são chamadas curvas de

Engel. Isto é, a curva de Engel relaciona a quantidade de equilíbrio de um bem adquirida para um dado nível de renda monetário.

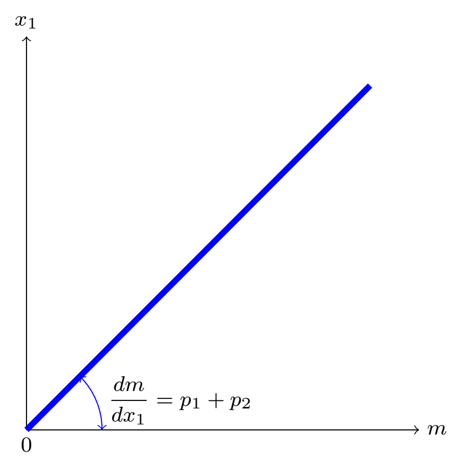
Exemplo 1.7.2 (Bens Substitutos)**.** *Suponha que* $p_1 < p_2$. *O consumidor é especializado em consumir o bem* 1. *A demanda pelo bem* 1 *é* $x_1^*(p_1, p_2, m) = \dfrac{m}{p_1}$. *Logo, a curva de Engel é uma linha reta,* $m = p_1 x_1$, *como inclinação* p_1, *como na Figura 1.37.*

Figura 1.37 – Curva de Engel: bens substitutos



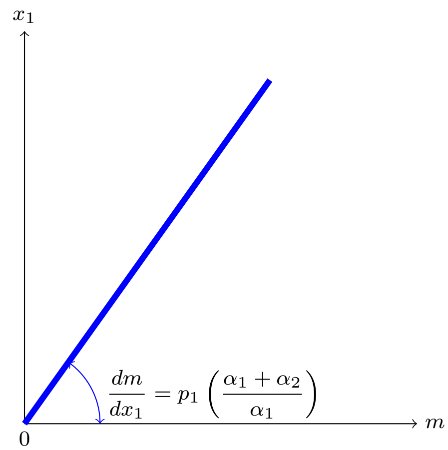
Exemplo 1.7.3 (Bens Complementares)**.** *A demanda pelo bem* 1 *é* $x_1^*(p_1, p_2, m) = \dfrac{m}{p_1 + p_2}$. *Logo, a curva de Engel é uma linha reta,* $m = (p_1 + p_2)x_1$, *com inclinação* $p_1 + p_2$, *como na Figura 1.38.*

Figura 1.38 – Curva de Engel: bens complementares



Exemplo 1.7.4 (Cobb-Douglas)**.** *A demanda pelo bem* 1 *é* $x_1^*(p_1, p_2, m) = \dfrac{\alpha_1}{\alpha_1 + \alpha_2}\dfrac{m}{p_1}$. *Logo, a curva de Engel é uma linha reta,* $m = \left(\dfrac{\alpha_1 + \alpha_2}{\alpha_1}\right) p_1 x_1$, *com inclinação* $p_1 \left(\dfrac{\alpha_1 + \alpha_2}{\alpha_1}\right)$, *como na Figura 1.39.*

Figura 1.39 – Curva de Engel: Cobb-Douglas



Vamos ver o mesmo conceito por meio da ideia de elasticidade-renda. A elasticidade-renda da demanda é a varição proporcional no consumo de um bem para uma dada variação proporcional na renda. De acordo com o formato da curva de Engel, conforme a Figura 1.40, os bens podem ser classificados como normais ou inferiores. Também temos que os bens podem ser luxuosos (superior) ou necessários, que são casos particulares de bens normais.

Figura 1.40 – Curvas de Engel

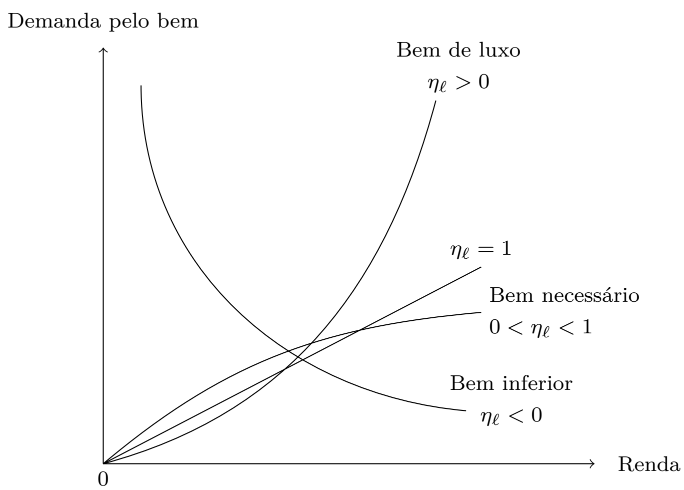
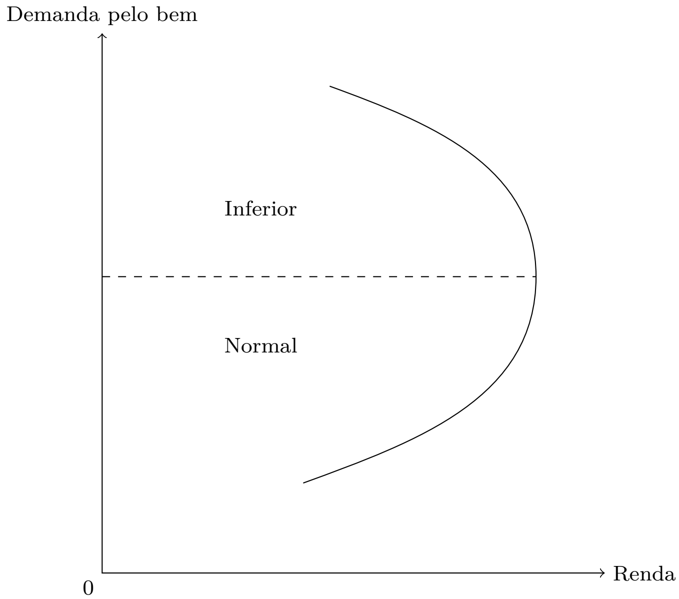


Figura 1.41 – Bem normal e bem inferior



A elasticidade-renda pode ser definida pela seguinte fórmula:

$$\eta_\ell = \frac{dx_\ell}{dm}\frac{m}{x_\ell}. \tag{1.195}$$

Assim, temos os seguintes casos:

1. Se $\frac{dx_\ell}{dm}$ for maior que zero temos um bem normal.
   (a) Se $\frac{dx_\ell}{dm}$ for maior que zero e menor que a unidade temos um bem necessário.
   (b) Se $\frac{dx_\ell}{dm}$ for maior que a unidade temos um bem superior.
2. Se $\frac{dx_\ell}{dm}$ for menor que zero temos um bem inferior.
3. Se $\frac{dx_\ell}{dm}$ for nulo o bem é neutro.

Note que todos os bens podem ser normais, mas nem todos podem ser inferiores. Para mostrar isso usando elasticidades, temos:

$$p_1x_1 + p_2x_2 + \ldots + p_\ell x_\ell = m. \tag{1.196}$$

Diferenciando com relação à renda, temos:

$$p_1\frac{dx_1}{dm} + p_2\frac{dx_2}{dm} + \ldots + p_\ell\frac{dx_\ell}{dm} = 1$$

$$p_1 \frac{dx_1}{dm} \frac{x_1}{x_1} \frac{m}{m} + p_2 \frac{dx_2}{dm} \frac{x_2}{x_2} \frac{m}{m} + \ldots + p_\ell \frac{dx_\ell}{dm} \frac{x_\ell}{x_\ell} \frac{m}{m} = 1$$
$$\frac{p_1 x_1}{m} \frac{dx_1}{dm} \frac{m}{x_1} + \frac{p_2 x_2}{m} \frac{dx_2}{dm} \frac{m}{x_2} + \ldots + \frac{p_\ell x_\ell}{m} \frac{dx_\ell}{dm} \frac{m}{x_\ell} = 1$$
$$\theta_1 \eta_1 + \theta_2 \eta_2 + \ldots + \theta_\ell \eta_\ell = 1. \tag{1.197}$$

Dado que $\theta_1, \theta_2, \ldots, \theta_\ell \geq 0$, não podemos ter $\eta_i < 0$ para todos os bens e ainda assim a expressão ser igual a 1. Isso significa que não é possível que todos os bens sejam inferiores; alguns, devem ser normais.

Exemplo 1.7.5. *Sabemos que da função de utilidade Cobb-Douglas $u(x_1, x_2) = x_1^\alpha x_2^\beta$, as demandas ótimas serão:*

$$x_1^* = \frac{\alpha}{\alpha + \beta} \frac{m}{p_1} \tag{1.198}$$
$$x_2^* = \frac{\alpha}{\alpha + \beta} \frac{m}{p_2} \tag{1.199}$$

Calculando as elasticidades-renda da demanda por cada bem, encontramos:

$$\eta_1 = \frac{dx_1}{dm} \frac{m}{x_1} = \frac{\alpha}{\alpha + \beta} \frac{1}{p_1} \frac{m}{\frac{\alpha}{\alpha + \beta} \frac{m}{p_1}} = 1 \tag{1.200}$$
$$\eta_2 = \frac{dx_2}{dm} \frac{m}{x_2} = \frac{\beta}{\alpha + \beta} \frac{1}{p_2} \frac{m}{\frac{\beta}{\alpha + \beta} \frac{m}{p_2}} = 1 \tag{1.201}$$

Assim, verificamos que os bens são normais.

Também verificamos que:

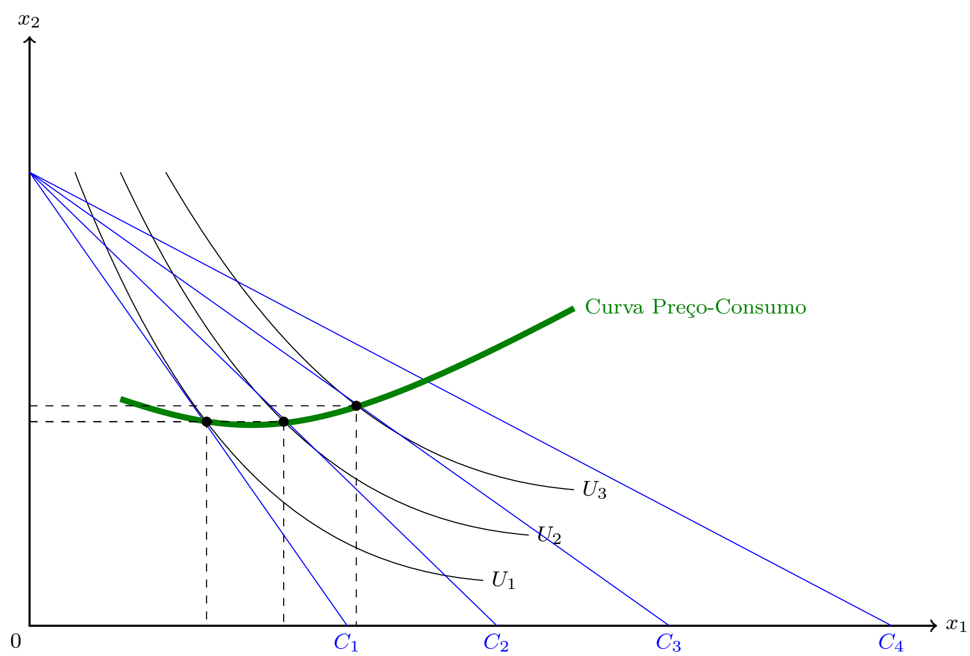
$$\theta_1 \eta_1 + \theta_2 \eta_2 = 1$$
$$1 \left(\frac{\alpha}{\alpha + \beta} \right) + 1 \left(\frac{\beta}{\alpha + \beta} \right) = 1 \tag{1.202}$$

1.7.2 Mudanças no Preço

Também podemos manter a renda fixa e permitir que os preços variem. Se deixarmos p_1 e m fixos e variarmos p_2, o lócus das tangências varrerá uma curva conhecida como a curva de oferta de renda. No caso ordinário em que um preço mais baixo para o bem 1 leva a uma maior demanda pelo bem, a Lei da Demanda é satisfeita; no caso em que uma diminuição no preço do bem 1 provoca uma diminuição da demanda pelo bem 1, esse bem é chamado de bem de Giffen.

A curva de oferta é uma forma alternativa de descrever o comportamento de demanda de um indivíduo, ou seja, sua função de demanda. E resumindo o comportamento da demanda individual, também podemos usar a curva de oferta para descrever a função de demanda do mercado. A curva de oferta é geralmente bem definida para qualquer número de mercadorias, mas queremos nos concentrar no caso de dois bens. A ideia por trás da curva de oferta é descrever o comportamento de demanda do indivíduo no mesmo espaço que usamos para descrever suas preferências (seu mapa de indiferença) - ou seja, o espaço da mercadoria, que, no caso dos dois bens, é bidimensional. Quando você altera o preço de um dos bens (digamos, o bem x_1) mantendo constante a riqueza do consumidor e o preço do outro bem, você gira geometricamente a restrição orçamentária e traça o lócus dos feixes (x_1, x_2) no qual o consumidor compraria aos vários preços p_1, como na Figura 1.42. Essa curva preço-consumo mostra exatamente as cestas que esse consumidor potencialmente compraria nos vários preços possíveis p_1; todos as outras cestas em seu espaço de bens são aqueles que ele não compraria, a qualquer preço (assumindo que o preço do outro bem e a riqueza do consumidor, p_2 e m permaneçam fixos em seus valores originais).

Figura 1.42 – Curva preço-consumo

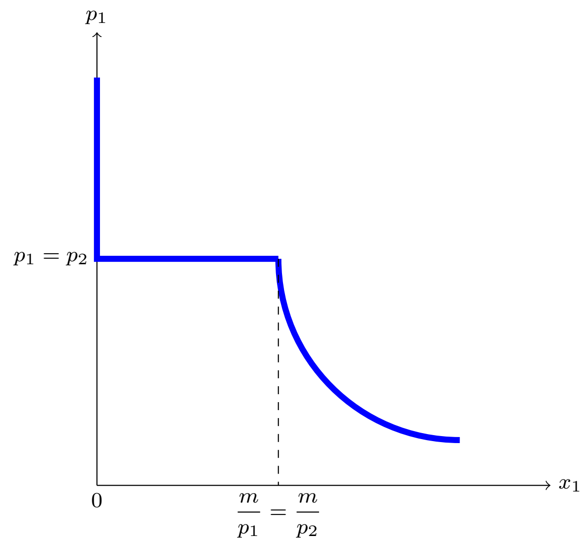


Exemplo 1.7.6 (Bens Substitutos)**.** *A demanda pelo bem* 1 *é*

$$x_1^*(p_1, p_2, m) = \begin{cases} \dfrac{m}{p_1}, & se\ p_1 < p_2 \\ \left[0, \dfrac{m}{p_1}\right), & se\ p_1 = p_2 \\ 0, & se\ p_1 > p_2 \end{cases} \tag{1.203}$$

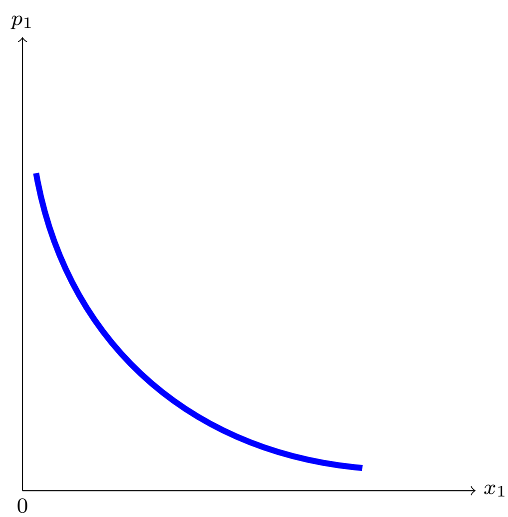
A função demanda pelo bem 1 *é mostrada na Figura 1.43.*

Figura 1.43 – Curva de demanda: bens substitutos



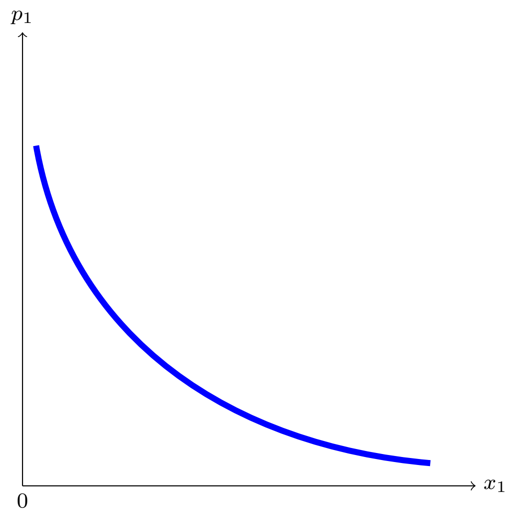
Exemplo 1.7.7 (Bens Complementares)**.** *A demanda pelo bem 1 é* $x_1^*(p_1, p_2, m) = \dfrac{m}{p_1 + p_2}$*. Logo, a curva de demanda é dada pela Figura 1.44. O formato decorre do fato de que* $\dfrac{\partial x_1^*(p_1, p_2, m)}{\partial p_1} = -\dfrac{m}{(p_1 + p_2)^2} < 0$.

Figura 1.44 – Curva de demanda: bens complementares



Exemplo 1.7.8 (Cobb-Douglas)**.** *A demanda pelo bem 1 é* $x_1^*(p_1, p_2, m) = \dfrac{\alpha_1}{\alpha_1 + \alpha_2} \dfrac{m}{p_1}$*, como mostrado na Figura 1.45. O formato decorre do fato de que* $\dfrac{\partial x_1^*(p_1, p_2, m)}{\partial p_1} = -\left(\dfrac{\alpha_1}{\alpha_1 + \alpha_2}\right)\left(\dfrac{m}{p_1^2}\right) < 0$.

Figura 1.45 – Curva de demanda: Cobb-Douglas



Um bem de Giffen é um bem para o qual a curva de demanda apresenta uma região de preços na qual a quantidade demandada varia positivamente com os preços. Logo, ser de Giffen é uma propriedade local. E não é uma propriedade do bem em si. Não é o bem que é de Giffen, são as preferências do consumidor que têm uma forma tal que a demanda dela resultante apresenta a característica de Giffen para alguma região de preços.

O bem de Giffen viola, assim, a primeira lei fundamental da demanda, segundo a qual a quantidade demandada varia inversamente com o preço. O preço de demanda, isto é, o preço que um consumidor está disposto a pagar por uma unidade qualquer do bem, reflete o valor que ele atribui às alternativas sacrificadas para a obtenção dessa unidade. Afinal, se não houvesse sacrifício para a aquisição e consumo da unidade adicional, então o bem não seria escasso e, com isso, teria preço zero, o que nunca acontece.

Suponha que, para um consumidor, um bem é de Giffen e que essa propriedade se verifica para as quantidades iniciais consumidas. Inicialmente ele está disposto a pagar \$100 pela 1ª unidade. O que isso quer dizer? Em primeiro lugar, se é a 1ª unidade do bem, então é porque antes ele não consumia esse bem. Sua situação então é de ajuste na margem extensiva, ele está entrando nesse mercado. Para consumir essa 1ª unidade do bem, ele sacrificou o consumo de outros bens em um valor de \$100. Logo, para consumir a 1ª unidade do bem, ele já abriu mão de outros bens, já que a renda para gastar é constante. Ele agora pensa em consumir a 2ª unidade do bem, já um ajuste na margem intensiva. Para isso, ele tem que sacrificar novamente o consumo de outros bens. Qual o valor do que ele está disposto a sacrificar em troca da 2ª unidade do bem? Digamos que \$110. Então da 1ª para a 2ª unidade o preço de demanda vai de \$100 para \$110. Nesse trecho, a curva de demanda é crescente. Vejamos a razão pela qual isso contraria a lei de demanda. O nosso consumidor atribui um valor à primeira unidade do bem. Esse valor é \$100. É importante entender que esses \$100 representam o valor dos bens cujo consumo foi sacrificado para o consumo da 1ª unidade do bem em questão. Ele economiza por um lado para gastar por outro. Então ele já consumiu a 1ª unidade. Para consumir a 2ª unidade, ele está disposto a pagar \$110. O valor do que ele está disposto a sacrificar para consumir a 2ª unidade aumentou! Isso significa uma violação do princípio da utilidade marginal decrescente para esse bem. Em outras palavras, a 2ª unidade vale tão mais que a 1ª, que ele está disposto a sacrificar ainda mais dos bens restantes em troca desse bem. E os bens restantes, apesar de se tornarem mais escassos a cada vez, ainda assim perdem valor, pois são facilmente trocados pelas unidades adicionais dos bens. É uma violação geral do princípio da utilidade marginal decrescente!

O princípio da utilidade marginal decrescente é uma lei geral justamente porque é válida sob a égide do aspecto temporal. Essa temporalidade essencial da lei de demanda é a chave para mostrar que a constatação observacional de um bem de Giffen pode ser apenas uma consideração defeituosa ou incompleta do aspecto temporal da demanda.

Suponha que o preço de um bem cai. O consumidor pode pensar assim: "O preço caiu. Tenho razões para esperar que o preço vai cair ainda mais e, portanto, vou reter meu consumo até o preço cair mais. Como isso não deve demorar muito, o custo de oportunidade de não consumir agora e esperar não é maior do que o benefício que eu espero obter ao consumir mais tarde ao preço menor. Por isso, é racional que eu retenha o consumo mesmo com a atual queda de preço".

O efeito disso é uma queda na quantidade agregada demandada, pois a disposição marginal a pagar pelas unidades adicionais caiu. Um observador externo pode achar que esse bem apresenta características de bem de Giffen, isto é, que uma lei básica da demanda foi violada e que o consumidor é, na verdade, irracional.

Mas é essa a maneira correta de um economista interpretar o fenômeno? Não! A decisão do consumidor não foi a de consumir uma quantidade menor em razão da queda observada do preço hoje. A decisão verdadeira é a de aumentar a quantidade consumida após um horizonte mais ou menos determinado de tempo em função da queda esperada do preço ao fim desse horizonte de tempo. Logo, a lei básica da demanda não foi violada.

O problema é que o observador externo não entende que o indivíduo, ao tomar uma decisão hoje, o faz com o propósito de mudar seu estado no futuro. Não tem sentido pressupor que a decisão econômica tem por propósito a mudança do passado. E mesmo que se admita que possa mudar o seu estado instantaneamente, no momento da queda de preço, é preciso admitir que o consumidor também vislumbra possíveis estados futuros.

Existe na literatura uma rationale para o bem de Giffen que poderia retirar da explicação acima o seu caráter de generalidade: bens indivisíveis. Rod Garrat ("Indivisibilities, inferior goods, and Giffen goods", Canadian Journal of Economics, 30: 246–251 (1997)) provou que, se um bem é indivisível e outro não, então a indivisibilidade pode gerar demanda com natureza de Giffen. A ideia vem de John Hicks (A Revision of Demand Theory, 1959) e é simples. Suponha que um aumento de renda é suficiente para induzir um consumidor a comprar um carro. Mas a compra do carro traz consigo uma série de outros gastos adicionais que, ao fim, acabam reduzindo o dispêndio em vários outros bens de forma significativa. Todos esses bens sacrificados pelo consumidor em troca do consumo do carro acabam apresentando característica de bens inferiores. Ora, para ser um bem de Giffen, é necessário que primeiro o bem seja inferior. Portanto, a inferioridade de uma gama enorme de bens pode fazer com que alguns deles se tornem de Giffen. A razão disso está no fato de que o carro é um bem indivisível. O aumento de renda tem de ser suficiente para induzir a compra do carro, mas, ao ocorrer, transfere para o carro parte do dispêndio com outros bens.

A indivisibilidade do bem esconde o horizonte de tempo relevante para a decisão de comprar o carro ou não. Esse horizonte é de anos. As quantidades demandadas por cada bem por um consumidor, para efeito de análise com base nas taxas marginais de troca, elasticidades diretas e cruzadas, ou seja, qualquer análise microeconômica coerente, essas quantidades, como dizia, devem vir expressas em unidades demandadas ao longo da MESMA unidade de tempo. Se não for assim, as quantidades são incomparáveis. É assim na produção e é assim também no consumo. Ora, a janela de tempo associada ao carro, que é o bem indivisível, é de vários anos. Entretanto, os demais bens estão sendo avaliados instantaneamente! A forma correta de avaliá-los é ver como são demandados ao longo dos mesmos vários anos. Para além dessa longa janela de tempo, o dispêndio médio nos outros bens, a partir do aumento de renda hoje, é constante. Não apresentarão, portanto, natureza de bem inferior e muito menos de Giffen. Não há absolutamente sentido algum em comparar a alternativa de comprar um carro, cujo valor é obtido pela diferença entre os valores presentes de benefícios e custos de um fluxo de, digamos, 5 anos, e o consumo vislumbrado, hoje, de arroz e feijão pelo mês à frente. Deve-se comparar com o consumo de arroz e feijão por 5 anos. Assim, a taxa de troca é entre o carro por 5 anos e comida por 5 anos. O consumidor consome uma unidade de carro, mas o consumo mensal de comida deve ser reescalado para o lustro.

1.7.3 Efeitos Renda e Substituição: A Equação de Slutsky

Já introduzimos as noções de taxa marginal de transformação a partir da restrição orçamentária e de taxa marginal de substituição a partir das preferências do consumidor. É a relação entre os preços que determina a taxa marginal de transformação e, portanto, determina a localização ideal ao longo da linha de orçamento para o consumidor. O problema, no entanto, é que se começarmos a partir de um ponto de solução sob um orçamento específico e um preço mudar, a taxa marginal de transformação não é a única coisa a mudar. Não é possível alterar a inclinação de uma linha sem alterar sua posição inteira, exceto um ponto. Para a mudança de preço típica, esse ponto é um intercepto, o que significa que a alteração de preço inclui um efeito significativo de riqueza ou rendimento. Vamos imaginar a mudança na linha do orçamento a partir de uma mudança de preço ocorrendo em duas etapas:

- uma rotação para a nova taxa marginal de transformação girando através da escolha original (e, portanto, eliminando o efeito renda)
- seguida de uma mudança paralela para a nova linha orçamentária, que "adiciona" de volta o efeito renda que nós eliminamos na rotação.

Se a mudança de preço realmente produziu uma nova linha orçamentária girando em torno do seu antigo ponto de solução, você se moveria na direção do bem cujo preço relativo caiu. Em segundo lugar, o deslocamento paralelo. Nesta nota, decomporemos a resposta da demanda às mudanças de preço em efeitos de substituição e efeitos de renda. O efeito substituição é a direção e a distância que um consumidor moveria se a linha orçamentária mudasse para sua nova inclinação ditada pelos novos preços, mas fê-lo girando em cima da solução de demanda original. Esta hipotética linha orçamentária intermediária é a linha orçamentária compensada de Slutsky. Dito de outra forma, o efeito substituição representa o movimento ao longo da curva de indiferença. O efeito renda é o restante da resposta à mudança de preço. É a direcção e magnitude da resposta da linha orçamentária compensada à nova linha orçamentária.

Seja $x_1(p_1, p_2, m)$ a demanda pelo bem 1 quando o preço do bem 1 é p_1, o preço do bem 2 é p_2 e m é o nível de renda do consumidor. Considere que o preço do bem 1 mude de p_1^0 para p_1^1. Pode ser um aumento ou uma redução do preço do bem. O efeito total da mudança de preço na demanda do bem 1 é definida como segue:

$$\text{Efeito Total} = \underbrace{x_1(p_1^1, p_2, m)}_{\text{Nova Demanda}} - \underbrace{x_1(p_1^0, p_2, m)}_{\text{Velha Demanda}} . \tag{1.204}$$

Observe que, se o bem for normal, e a alteração de preço for uma redução de preço, o efeito total será uma quantidade positiva. Se a variação de preço for o aumento de preço, o efeito total será uma quantidade negativa.

O primeiro passo em nossa decomposição é construir um orçamento compensado, (p_1^1, p_2, m^c). É extremamente importante ter em mente que esta linha orçamentária é hipotética. Isto é, tudo está em nossas cabeças. O consumidor nunca experimentou este orçamento em momento algum na transição de preços, embora nos perguntemos o que aconteceria se o fizesse.

Neste orçamento compensado, os preços serão p_1^1 e p_2. Em outras palavras, a taxa marginal de transformação dado este orçamento hipotético será a nova taxa marginal de transformação, uma vez que p_1^1 é o novo preço do bem 1. A diferença entre o novo orçamento real e este hipotético orçamento compensado é o nível de renda. No orçamento compensado, forneceremos exatamente a renda necessária para comprar a cesta de consumo antiga - ou seja, a cesta de consumo exigido pelo orçamento antigo. Portanto, o nível de renda compensada, m^c, é dado por:

$$m^c = p_1^1 x_1(p_1^0, p_2, m) + p_2 x_2(p_1^0, p_2, m). \tag{1.205}$$

Em palavras, o nível de renda compensado é o novo preço do bem 1 vezes a quantidade do bem 1 exigido sob o preço antigo do bem 1 mais o dispêndio necessário para comprar a quantidade do bem 2 exigida sob o preço antigo do bem 1.

Outra forma de calcular o nível compensado de renda é:

$$m^c = m + (p_1^1 - p_1^0) x_1(p_1^0, p_2, m). \tag{1.206}$$

Pensando desta forma, o nível compensatório de renda, para um aumento de preço, é simplesmente o antigo nível de renda mais o dinheiro adicional necessário para continuar comprando a velha demanda do bem 1. Para uma redução de preço, é o antigo nível de renda menos a economia experimentada na compra da antiga quantidade do bem 1 devido à diminuição do preço.

Se você fosse desenhar a linha de orçamento compensada, ela interceptaria o antigo ponto de demanda - ou seja, a cesta de consumo ótimo sob o preço antigo - e teria a inclinação da nova linha de orçamento, como nas Figuras 1.46 e 1.47

Figura 1.46 – Aumento do preço do bem 1

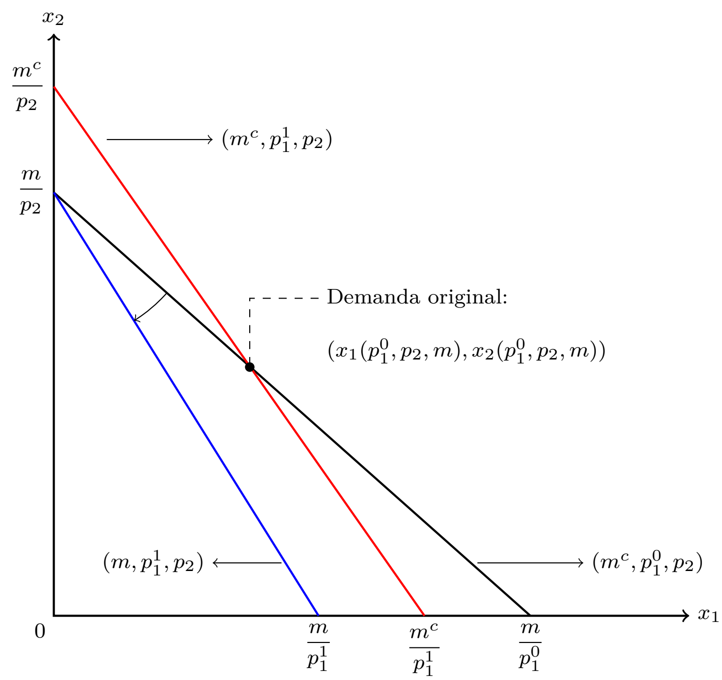
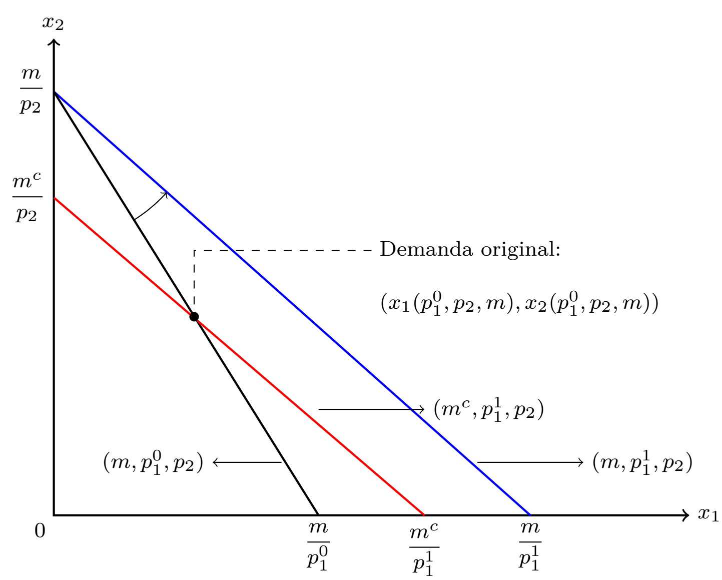


Figura 1.47 – Redução do preço do bem 1



A demanda compensada pelo bem 1, $x_1(p_1^1, p_2, m^c)$, é a solução do seguinte problema:

$$\begin{aligned}(x_1(p_1^1, p_2, m^c), x_2(p_1^1, p_2, m^c)) = \arg \max_{(x_1, x_2)} u(x_1, x_2) \\ \text{sujeito a} \quad p_1^1 x_1 + p_2 x_2 < m^c.\end{aligned} \tag{1.207}$$

As Figuras 1.48 e 1.49 mostram como encontrar a demanda compensada de forma gráfica. Há algumas coisas a se notar sobre a demanda compensada nas imagens abaixo. A primeira delas é a "lei da demanda compensada": as mudanças direcionais necessárias na demanda compensada são resultado da convexidade das preferências. A cesta de consumo ótimo para um aumento do preço do bem 1 (compensado) deve ser para cima e para a esquerda (menos bem 1, mais bem 2) a partir do ponto de solução original. A cesta de consumo ótimo para uma redução do preço do bem 1 (compensada) deve ser para baixo e à direita (mais bem 1, menos bem 2) a partir do ponto de solução original.

As seções verdes da linha orçamentária compensada mostram onde a solução de demanda compensada deve estar de acordo com a solução de demanda original e sua curva de indiferença associada.

Observe que $U^c > U_0$, isto é, a compensação de Slutsky é excessiva no sentido de bem-estar. Note que em ambas as Figuras o nível de utilidade que se alcançaria se confrontado com a linha de orçamento compensada seria pelo menos tão alto e tipicamente mais alto do que o nível de utilidade alcançado sob o orçamento original. Há outra noção de compensação, conhecida como compensação Hicksiana, que corrige isso.

Figura 1.48 – Encontrando a demanda compensada: aumento do preço do bem 1

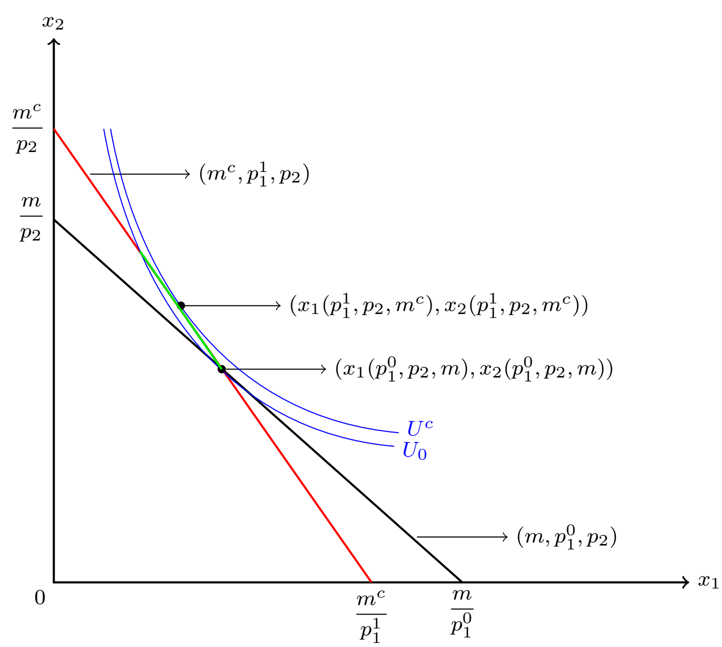
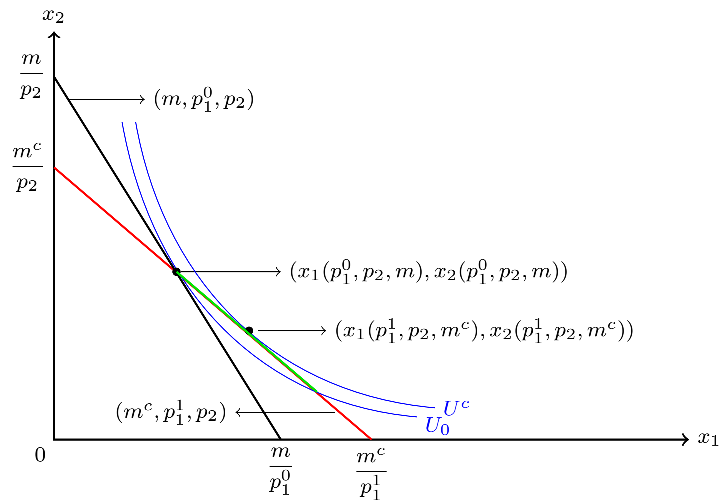


Figura 1.49 – Encontrando a demanda compensada: redução do preço do bem 1



O efeito substituição de uma mudança de preço descreve a transição entre a solução de demanda original e a solução de demanda compensada. O efeito renda descreve a direção e a magnitude da transição da demanda compensada para a nova demanda. Desse modo:

$$\text{Efeito Susbtituição} = \text{Demanda Compensada} - \text{Velha Demanda} \tag{1.208}$$
$$\text{Efeito Renda} = \text{Nova Demanda} - \text{Demanda Compensada}. \tag{1.209}$$

Se olharmos apenas para o bem 1, temos:

$$\text{Efeito Susbtituição} = x_1(p_1^1, p_2, m^c) - x_1(p_1^0, p_2, m) \tag{1.210}$$
$$\text{Efeito Renda} = x_1(p_1^1, p_2, m) - x_1(p_1^1, p_2, m^c). \tag{1.211}$$

Observe que a soma dos efeitos de renda e substituição é o efeito total:

$$\begin{aligned}
\text{Efeito Total} &= \text{Efeito Renda} + \text{Efeito Susbtituição} \\
\text{Nova Demanda} - \text{Velha Demanda} &= [\text{Nova Demanda} - \text{Demanda Compensada}] + \\
&+ [\text{Demanda Compensada} - \text{Demanda Velha}] \\
x_1(p_1^1, p_2, m) - x_1(p_1^0, p_2, m) &= [x_1(p_1^1, p_2, m) - x_1(p_1^1, p_2, m^c)] + \\
&+ [x_1(p_1^1, p_2, m^c) - x_1(p_1^0, p_2, m)]. \qquad (1.212)
\end{aligned}$$

Embora a função de demanda compensada não seja diretamente observável, veremos que sua derivada pode ser facilmente calculada a partir de itens observáveis, ou seja, a derivada da demanda Marshalliana em relação a preço e renda. Essa relação é conhecida como a equação de

Slutsky. Podemos utilizar cálculo para descrever o efeito total de uma mudança de preços por meio da equação de Slutsky como segue:

$$\frac{\partial x_j(p,m)}{\partial p_i} = \underbrace{\frac{\partial h_j(p,v(p,m))}{\partial p_i}}_{\text{ES }(<0)} - \underbrace{\frac{\partial x_j(p,m)}{\partial m} x_i(p,m)}_{\text{ER }(\gtreqless 0)}. \tag{1.213}$$

Demonstração. Seja x^* a cesta que maximiza utilidade em (p^*, m^*) e seja $u^* = u(x^*)$. É verdade que

$$h_j(p^*, u^*) \equiv x_j(p, e(p, u^*)). \tag{1.214}$$

Podemos diferenciar essa expressão em relação a p_i e avaliar a derivada em p^* para obter:

$$\frac{\partial h_j(p^*,u^*)}{\partial p_i} = \frac{\partial x_j(p^*,m^*)}{\partial p_i} + \frac{\partial x_j(p^*,m^*)}{\partial m}\frac{\partial e(p^*,u^*)}{\partial p_i}. \tag{1.215}$$

Observe atentamente o significado dessa expressão. O lado esquerdo é como a demanda compensada muda quando p_i muda. O lado direito diz que essa mudança é igual à mudança na demanda mantendo o gasto fixo em m^* mais a mudança na demanda quando a renda muda vezes quanto a renda precisa mudar para manter a utilidade constante. Mas este último termo, $\frac{\partial e(p^*,u^*)}{\partial p_i}$, é x_i^*. Rearranjando, obtemos:

$$\frac{\partial x_j(p^*,m^*)}{\partial p_i} = \frac{\partial h_j(p^*,u^*)}{\partial p_i} - \frac{\partial x_j(p^*,m^*)}{\partial m} x_i^*, \tag{1.216}$$

que é a equação de Slutsky. ■

Há outras variações da equação de Slutsky, como a que está disponível em Varian (1994):

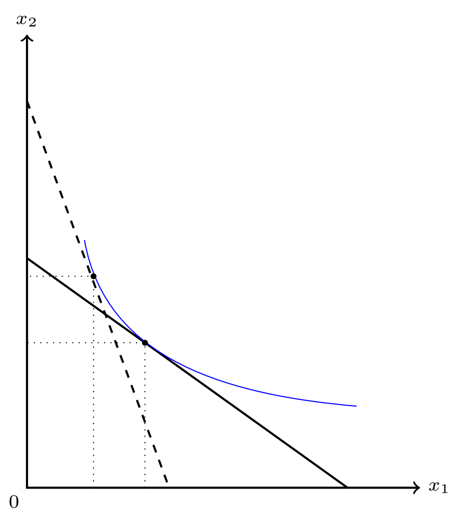
$$\Delta x_j \approx \frac{\partial x_j(p,m)}{\partial p_i}\Delta p_i = \frac{\partial h_j(p,u)}{\partial p_i}\Delta p_i - \frac{\partial x_j(p,m)}{\partial m} x_i^* \Delta p_i. \tag{1.217}$$

Lembre-se de que o objetivo do problema de minimização do dispêndio era gerar conceitos que pudéssemos usar para avaliar as mudanças no bem-estar. O objetivo da função dispêndio era nos dar uma maneira de medir o impacto de uma mudança de preços em termos de unidades monetárias. Enquanto a função dispêndio faz isso (você pode apenas olhar para $e(p^1, u) - e(p^0, u)$), ela sofre de outro problema. A função dispêndio é baseada na função de demanda Hicksiana, e a função de demanda Hicksiana leva como seus argumentos os preços e o nível de utilidade u. O problema é que, embora os preços sejam observáveis, os níveis de utilidade certamente não são. E, embora possamos gerar algumas informações perguntando às pessoas repetidas vezes como elas comparam certas cestas, essa não é uma maneira muito boa de fazer comparações de bem-estar. Para resumir nosso problema. As funções de demanda Marshallianas são baseadas em dados observáveis (p, w), mas não podem ser usadas para comparações de bem-estar. As funções de demanda Hicksianas, por outro lado, podem ser usadas para fazer comparações de bem-estar, mas são baseadas em dados não observáveis. Como mencionamos anteriormente, as restrições sobre as funções de demanda Hicksianas não são diretamente observáveis. No entanto, como indicado pela equação de Slutsky, podemos expressar as derivadas de h em relação a p como derivadas de x em relação a p e m, e estas são observáveis.

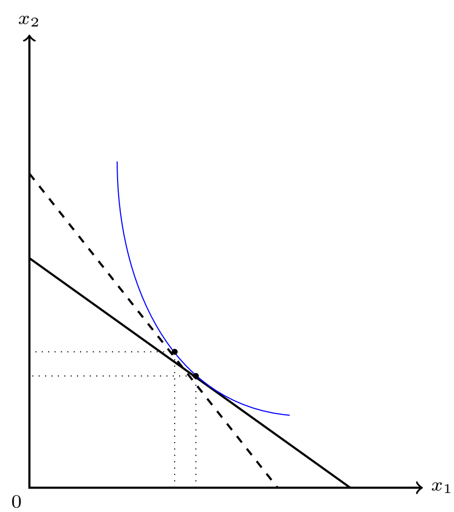
Vendo de outra forma, o efeito substituição, é a variação na demanda devido à variação da taxa à qual os dois bens são trocados, ou seja, a variação da demanda decorrente de uma mudança nos preços relativos. Neste caso, compensa-se a renda de modo a manter o poder de compra constante (à la Slutsky) ou a utilidade constante (à la Hicks). O efeito renda é a variação na demanda devido a uma mudança na renda real. Os preços relativos são mantidos fixos. Neste caso, retira-se a compensação na renda dada para se capturar o efeito substituição. O efeito substituição é sempre negativo, já que corresponde ao oposto da variação dos preços, mas o efeito total pode ser negativo ou positivo, a depender do efeito renda.

Obviamente, então, se as curvas de indiferença são relativamente planas, você tem que percorrer um longo caminho antes que a taxa marginal de substituição seja igual à nova relação de preços, e o efeito substituição é considerável. Se as curvas de indiferença são altamente convexas, a taxa marginal de substituição muda rapidamente e você não precisa ir longe: o efeito substituição é pequeno. Veja a Figura 1.50.

Figura 1.50 – Efeito substituição e curvatura da curva de utilidade



(a) Substancial



(b) Reduzido

Tabela 1.2 – Resumo

Efeito total	Efeito substituição	Efeito renda	Tipo de bem
-	-	+	Normal
-	-	-	Inferior, mas não Giffen
+	-	-	Giffen

Podemos derivar esse problema de forma gráfica também, como segue nas Figuras 1.51 e 1.52.

Figura 1.51 – Identificando os efeitos renda e substituição: aumento do preço do bem 1

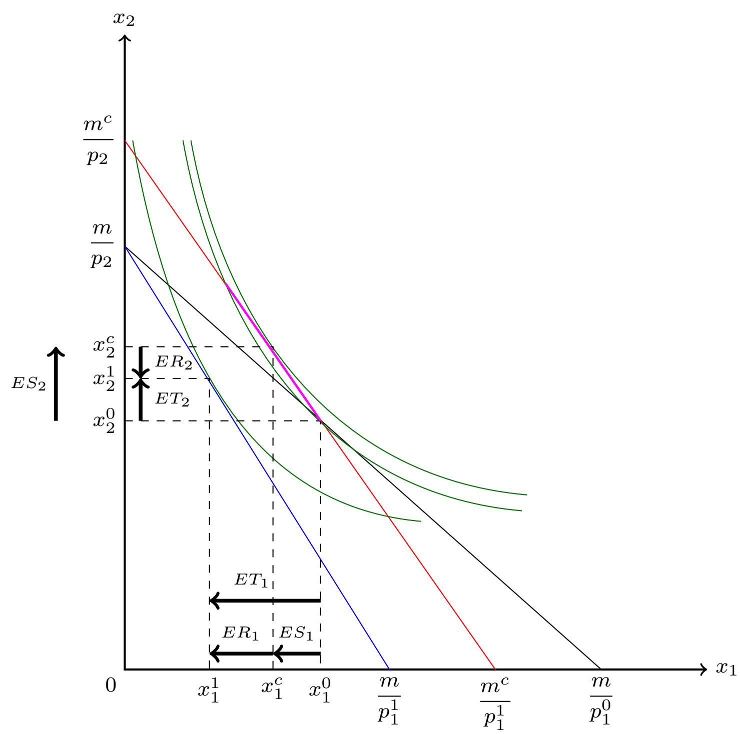
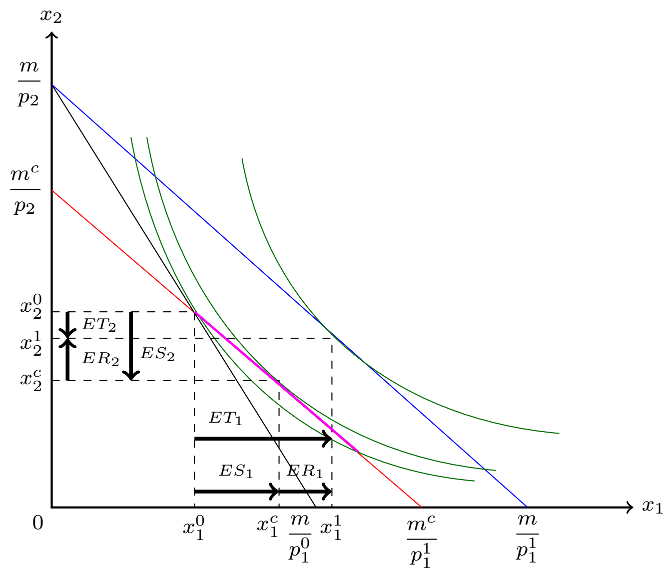


Figura 1.52 – Identificando os efeitos renda e substituição: redução do preço do bem 1



Note que a equação de Slutsky fornece a ligação entre as funções de demanda Marshallianas $x(p, m)$ e as funções de demanda Hicksianas, $h(p, u)$. Assim, se estimarmos o lado direito desta equação, que é uma função das variáveis observáveis p e m, então podemos derivar o valor do lado esquerdo da equação, mesmo que seja baseado em u, que não é observável. Lembre-se de que o que está implícito na ideia da função de demanda Hicksiana é a ideia de que a riqueza do consumidor seria ajustada para que ele pudesse obter a mesma utilidade depois de uma mudança de preço, como fazia antes. Essa ideia é aparente quando olhamos para a equação de Slutsky. Ele diz que a mudança na demanda quando a riqueza do consumidor é ajustada é composta de duas partes. O termo $\dfrac{\partial x_j(p, m)}{\partial p_i}$ é igual a quanto o consumidor mudaria a demanda se a riqueza fosse mantida constante. O termo $\dfrac{\partial x_j(p, m)}{\partial m} x_i(p, m)$ é a mudança adicional na demanda após a compensação na riqueza.

Além disso, a equação de Slutsky e a matriz semi-definida negativa nas funções de demanda Hicksianas nos dão o seguinte resultado nas funções de demanda Marshallianas:

Proposição 1.7.1. *A matriz de substituição* $\left(\dfrac{\partial x_j(p^*, m^*)}{\partial p_i} + \dfrac{\partial x_j(p^*, m^*)}{\partial m} x_i^*\right)$ *é simétrica e semi-definida negativa.*

A partir da equação de Slutsky podemos obter algumas relações interessantes:

- Seja a equação de Slutsky

$$\frac{\partial x_i^*(p, m)}{\partial p_i} = \frac{\partial h_i^*(p, v(p, m))}{\partial p_i} - \frac{\partial x_i^*(p, m)}{\partial m} x_i^*(p, m). \tag{1.218}$$

Multiplique ambos os lados por $\dfrac{p_i}{x_i^*(p, m)}$ e reescreva como:

$$\frac{\partial x_i^*(p,m)}{\partial p_i}\frac{p_i}{x_i^*(p,m)} = \frac{\partial h_i^*(p,v(p,m))}{\partial p_i}\frac{p_i}{x_i^*(p,m)} - \frac{\partial x_i^*(p,m)}{\partial m}x_i^*(p,m)\frac{p_i}{x_i^*(p,m)}\frac{m}{m}$$
$$\varepsilon_{x_i,p_i} = \varepsilon_{h_i,p_i} - \varepsilon_{x_i,m}\theta_{x_i}. \tag{1.219}$$

A elasticidade da demanda Marshalliana (não compensada) é igual à elasticidade da demanda Hicksiana (compensada) menos a elasticidade-renda da demanda multiplicada pela participação que o bem tem no orçamento.

- A equação de Slutsky (cruzada) é análoga:

$$\frac{\partial x_j^*(p,m)}{\partial p_i} = \frac{\partial h_j^*(p,v(p,m))}{\partial p_i} - \frac{\partial x_j^*(p,m)}{\partial m}x_j^*(p,m). \tag{1.220}$$

Intuitivamente, pensamos em dois bens como substitutos se a demanda por um aumenta quando o preço do outro aumenta. Assim, dizemos que o bem j é um substituto bruto para o bem i se $\frac{\partial x_j^*(p,m)}{\partial p_i} \geq 0$. Infelizmente, a definição de substitutos baseados na demanda Marshalliana (não compensada) não é muito útil, uma vez que não satisfaz a simetria: podemos ter $\frac{\partial x_j^*(p,m)}{\partial p_i} > 0$ e $\frac{\partial x_i^*(p,m)}{\partial p_j} < 0$. Existe uma definição melhor baseada na demanda Hicksiana (compensada). Nós dizemos que os bens j e i são substitutos líquidos se $\frac{\partial h_j^*(p,v(p,m))}{\partial p_i} \geq 0$ e complementares líquidos se $\frac{\partial h_j^*(p,v(p,m))}{\partial p_i} \leq 0$. Esta definição atende a simetria: $\frac{\partial h_j^*(p,v(p,m))}{\partial p_i} \geq 0$ se e somente se $\frac{\partial h_i^*(p,v(p,m))}{\partial p_j} \geq 0$.

Exemplo 1.7.9. *Seja a seguinte função de demanda Cobb-Douglas: $u(x_1,x_2) = x_1^\alpha x_2^{1-\alpha}$. Vamos verificar como é a equação de Slutsky nesse exemplo. Para tanto, sabemos que:*

$$v(p_1,p_2,m) = mp_1^{-\alpha}p_2^{\alpha-1} \tag{1.221}$$
$$e(p_1,p_2,u) = up_1^\alpha p_2^{1-\alpha} \tag{1.222}$$
$$x_1(p_1,p_2,m) = \frac{\alpha m}{p_1} \tag{1.223}$$
$$h_1(p_1,p_2,u) = \alpha p_1^{\alpha-1}p_2^{1-\alpha}u \tag{1.224}$$

Logo,

$$\frac{\partial x_1(p,m)}{\partial p_1} = -\frac{\alpha m}{p_1^2} \tag{1.225}$$
$$\frac{\partial x_1(p,m)}{\partial m} = \frac{\alpha}{p_1} \tag{1.226}$$

$$\frac{\partial h_1(p,u)}{\partial p_1} = \alpha(\alpha-1)p_1^{\alpha-2}p_2^{1-\alpha}u \tag{1.227}$$

$$\begin{aligned}\frac{\partial h_1(p,v(p,m))}{\partial p_1} &= \alpha(\alpha-1)p_1^{\alpha-2}p_2^{1-\alpha}mp_1^{-\alpha}p_2^{\alpha-1}\\ &= \alpha(\alpha-1)p_1^{-2}m\end{aligned} \tag{1.228}$$

Substituindo estes resultados na equação de Slutsky, obtemos:

$$\begin{aligned}\frac{\partial h_1}{\partial p_1} - \frac{\partial x_1}{\partial m}x_1 &= \frac{\alpha(\alpha-1)m}{p_1^2} - \frac{\alpha}{p_1} \times \frac{\alpha m}{p_1}\\ &= -\frac{\alpha m}{p_1^2}\\ &= \frac{\partial x_1}{\partial p_1}\end{aligned} \tag{1.229}$$

Exemplo 1.7.10. *Seja a seguinte função de quase-linear: $u(x_1,x_2) = x_1 + \ln x_2$. Vamos verificar como é a equação de Slutsky nesse exemplo. Para tanto, sabemos que:*

$$v(p_1,p_2,m) = \frac{m}{p_1} + \ln\left(\frac{p_1}{p_2}\right) - 1 \tag{1.230}$$

$$e(p_1,p_2,u) = p_1\left[u - \ln\left(\frac{p_1}{p_2}\right) + 1\right] \tag{1.231}$$

$$x_1(p_1,p_2,m) = \frac{m}{p_1} - 1 \tag{1.232}$$

$$x_2(p_1,p_2,m) = \frac{p_1}{p_2} \tag{1.233}$$

$$h_1(p_1,p_2,u) = u - \ln\left(\frac{p_1}{p_2}\right) \tag{1.234}$$

$$h_2(p_1,p_2,u) = \frac{p_1}{p_2} \tag{1.235}$$

Substituindo estes resultados na equação de Slutsky, obtemos:

$$\begin{aligned}\frac{\partial h_1}{\partial p_1} - \frac{\partial x_1}{\partial m}x_1 &= \underbrace{-\frac{1}{p_1}}_{ES} \underbrace{- \frac{1}{p_1} \times \left(\frac{m}{p_1} - 1\right)}_{ER}\\ &= -\frac{m}{p_1^2} < 0\end{aligned} \tag{1.236}$$

Observe que o efeito substituição é negativo, o efeito renda é positivo e o efeito total é negativo.

Fazendo o mesmo para o bem 2, temos:

$$\begin{aligned}\frac{\partial h_2}{\partial p_2} - \frac{\partial x_2}{\partial m}x_2 &= \underbrace{-\frac{p_1}{p_2^2}}_{ES} \underbrace{-0 \times \frac{p_1}{p_2}}_{ER} \\ &= -\frac{p_1}{p_2^2} < 0 \end{aligned} \tag{1.237}$$

Observe que o efeito substituição é negativo, o efeito renda é nulo e o efeito total é negativo.

Exemplo 1.7.11. *Suponha que aos preços* $(p_1, p_2) = (5, 10)$*, um consumidor racional, com renda de \$ 100, adquire a cesta* $(x_1, x_2) = (6, 7)$*. Suponha que um economista tenha estimado as seguintes derivadas:*

$$\frac{\partial h_1(5, 10, V(5, 10, 100))}{\partial p_1} = -2 \tag{1.238}$$

$$\frac{\partial h_1(5, 10, V(5, 10, 100))}{\partial p_2} = 1 \tag{1.239}$$

$$\frac{\partial x_1(5, 10, 100)}{\partial m} = \frac{2}{7} \tag{1.240}$$

Estime a cesta que o consumidor escolheria se ele se defrontasse com os preços $(p_1, p_2) = (5, 11)$*.*

O preço do bem 1 *e a renda não mudaram. Portanto, nós precisamos estimar o impacto de uma mudança (pequena) no preço do bem* 2 *sobre a demanda dos dois bens.*

A mudança na demanda do bem 1*,* Δx_1*, é dada por:*

$$\begin{aligned}\Delta x_1 &= \frac{\partial x_1(5, 10, 100)}{\partial p_2}\Delta p_2 \\ &= \frac{\partial x_1(5, 10, 100)}{\partial p_2} \\ &= \frac{\partial h_1(5, 10, V(5, 10, 100))}{\partial p_2} - x_2\frac{\partial x_1(5, 10, 100)}{\partial m} \\ &= 1 - 7\left(\frac{2}{7}\right) = -1 \end{aligned} \tag{1.241}$$

Portanto, o novo consumo do bem 1 *será* 5*. Para estimar o novo consumo do bem* 2*, usamos a restrição orçamentária:*

$$x_2 = \frac{100}{11} - \frac{25}{11} = \frac{75}{11} \tag{1.242}$$

1.7.4 Implicações da Homogeneidade e da Lei de Walras nos Efeitos Renda e Preço

A homogeneidade de grau zero da demanda marshalliana implica que

$$x(\lambda p, \lambda m) = x(p, m) \quad \therefore \quad x(\lambda p, \lambda m) - x(p, m) = 0. \tag{1.243}$$

A partir desta relação, podemos tomar a derivada total da equação acima, e assim, temos a seguinte relação de Euler:

$$\begin{aligned}
\alpha \frac{\partial x_1(p,m)}{\partial p_1} p_1 + \alpha \frac{\partial x_1(p,m)}{\partial p_2} p_2 + \ldots + \alpha \frac{\partial x_1(p,m)}{\partial p_L} p_L + \\
\alpha \frac{\partial x_2(p,m)}{\partial p_1} p_1 + \alpha \frac{\partial x_2(p,m)}{\partial p_2} p_2 + \ldots + \alpha \frac{\partial x_2(p,m)}{\partial p_L} p_L + \\
\alpha \frac{\partial x_L(p,m)}{\partial p_1} p_1 + \alpha \frac{\partial x_L(p,m)}{\partial p_2} p_2 + \ldots + \alpha \frac{\partial x_L(p,m)}{\partial p_L} p_L + \\
\alpha \frac{\partial x_1(p,m)}{\partial w} w + \alpha \frac{\partial x_2(p,m)}{\partial w} w + \ldots + \alpha \frac{\partial x_L(p,m)}{\partial w} w + \\
- \frac{\partial x_1(p,m)}{\partial p_1} p_1 - \frac{\partial x_1(p,m)}{\partial p_2} p_2 - \ldots - \frac{\partial x_1(p,m)}{\partial p_L} p_L + \\
- \frac{\partial x_2(p,m)}{\partial p_1} p_1 - \frac{\partial x_2(p,m)}{\partial p_2} p_2 - \ldots - \frac{\partial x_2(p,m)}{\partial p_L} p_L + \\
- \frac{\partial x_L(p,m)}{\partial p_1} p_1 - \frac{\partial x_L(p,m)}{\partial p_2} p_2 - \ldots - \frac{\partial x_L(p,m)}{\partial p_L} p_L + \\
- \frac{\partial x_1(p,m)}{\partial w} w - \frac{\partial x_2(p,m)}{\partial w} w - \ldots - \frac{\partial x_L(p,m)}{\partial w} &= 0 \\
\alpha \sum_{k=1}^{L} \frac{\partial x_\ell(p,m)}{\partial p_k} p_k + \alpha \sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial m} m - \sum_{k=1}^{L} \frac{\partial x_\ell(p,m)}{\partial p_k} p_k - \sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial m} m &= 0 \\
(\alpha - 1) \left(\sum_{k=1}^{L} \frac{\partial x_\ell(p,m)}{\partial p_k} p_k \right) + (\alpha - 1) \left(\sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial m} m \right) &= 0 \\
(\alpha - 1) \left(\sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial p_k} p_k + \sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial m} m \right) &= 0 \\
\left(\sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial p_k} p_k + \sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial m} m \right) &= 0 \\
\left(\sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial p_k} \frac{p_k}{x_\ell(p,m)} + \sum_{\ell=1}^{L} \frac{\partial x_\ell(p,m)}{\partial m} \frac{m}{x_\ell(p,m)} \right) &= 0 \\
\sum_{\ell=1}^{L} \eta_{\ell k}(p,m) + \eta_{\ell m}(p,m) &= 0.
\end{aligned} \tag{1.244}$$

Ou seja, uma mesma mudança em todos os preços e renda leva a nenhuma mudança na demanda.

Da exaustão da lei de Walras podemos ver que:

$$\begin{aligned}
\frac{\partial w}{\partial w} &= \sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p,m)}{\partial m} \\
1 &= \sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p,m)}{\partial m} \frac{x_\ell}{m} \frac{m}{x_\ell} \\
1 &= \sum_{\ell=1}^{L} \frac{p_\ell x_\ell}{m} \frac{\partial x_\ell(p,m)}{\partial m} \frac{m}{x_\ell} \\
1 &= \sum_{\ell=1}^{L} \theta_\ell \eta_{\ell m}. \qquad (1.245)
\end{aligned}$$

Esta relação, conhecida como propriedade de agregação de Engel, implica que a despesa total deve mudar na mesma proporção que mudanças na renda.

Ademais, temos a seguinte relação:

$$\begin{aligned}
\frac{\partial w}{\partial p_k} &= \sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p,m)}{\partial p_k} + x_k(p,m) \\
-x_k(p,m) &= \sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p,m)}{\partial p_k} \\
-\frac{p_k x_k(p,m)}{m} &= \sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p,m)}{\partial p_k} \frac{p_k}{m} \\
-\theta_k &= \sum_{\ell=1}^{L} \theta_\ell \eta_{\ell k} \\
\sum_{\ell=1}^{L} \theta_\ell \eta_{\ell k} + \theta_k &= 0. \qquad (1.246)
\end{aligned}$$

Esta relação, conhecida como propriedade de Cournot, implica que a despesa total não muda quando os preços mudam.

1.8 As Curvas de Demanda Marshalliana e de Demanda Hicksiana

Já sabemos que a curva de demanda Marshalliana descreve a relação entre os valores assumidos pela função de demanda de um consumidor e o preço de um bem quando a renda do consumidor e os preços dos demais bens são mantidos constantes. De modo similar, chamamos de curva de demanda compensada ou Hicksiana o gráfico da relação entre o preço de um bem e a função de demanda compensada de um consumidor por esse bem quando os preços dos outros bens são mantidos constantes e a renda do consumidor é constantemente reajustada para garantir que seu nível de utilidade permaneça inalterado. Assim, uma curva de demanda Marshalliana do bem 1, por exemplo, é dada pelo gráfico da função $x_1(p_1, p_2, m)$ quando se supõe p_2 e m constantes e uma curva de demanda Hicksiana do mesmo bem é dada pelo gráfico da função $h_1(p_1, p_2, u)$ quando se supõe p_2 e u constantes.

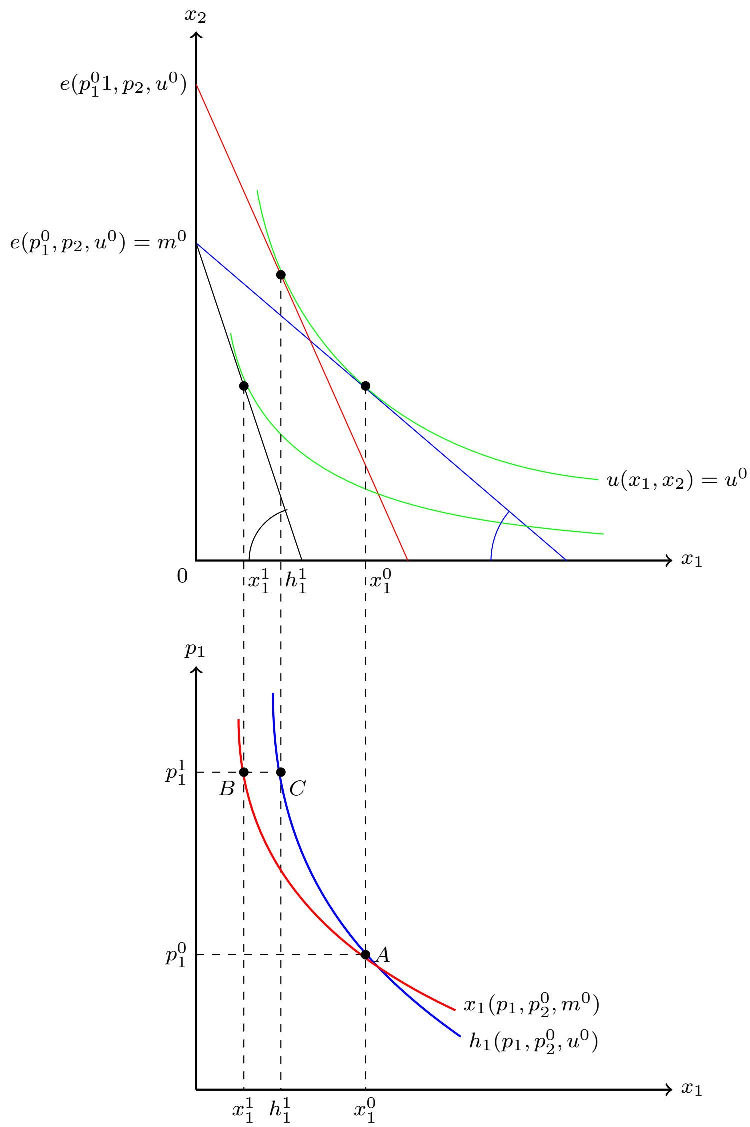
Na Figura 1.53 ilustramos a diferença na obtenção dessas duas curvas. O gráfico na parte de cima dessa figura mostra uma mapa de curvas de indiferença e algumas linhas de restrição orçamentária. Num primeiro momento o consumidor se defronta com a linha de restrição orçamentária em azul, sendo o preço do bem 1 p_1^0, a renda do consumidor m^0 e o preço do bem 2 $p_2^0 = 1$. Ao maximizar sua função de utilidade, o consumidor adquire a quantidade x_1^0 do bem 1, obtendo o nível de utilidade u^0. Nessas condições, o ponto A no gráfico abaixo, com coordenadas x_1^0 e p_1^0, é tanto um ponto da curva de demanda Marshalliana associada a $p_2 = p_2^0$ e $m = m^0$, pois x_1^0 é a quantidade do bem 1 que maximiza a utilidade do consumidor quando $p_1 = p_1^0$ e a renda do consumidor é m^0, quanto um ponto sobre a curva de demanda compensada associada ao nível de utilidade u^0, pois x_1^0 é a quantidade demandada do bem 1 quando o preço é p_1^0 e sua renda é ajustada para garantir que ele obtenha o nível de utilidade u^0.

Quando o preço do bem 1 sobe para p_1^1 e a renda do consumidor assim como o preço do bem 2 permanecem inalterados, a linha de restrição orçamentária de nosso consumidor se desloca para a linha em preto e o consumidor obtém um novo equilíbrio consumindo uma quantidade x_1^1 do bem 1. No gráfico abaixo, ponto B, cujas coordenadas são x_1^1 e p_1^1, representa, portanto, um ponto sobre a curva de demanda Marshalliana do bem 1 associada a $m = m^0$ e $p_2 = p_2^0$. Esta curva, representada pela curva em vermelho, deve, portanto, passar pelos pontos B e A. Porém, o ponto B não é um ponto da curva de demanda compensada associada a $u = u^0$ (e $p_2 = p_2^0$), pois, o equilíbrio obtido sobre a linha de restrição orçamentária em preto não se dá sobre a curva de indiferença original, ou seja, nesse equilíbrio, o consumidor não obtém o nível de utilidade original u^0. Para que ele possa obter esse nível de utilidade ao novo preço, é necessário que sua renda seja reajustada para $e(p_1^1, p_2^0, m^0)$, de tal sorte que a linha de restrição orçamentária volte a tangenciar a curva de indiferença original (linha de restrição orçamentária em vermelho). Caso isso ocorra, o consumidor irá demandar, após a compensação em sua renda, a quantidade h_1^1 do bem 1. Desse modo, o ponto C ou (h_1^1, p_1^1) no gráfico abaixo representa um ponto sobre a curva de demanda compensada de nosso consumidor. Essa curva, em azul, deve passar portanto, pelos pontos A e C. Por que C está à direita de B? Porque o bem em questão é um bem normal, de tal sorte que o aumento na renda decorrente da compensação pela elevação no preço do bem 1 deve provocar um aumento na demanda pelo bem 1.

Se, ao contrário de uma elevação no preço do bem, ocorresse uma redução no preço desse bem, então, a variação na quantidade demandada seria maior na curva de demanda Marshalliana do que sobre a curva de demanda Hicksiana, pois, nesse caso, a compensação na renda necessária para manter a utilidade do consumidor constante seria negativa, de tal sorte que, em se tratando de um bem normal, a demanda desse bem após a compensação seria inferior à demanda verificada antes da compensação.

Concluímos, portanto, que no caso de um bem normal, para preços superiores ao correspon-

Figura 1.53 – A construção das curvas de demanda Marshalliana e Hicksiana



dente ao cruzamento de uma curva de demanda compensada com uma curva de demanda Marshalliana, a curva de demanda compensada deve ficar à direita da curva de demanda Marshalliana e, para preços inferiores ao preço de cruzamento dessas curvas, a curva de demanda compensada deve ficar à esquerda da curva de demanda Marshalliana. Em outras palavras, a curva de demanda compensada deve ter inclinação mais acentuada que a curva de demanda Marshalliana.

1.9 Variação Compensatória e Variação Equivalente

Como podemos medir a variação no bem estar de um consumidor decorrente de uma mudança nos parâmetros que definem a posição de sua linha de restrição orçamentária? Como sabemos, se considerarmos que a variação na função de utilidade do consumidor é a medida procurada, há várias medidas possíveis, visto que as preferências do consumidor podem ser representadas por incontáveis funções de utilidade alternativas. Mas nós não queremos uma medida de variação de bem estar qualquer. Queremos uma medida que seja conveniente para propósitos práticos. Se pesquisarmos a literatura acerca de medidas de variação no bem-estar do consumidor, verificaremos que, usualmente, duas medidas são consideradas convenientes. A primeira delas é a chamada variação compensatória e, a segunda, variação equivalente. Mais adiante, apresentaremos uma definição técnica dessas duas medidas, mas, por ora, contentar-nos-emos com definições conceituais.

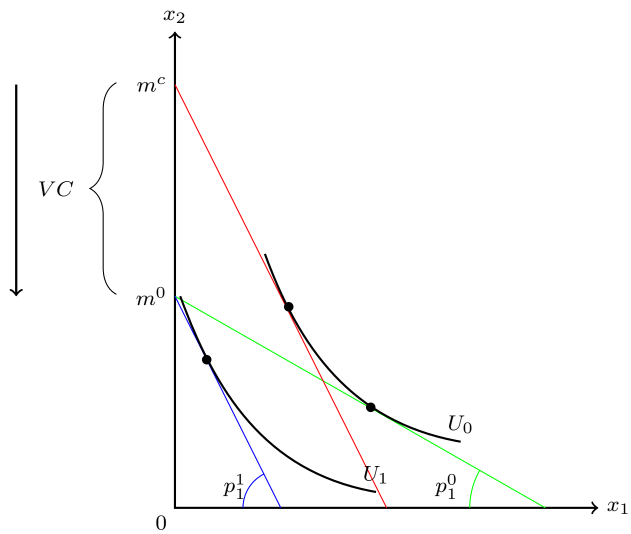
Definição 1.9.1 (Variação Compensatória)**.** *Suponha que um consumidor experimente uma mudança nos preços e em sua renda de* (p_1^0, p_2^0, m^0) *para* (p_1^1, p_2^1, m^1)*. A variação compensatória (VC) associada a essa mudança é definida como o negativo da variação na renda* m^1 *necessária para fazer com que, após essa variação, o consumidor, defrontando-se com os preços* p_1^1 *e* p_2^1*, obtenha ao maximizar sua utilidade exatamente o mesmo nível de utilidade que obteria caso sua renda fosse* m^0 *e os preços dos dois bens fossem* p_1^0 *e* p_2^0*. Em outras palavras, a variação compensatória é o valor que faz com que o nível de utilidade máxima do consumidor quando os preços são* p_1^1 *e* p_2^1 *e sua renda é* $m^1 - VC$ *seja igual ao nível de utilidade máximo atingido quando os preços são* p_1^0 *e* p_2^0 *e sua renda é* m^0*.*

Definição 1.9.2 (Variação Equivalente)**.** *Suponha que um consumidor experimente uma mudança nos preços e em sua renda de* (p_1^0, p_2^0, m^0) *para* (p_1^1, p_2^1, m^1)*. A variação equivalente (VE) associada a essa mudança é definida como a variação na renda* m^0 *que faria que, após essa variação, o consumidor, defrontando-se com os preços* p_1^0 *e* p_2^0*, obtenha ao maximizar sua utilidade exatamente o mesmo nível de utilidade que obteria caso sua renda fosse* m^1 *e os preços dos dois bens fossem* p_1^1 *e* p_2^1*. Em outras palavras, a variação equivalente é o valor que faz com que o nível de utilidade máxima do consumidor quando os preços são* p_1^0 *e* p_2^0 *e sua renda é* $m_0 + VE$ *seja igual ao nível de utilidade máximo atingido quando os preços são* p_1^1 *e* p_2^1 *e sua renda é* m^1*.*

A variação compensatória tem tal nome porque ela tem o mesmo valor absoluto da variação na renda necessária para compensar qualquer variação no bem estar do consumidor provocada pela mudança nos parâmetros de sua linha de restrição orçamentária. Já o nome “variação equivalente” advém do fato de que a variação equivalente é a variação de renda que faz com que, aos preços e renda iniciais, haja uma variação de bem-estar equivalente à gerada pela alteração para os preços e renda finais.

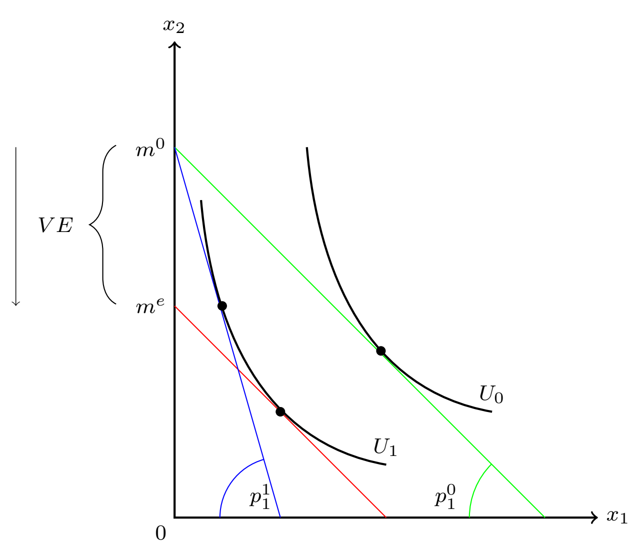
A Figura 1.54 ilustra o conceito de variação compensatória tal como ele se aplica a um aumento no preço do bem 1 quando a renda do consumidor e o preço do bem 2 são mantidos constantes em m^0 e 1, respectivamente. Na Figura, a linha em verde é a linha de restrição orçamentária inicial, quando o preço do bem 1 é p_1^0. Após uma elevação nesse preço para p_1^1, a linha de restrição orçamentária passaria a ser a linha em azul. Para devolver o consumidor ao nível de bem estar original seria necessário um aumento de renda que deslocasse essa linha de restrição orçamentária até a linha em vermelho, quando ela voltaria a tangenciar a curva de indiferença inicial. Com esse aumento, a renda do consumidor assumiria o valor m^c. A variação compensatória é o negativo desse aumento de renda e é dada pela diferença $m^0 - m^c$.

Figura 1.54 – Variação compensatória e curvas de indiferença: aumento em p_1



A Figura 1.55 mostra de modo similar a interpretação gráfica do conceito de variação equivalente. Na Figura, a linha de restrição orçamentária inicial é a linha em verde. Após uma elevação no preço do bem 1 de p_1^0 para p_1^1, o consumidor se defronta com uma nova linha de restrição orçamentária (em azul) e se vê forçado a consumir sobre uma curva de indiferença mais baixa que a inicial. A variação equivalente é dada pela variação na renda do consumidor que faria com que, ao preço inicial p_1^0, o consumidor passasse a consumir sobre a mesma curva de indiferença da situação final. Em outras palavras, ela é dada pela diferença entre a renda m^e, que faria com que, ao preço inicial p_1^0, o consumidor obtivesse o mesmo nível de utilidade que obtém na situação final (com renda m^0 e preço p_1^1), e a renda da situação final, m^0.

Figura 1.55 – Variação equivalente e curvas de indiferença: aumento em p_1



É interessante comparar a representação gráfica dos conceitos de variação compensatória e variação equivalente para o caso em que há um aumento no preço do bem 1 com a representação gráfica desses mesmos conceitos no caso em que há uma redução no preço do bem 1. Fazemos isso nas Figuras 1.56 e 1.57. Estas diferem das Figuras 1.54 e 1.55 por trocarem de posição o preço inicial p_1^0 com o preço final p_1^1, de tal sorte que, aquilo que representa um aumento de preço nas Figuras 1.54 e 1.55 representa, nas Figuras 1.56 e 1.57 uma redução de preços em igual magnitude.

Figura 1.56 – Variação compensatória e curvas de indiferença: redução em p_1

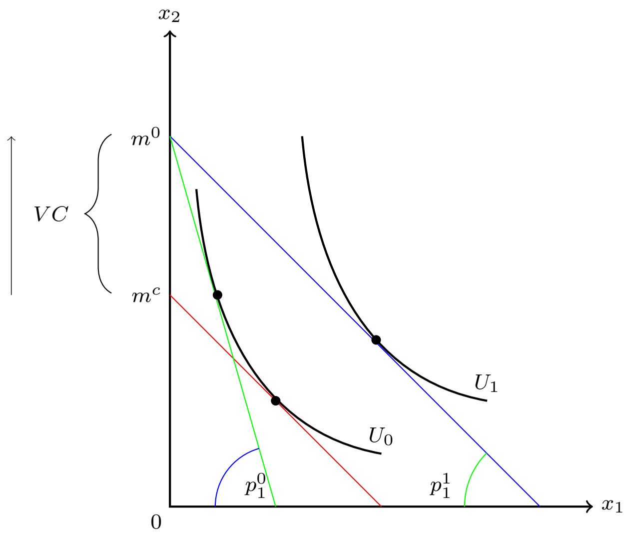
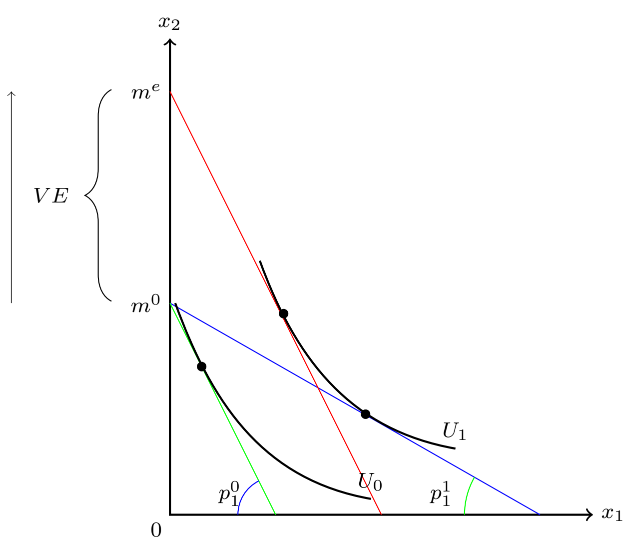


Figura 1.57 – Variação equivalente e curvas de indiferença: redução em p_1



Na Figura 1.56, a linha de restrição orçamentária inicial, quando o preço do bem 1 é p_1^0 é a linha em verde. Com a redução desse preço para p_1^1, a linha de restrição orçamentária passa a ser a linha em verde, de tal forma que o consumidor, que inicialmente consumia uma cesta de bens sobre a curva de indiferença U_0, obtém um novo equilíbrio na curva de indiferença mais elevada $U_1 1$. A variação compensatória na renda associada a essa redução no preço do bem 1 é a redução na renda que faria com que, ao preço final p_1^1, o consumidor voltasse a consumir sobre a curva de indiferença original U_0. Se m^c é a renda necessária para fazer com que a linha de restrição orçamentária, ao preço p_1^1, volte a tangenciar a curva de indiferença inicial U_0, então, a variação compensatória será a diferença entre a renda efetiva do consumidor após a redução no preço, m^0 e a renda m^c.

Na Figura 1.57 vemos que, após uma redução no preço do bem 1 de p_1^0 para p_1^1, o consumidor passa de um equilíbrio sobre a curva de indiferença U_0 para a curva de indiferença mais elevada U_1. Caso não ocorresse essa redução no preço seria necessário uma elevação na renda do consumidor de m^0 para m^e para fazer com que a linha de restrição orçamentária (destacada em vermelho) tangenciasse a curva de indiferença U_1. Assim, a variação equivalente é $VE = m^e - m^0$.

Se você comparar as Figuras, verá que em nosso exemplo, para a mesma elevação no preço do bem 1 a variação compensatória é em módulo superior à variação equivalente. Como essas duas variações têm sinal negativo, isso significa que a variação compensatória é menor que a variação equivalente. Quando há uma redução no preço do bem 1, a variação compensatória é inferior à variação equivalente. Por que isso acontece? No nosso exemplo, isso acontece porque as duas curvas de indiferença estão tão mais distantes entre si quanto mais à direita medimos essa distância. Isso ocorre caso o bem em questão seja um bem normal. Desse modo, podemos concluir que, no caso de um bem normal, a variação equivalente associada a uma variação no preço desse bem é maior que a variação compensatória associada a essa mudança.

Exemplo 1.9.1. *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,* $m = 1000$ *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

$$x_1(p_1, 1, 100) = \frac{100}{p} \tag{1.247}$$

A variação no excedente do consumidor é encontrada integrando-se essa função entre $p_1 = p_1^0 = 8$ *e* $p_1 = p_1^1 = 10$*:*

$$CS = \int_8^{10} \frac{100}{p} dp = 100(\ln 10 - \ln 8) \approx 22,31 \tag{1.248}$$

Para calcularmos a variação equivalente e a variação compensatória, primeiramente determinamos os níveis de utilidade inicial u^0 *e final* u^1*:*

$$u^0 = a^a(1-a)^{1-a}\frac{m}{(p_1^0)^a(p_2^0)^{1-a}} = \left(\frac{1}{10}\right)^{\frac{1}{10}}\left(\frac{9}{10}\right)^{\frac{9}{10}}\frac{1000}{10^{\frac{1}{10}}1^{\frac{9}{10}}} \approx 573,88 \tag{1.249}$$

$$u^1 = a^a(1-a)^{1-a}\frac{m}{(p_1^1)^a(p_2^1)^{1-a}} = \left(\frac{1}{10}\right)^{\frac{1}{10}}\left(\frac{9}{10}\right)^{\frac{9}{10}}\frac{1000}{8^{\frac{1}{10}}1^{\frac{9}{10}}} \approx 586,83 \tag{1.250}$$

A variação compensatória e a variação equivalente são, respectivamente:

$$CV = m - e(p_1^1, p_2, u^0) = 1000 - 573,88\frac{8^{\frac{1}{10}}1^{\frac{9}{10}}}{\left(\frac{1}{10}\right)^{\frac{1}{10}}\left(\frac{9}{10}\right)^{\frac{9}{10}}} \approx 22,07 \tag{1.251}$$

$$EV = e(p_1^0, p_2, u^1) - m = 586,83\frac{10^{\frac{1}{10}}1^{\frac{9}{10}}}{\left(\frac{1}{10}\right)^{\frac{1}{10}}\left(\frac{9}{10}\right)^{\frac{9}{10}}} - 1000 \approx 22,57 \tag{1.252}$$

Nesse exemplo, a variação no excedente do consumidor difere pouco - aproximadamente 1% - tanto da variação compensatória quanto da variação equivalente.

Tanto a variação equivalente quanto a variação compensatória fornecem medidas em unidades monetárias do impacto de uma mudança de preço no bem-estar do consumidor, e há circunstâncias em que cada uma é a medida apropriada a ser usada. Por exemplo, considere o preço inicial p^0 e dois vetores de preços alternativos p^a e p^b. As grandezas $VE(p^0, p^a, m)$ e $VE(p^0, p^b, m)$ são medidas em termos de riqueza a preços p^0 e, portanto, podem ser comparadas. Por outro lado, a variação compensatória, em diferentes níveis de preço, não podem ser facilmente comparados. Essa distinção é importante em questões de política, como decidir qual mercadoria será tributada. O impacto de colocar um imposto sobre a gasolina versus o impacto da aplicação de um imposto sobre a eletricidade precisa ser medido com relação ao mesmo preço de referência, se quisermos comparar os dois de maneira significativa. Isso significa usar a variação equivalente.

1.10 Efeito Renda e o Problema de Path Dependence

Esta exposição é 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 um imposto, por exemplo). O excedente do consumidor (Marshalliano) é definido como:

$$CS = \int_{p^0}^{p^1} x(p, m)dp. \tag{1.253}$$

Qual o problema com esta definição? O excedente do consumidor é dependente do caminho quando mais de um preço muda: de p^0 para p' e de p' para p^1. Ou seja,

$$CS(p^0 \to p') + CS(p' \to p^1) \neq CS(p^0 \to p^1). \tag{1.254}$$

Agora considere o caso de dois bens com a incidência de dois impostos (um em cada bem). Então, temos:

$$CS = \int_{p_1^0}^{p_1^1} x_1(p_1, p_2^0, m)dp_1 + \int_{p_2^0}^{p_2^1} x_2(p_1^1, p_2, m)dp_2, \tag{1.255}$$

ou

$$CS = \int_{p_2^0}^{p_2^1} x_2(p_1^0, p_2, m)dp_2 + \int_{p_1^0}^{p_1^1} x_1(p_1, p_2^1, m)dp_1. \tag{1.256}$$

Problema matemático: para estas duas definições serem equivalentes (path independence), as derivadas parciais cruzadas devem ser iguais, isto é, $\dfrac{dx_2}{dp_1} = \dfrac{dx_1}{dp_2}$. 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 Marshallianas 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 utililidade $h(p, u)$. 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:

$$e(p, U) = \min px \tag{1.257}$$

$$\text{sujeito a} \quad u(x) \geq U. \tag{1.258}$$

Seja μ o multiplicador na restrição de utilidade, então as condições de primeira ordem dadas por $p_i = \mu u_{x_i}$ e a restrição geram funções de demanda Hicksianas (ou compensadas), h, que mapeiam preços e utilidade na demanda

$$x_i = h_i(p, u). \tag{1.259}$$

Defina a perda para o consumidor de aumentar as taxas de imposto como

$$e(p^1, u) - e(p^0, u). \tag{1.260}$$

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 m. Temos que:

- $u^0 = v(p^0, m)$, utilidade inicial
- $u^1 = v(p^1, m)$, utilidade ao novo preço p^1

Usando estas, nós definimos a variação equivalente e a variação compensatória.

Variação compensatória:

$$CV = e(p^1, u^0) - e(p^0, u^0) = m - e(p^1, u^0). \tag{1.261}$$

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.

Variação equivalente:

$$EV = e(p^1, u^1) - e(p^0, u^1) = e(p^0, u^1) - m. \tag{1.262}$$

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.

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.

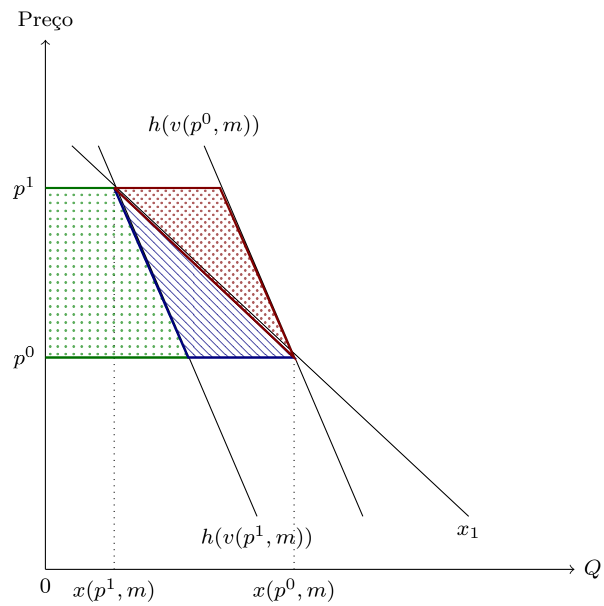
Sejam as funções de demanda hicksianas. Primeiro note que o teorema do envelope implica que $h_i = e_{p_i}(p,u)$. Por isso, pode definir CV ou EV como:

$$e(p^1,u) - e(p^0,u) = \int_{p^0}^{p^1} h(p,u)dp. \tag{1.263}$$

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 (Marshalliano) 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 1.58 – Excedente do consumidor marshalliano e tributação



Note que $h(p,v(p,m)) = x(p,m)$ por causa da dualidade (a solução para o problema max deve coincidir com a solução para o problema min de gasto no mesmo nível de utilidade indireta). Portanto, os correspondentes hicksianos de CV e EV devem cruzar os Marshallianos nos dois preços ($CV : p^0$ e $EV : p^1$). Observe que $h(p,u)$ tem uma inclinação mais acentuada que $x(p,m)$. 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:

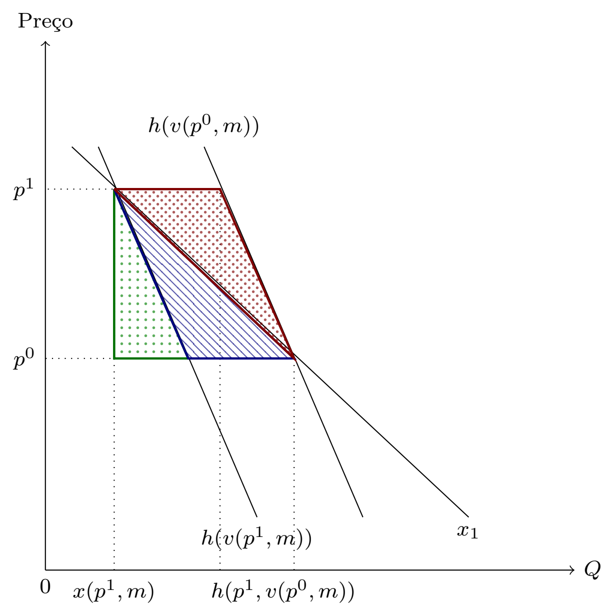
$$CS(u^1) = EV - (p^1 - p^0)h(p^1, u^1) \quad \text{[Mohring 1971]} \tag{1.264}$$
$$CS(u^0) = CV - (p^1 - p^0)h(p^1, u^0) \quad \text{[Diamond e McFadden 1974]} \tag{1.265}$$

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 quasilinear), então $CV = EV$ e há uma definição única de excedente do consumidor e peso morto.

Figura 1.59 – 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 demandaMarshalliana, 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

$$CV = CS = EV. \tag{1.266}$$

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

$$CV < CS < EV. \tag{1.267}$$

3. Caso, o bem seja um bem inferior, ou seja, caso sua demanda seja decrescente em relação à renda do consumidor, então teremos

$$CV > CS > EV. \tag{1.268}$$

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 $CS \approx CV \approx EV$.

1.11 Preferência Revelada

Até agora, começamos com preferência e, em seguida, descrevemos o comportamento. Mas na vida real, as preferências não são diretamente observáveis. A preferência revelada é uma forma de "trabalhar de trás para frente" - comece com o comportamento e descreva as preferências. Se queremos "recuperar" as preferências do comportamento das pessoas, temos que assumir que as preferências não mudam com o tempo. Nesta seção, também assumimos que as preferências são estritamente convexas - obtemos uma cesta única (não é necessário para a teoria das preferências reveladas, mas simplifica a exposição).

Os axiomas de preferência básica às vezes são criticados por serem muito fortes, alegando que é improvável que os indivíduos façam escolhas por meio do uso consciente de uma relação de preferência. Uma resposta a essa crítica é desenvolver uma teoria alternativa com base em um conjunto mais fraco de hipóteses. Uma das teorias alternativas mais interessantes é a da preferência revelada, que é discutida nesta seção. O princípio básico da teoria da preferência revelada é que as preferência devem ser construídas apenas a partir de decisões observáveis, isto é, a partir da escolha real feita por um consumidor. Uma relação de preferência individual, mesmo que exista, nunca pode ser diretamente observada no mercado. O melhor que podemos esperar na prática é uma lista das escolhas feitas em diferentes circunstâncias. Como podemos saber se esses dados foram gerados por um consumidor que maximiza a utilidade? A teoria da preferência revelada concentra-se nas escolhas feitas por um consumidor, não em uma relação de preferência subjacente.

Seja x^0 a cesta escolhida quando os preços eram p^0 e x^1 a cesta escolhida quando os preços eram p^1. Defina a relação binária "revelada preferível a" $\mathcal{R}$ de forma que $x^0 \mathcal{R} x^1$ se e somente se $p^0 x^0 \geq p^0 x^1$.

Definição 1.11.1 (Axioma Fraco da Preferência Revelada (AFrPR))**.** *Se $x^0 \mathcal{R} x^1$, então não é possível que $x^1 \mathcal{R} x^0$. Uma forma equivalente:*

$$p^0 x^0 \geq p^0 x^1 \quad \Longrightarrow \quad p^1 x^0 \geq p^1 x^1 \tag{1.269}$$

Traduzindo. Se o agente escolheu x^0 quando x^1 era factível, x^0 se revelou preferível a x^1. Então quando x^1 foi escolhido é porque x^0 não era factível. Note que aqui se trata simplesmente de uma regra que associa um vetor de preços e um nível de renda (logo um conjunto orçamentário) a uma escolha.

O axioma fraco parece gerar todas as propriedades da nossa teoria da escolha racional: homogeneidade da função escolha; matriz de Slutsky para a função de escolha $x^e(p,m)$ é negativa semi-definida; a demanda marshalliana $x^M(p,m)$ satisfaz ao AFrPR.

Somente no caso em que as preferências estão definidas somente para dois bens a matriz de Slutsky associada a $x^e(p,m)$ é necessariamente simétrica. No caso de mais de dois bens, o AFrPR não implica transitividade, que é o axioma da teoria baseada em preferências associado à simetria da matriz de Slutsky. Portanto, a função de escolha $x^e(p,m)$ não é necessariamente uma função de demanda Marshalliana.

MasCollel *et al.* (1995: 12–14) mostram que, se o consumidor atende o requisito de racionalidade a partir da teoria das preferências subjetivas, ou seja, às hipóteses de preferências completas e transitivas, ele obrigatoriamente passará no teste do axioma fraco da preferência revelada. Contudo, o inverso só será verdadeiro se o consumidor se confrontar com todas as escolhas possíveis duas a duas e passar no teste. Nesse sentido, as hipóteses de preferências completas e transitivas são mais fortes do que o axioma forte da preferência revelada.

O consumidor atenderá às hipóteses de preferências completas e transitivas se passar no teste do axioma forte da preferência revelada. O axioma forte da preferência revelada exige que se x for revelada preferível a y e y for revelada preferível a z, então x também será revelada preferível a z, ou seja, será indiretamente revelada preferível, enquanto o axioma fraco só implica a revelação direta de preferência.

Substituiremos, então, o AFrPR pelo Axioma Forte da Preferência Revelada (AFoPR) como forma de impor transitividade na nossa teoria da escolha.

Definição 1.11.2 (Axioma Forte da Preferência Revelada (AFoPR))**.** *Para qualquer sequência de cestas $x^0, x^1, \ldots, x^k$, tal que $x^0 \mathcal{R} x^1, x^1 \mathcal{R} x^2, \ldots, x^{k-1} \mathcal{R} x^k$, não é possível que $x^k \mathcal{R} x^0$.*

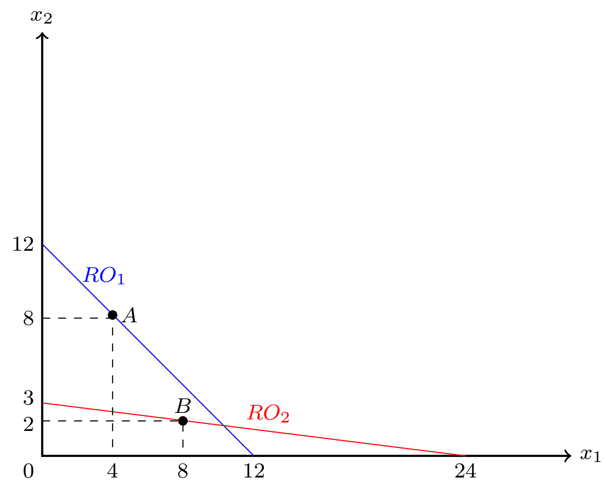
O AFoPR é simplesmente uma forma de impor transitividade na relação de preferência revelada. Na verdade, o AFoPR é o fecho transitivo do AFrPR.

Exemplo 1.11.1. *Um indivíduo com 12 reais, tendo que escolher combinações dos bens (x, y), comprou a cesta $(4, 8)$, quando o preço dos dois bens era de 1 real. Quando o preço do primeiro bem caiu para 50 centavos e o do segundo subiu para 4 reais, ele comprou a cesta $(8, 2)$. Ele está melhor ou pior após a mudança de preços?*

Temos que testar duas condições:

- *Se, aos preços antigos, ele consegue comprar a nova cesta (cesta B ou cesta no período $t = 1$), é porque, à época, poderia ter comprado a cesta nova B, assim como a antiga (cesta A ou cesta no período $t = 0$), mas escolheu a cesta A. Então, é porque ele está revelando a sua preferência direta pela cesta antiga A. Isto é, $\sum_{i=1}^{2} p_i^0 x_i^0 \leq \sum_{i=1}^{2} p_i^0 x_i^1$.*
- *Além disso, ele também tem que testar se, aos preços novos, ele não consegue comprar a cesta abtiga, isto é, $\sum_{i=1}^{2} p_i^0 x_i^0 \leq \sum_{i=1}^{2} p_i^1 x_i^1$.*

Se as duas condições forem satisfeitas, será possível afirmarmos que o indivíduo piorou de situação, pois, na nova situação (preços), a cesta antiga não é mais factível e, antes, quando as duas podiam ser compradas, ele preferiu a antiga.



- *Situação inicial* ($t = 0$)*:*
  - $x_0 = 4$ *e* $p_1^0 = 1$*: gasto = 4*
  - $y_0 = 8$ *e* $p_2^0 = 1$*: gasto = 8*
  - *Gasto total no período* $t = 0 \;\Rightarrow\; 4 + 8 = 12$
- *Situação final* ($t = 1$)*:*
  - $x_1 = 8$ *e* $p_1^1 = 0,5$*: gasto = 4*
  - $y_1 = 2$ *e* $p_2^1 = 4$*: gasto = 8*
  - *Gasto total no período* $t = 1 \;\Rightarrow\; 4 + 8 = 12$
- *Situação final aos preços da situação inicial:*
  - $x_1 = 8$ *e* $p_1^0 = 1$*: gasto = 8*
  - $y_1 = 2$ *e* $p_2^0 = 1$*: gasto = 2*
  - *Gasto total* $\Rightarrow\; 8 + 2 = 10$
- *Situação inicial aos preços da situação final:*
  - $x_0 = 4$ *e* $p_1^1 = 0,5$*: gasto = 2*
  - $y_0 = 8$ *e* $p_2^1 = 4$*: gasto = 32*
  - *Gasto total* $\Rightarrow\; 2 + 32 = 34$

Assim, de fato, aos preços p_1*, ele poderia ter comprado a cesta* B*, como poderia ter comprado a cesta* A*, mas optou pela cesta* A*. Além disso, com a mudança de preços, ele não é mais capaz de comprar a cesta que realmente gosta,* A*. Dessa forma, é possível afirmar que o indivíduo piorou de situação. Na nova situação,* A *não é mais factível. E quando* A *e* B *eram, ele preferiu* A*.*

Aplicações do AFrPR:

- Testar racionalidade
- Formato das curvas de indiferença
- AFoPR e racionalização
- Integrabilidade
- Índice de preços de Laspeyres vs. Paasche
- Efeito de um imposto sobre o bem-estar
- Ganhos de bem-estar do comércio internacional

Resumindo. Seja $A, B, \ldots, Z$ um conjunto distinto de cestas, cada uma encontrando uma restrição orçamentária linear. Seguindo Varian (1993), temos os seguintes conceitos:

Definição 1.11.3 (Diretamente Revelado Preferido)**.** *A é revelada diretamente preferida a B se B estava no conjunto de escolha quando A foi escolhido.*

Definição 1.11.4 (Indiretamente Revelado Preferido)**.** *Se A é diretamente revelada preferida a B, B é diretamente revelada preferida a C, ..., até Y, e Y é diretamente revelada preferida a Z, então, A é indiretamente revelada preferida a Z.*

Os axiomas de preferência revelados clássicos são devidos a Samuelson (1938) e Hauthakker (1950):

Definição 1.11.5 (Axioma Fraco da Preferência Revelada (WARP))**.** *Se A é revelada diretamente preferida a B, então B não é diretamente revelada preferida a A.*

Definição 1.11.6 (Axioma Forte da Preferência Revelada (SARP))**.** *Se A é indiretamente revelada preferível a B, então B não é diretamente revelada preferida a A.*

O WARP é necessário e o SARP é necessário e suficiente para a existência de preferências estritamente convexas que poderiam ter produzido os dados observados. Varian (1982), ao aplicar os teoremas de Afriat (1967), generalizou a teoria para permitir curvas de indiferença que não são estritamente convexas.

Definição 1.11.7 (Axioma Generalizado da Preferência Revelada (SARP))**.** *Se A é indiretamente revelada preferida a B, então B não é estritamente revelada diretamente como preferida a A, isto é, A não está estritamente dentro do orçamento quando B é escolhida.*

Observe que, se as cestas violarem o WARP, elas também deverão violar o SARP e, se violarem o GARP, elas também deverão violar o SARP, mas o contrário não é verdadeiro. Como Varian mostra, satisfazer o GARP é uma condição necessária e suficiente para a existência de preferências bem comportadas, dadas as restrições orçamentárias lineares.

1.12 Mudanças Compensadas e Equação de Slutsky

O Axioma Fraco da Preferência Revelada pode ser usado para gerar previsões sobre o comportamento do consumidor. Imagine dois vetores diferentes, (p, m) e (p', m'), tal que a cesta $z = x(p, m)$ está na fronteira de ambos $B_{p,m}$ e $B_{p',m'}$. Isso corresponde à seguinte situação hipotética. Suponha que os preços originais sejam (p, m) e você escolha a cesta $z = x(p, m)$. Eu digo a vocês que vou mudar o vetor de preço para p'. Mas, eu sou justo, e então eu lhes digo que, para ter certeza de que você não será prejudicado pela mudança de preço, eu também vou mudar sua riqueza para m', onde m' é escolhido para que você ainda possa adquirir a cesta z com os novos preços e riqueza (p', m'). Assim $m' = p'z$. Nós chamamos isso de uma mudança compensada no preço, já que eu mudo sua riqueza para compensar os efeitos da mudança de preço. Desde que você possa adquirir z antes e depois da mudança de preço, sabemos que:

$$pz = m \quad \text{e} \quad p'z = m'. \tag{1.270}$$

Seja $y = x(p', m') \neq z$ seja a cesta escolhida em (p', m'). Como você realmente escolhe y quando temos p' e m', assumindo que sua demanda satisfaz a lei de Walras, sabemos que $p'y = m'$. Portanto:

$$\begin{aligned} 0 &= m' - m' = p'y - p'z \\ 0 &= p'(y - z). \end{aligned} \tag{1.271}$$

Além disso, como z é acessível em (p', m'), pelo AFPR deve ser que y não era acessível em (p, m):

$$\begin{aligned} py &> m \\ py - pz &> 0 \\ p(y - z) &> 0. \end{aligned} \tag{1.272}$$

Finalmente, subtraindo $p(y - z) > 0$ de $p'(y - z) = 0$ produz:

$$(p' - p)(y - z) < 0. \tag{1.273}$$

A equação acima capta a ideia de que, após uma mudança de preço compensada, os preços e a demanda se movem em direções opostas. Para colocar de outra forma, seja $\Delta p = p' - p$ o vetor de mudanças de preço e $\Delta x = x(p', m') - x(p, m)$ o vetor de mudança na demanda. A equação acima pode ser reescrita como

$$\Delta p \cdot \Delta x^c \leq 0, \tag{1.274}$$

quando substituímos a desigualdade estrita por uma desigualdade fraca, em reconhecimento de que pode ser o caso de $y = z$. Note que o sobrescrito c em Δx^c é para nos lembrar que esta é a mudança compensada em x. Esta é uma declaração da Lei de Demanda Compensada: se o preço de um bem aumenta, você reduz sua demanda por ele. Se tomarmos uma visão de cálculo das coisas, podemos reescrever isso em termos de diferenciais:

$$dp \cdot dx^c \leq 0. \tag{1.275}$$

Estamos quase lá. Agora, o que significa dar ao consumidor uma mudança de preço compensada? Seja $\widehat{x}$ o pacote de consumo inicial, ou seja, $\widehat{x} = \widehat{x}(p, m)$, em que p e m são os preços e a renda originais. Uma alteração de preço compensada significa que, a qualquer preço, p, o pacote $\widehat{x}$ ainda é acessível. Assim, após a mudança de preço, a riqueza é alterada para $\widehat{w} = p\widehat{x}$. Observe que $\widehat{x}$ nesta expressão é a cesta de consumo original, não a função de escolha $x(p, m)$. Considere a demanda do consumidor pelo bem i:

$$x_i^c = x_i(p, p\widehat{x}), \tag{1.276}$$

após uma mudança compensada no preço do bem j:

$$\begin{aligned} \frac{d}{dp_j}(x_i(p, p\widehat{x})) &= \frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial m}\frac{\partial (p\widehat{x})}{\partial p_j} \\ \frac{dx_i^c}{dp_j} &= \frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial m}\widehat{x}_j. \end{aligned} \tag{1.277}$$

Como $\widehat{x}_j = x(p, m)$, deixamos de lado o "chapéu" a partir de agora. Se escrevermos a equação anterior como um diferencial, isso é simplesmente:

$$dx_i^c = \left(\frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial m}x_j\right) dp_j = s_{ij}dp_j. \tag{1.278}$$

Se mudarmos mais de um p_j, a mudança em x_i^c seria simplesmente a soma das alterações devido às diferentes mudanças de preço:

$$dx_i^c = \sum_{j=1}^{L} \left(\frac{\partial x_i}{\partial p_j} + \frac{\partial x_i}{\partial m}x_j\right) dp_j = s_i dp, \tag{1.279}$$

ou

$$dx^c = S \cdot dp. \tag{1.280}$$

Agora, retornando ao AFPR:

$$\begin{aligned} dp \cdot dx^c &\leq 0 \\ dp \cdot S \cdot dp &\leq 0. \end{aligned} \tag{1.281}$$

A desigualdade acima tem um significado matemático: implica que a matriz S, que chamaremos de matriz de substituição, seja negativa semi-definida. O que isto significa é que, se você pré e pós-multiplicar S pelo mesmo vetor, o resultado será sempre um número não positivo. Isso é importante porque os matemáticos descobriram várias boas propriedades de matrizes semi-definidas negativas.

Os elementos s_{ii} são não-positivos. Este é um resultado fundamental em economia, porque diz que a mudança na demanda por um bem em resposta a um aumento de preço compensado é negativa. Em outras palavras, se o preço subir, a demanda cai. Esta é a Lei da Demanda Compensada. Você pode estar pensando que foi muito trabalhoso extrair algo tão óbvio, mas o fato de a Lei da Demanda Compensada ser derivada do AFPR e da Lei de Walras é realmente muito importante. Se estes não fossem suficientes para a Lei da Demanda Compensada, que sabemos ser verdadeira, então isso seria um forte indicativo de que deixamos algo fora do nosso modelo.

O fato de que $s_{ii} \leq 0$ pode ser usado para ajudar a explicar uma curiosidade da teoria econômica, o bem de Giffen. Normalmente, pensamos que, se o preço de um bem aumentar, mantendo a riqueza constante, a demanda por esse bem diminuirá. Teoricamente, é possível que, quando o preço de um bem aumenta, o consumidor realmente escolha consumir mais. Como esta história se manifesta na teoria que aprendemos até agora? Nós sabemos isso:

$$s_{ii} = \left(\frac{\partial x_i}{\partial p_i} + \frac{\partial x_i}{\partial m} x_i \right). \tag{1.282}$$

Rearranjando, temos:

$$\frac{\partial x_i}{\partial p_i} = s_{ii} - \frac{\partial x_i}{\partial m} x_i. \tag{1.283}$$

Sabemos que $s_{ii} \leq 0$, já que S é negativa semi-definida. Claramente, $x_i \geq 0$. Mas, o que acontece se x_i é um bem fortemente inferior? Neste caso, $\dfrac{\partial x_i}{\partial m} < 0$, significando $-\dfrac{\partial x_i}{\partial m} x_i > 0$. E, se a magnitude de $-\dfrac{\partial x_i}{\partial m} x_i$ for maior que s_{ii}, pode ser que $\dfrac{\partial x_i}{\partial p_i} > 0$, que é a expressão matemática para um bem de Giffen. O que a teoria nos diz? Bem, isso nos diz que, para que um bem seja um bem Giffen, deve ser um bem fortemente inferior. Ou, para colocar de outra forma, um bem normal não pode ser um bem de Giffen.

Referências Bibliográficas

[Afriat1967] Afriat, S. N. (1967). The Construction of Utility Functions from Expenditure Data. *International Economic Review*, 8(1):67–77.

[Chipman et al.1971] Chipman, J. S. et al. (1971). Preferences, Utility, and Demand.

[Debreu1954] Debreu, G. (1954). Representation of a Preference Ordering by a Numerical Function. *Decision Processes*, 3:159–165.

[Debreu1959] Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* Yale University Press.

[Hurwicz1971] Hurwicz, L. (1971). On the Integrability of Demand Functions. *Preferences, Utility and Demand.*

[Jaffray1975] Jaffray, J.-Y. (1975). Existence of a Continuous Utility Function: An Elementary Proof. *Econometrica*, 43(5-6):981.

[Jehle2001] Jehle, G. A. (2001). *Advanced Microeconomic Theory.* Pearson Education India.

[Mas-Colell et al.1995] Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic Theory*, volume 1. Oxford University Press, New York.

[Samuelson1948] Samuelson, P. A. (1948). Consumption Theory in Terms of Revealed Preference. *Economica*, 15(60):243–253.

[Takayama1985] Takayama, A. (1985). *Mathematical Economics.* Cambridge University Press.

Capítulo 2

Escolha sob Incerteza

Contents

2.1 Axiomatização e Propriedades da Teoria da Utilidade Esperada

Os primeiros desenvolvimentos da teoria da probabilidade nos século XVII pelos matemáticos Blaise Pascal (1623-1662) e Pierre de Fermat (1601-1665) permitiram o surgimento da primeira teoria matemática que tratou do comportamento humano. Foi o próprio Pascal que a formulou. Pascal tinha um problema nada mundano a resolver – queria estabelecer a correção de levar ou não uma vida devotada a Deus. A incógnita quanto à existência divina foi a motivação responsável pelo surgimento da primeira teoria da decisão: o princípio da expectância matemática, que levou Pascal a defender a fé em Deus em termos pragmáticos. Segundo Pascal, apesar da razão não poder determinar a existência ou não de Deus, este, de fato, existe ou não, independente da nossa crença. Porém, a verdade somente nos seria revelada na ocasião de nossa morte. O problema é que precisaríamos decidir por alguma alternativa de vida antes de morrer. Para propor uma decisão, Pascal retomou o conceito de valor esperado. O objetivo de Pascal foi apontar uma racionalidade para a aposta na existência de Deus. Assim, argumentou que a decisão deveria ser baseada na comparação entre os valores esperados. Esta comparação, que ficou conhecida como princípio da expectância matemática. Uma propriedade interessante deste princípio é que a variância dos possíveis retornos não importa – a única medida importante para a tomada de decisões é o valor esperado. Contemporaneamente, diríamos que este princípio não leva em conta as possíveis atitudes dos indivíduos frente ao risco. Apesar desta limitação, que hoje é algo bastante clara – e naquela época ainda não era - o princípio da expectância matemática podia ser aplicado a vários tipos de problemas e constituiu-se no primeiro desenvolvimento intelectual capaz de lidar com decisões em condições de incerteza. Apesar da praticidade do princípio da expectância matemática, suas limitações não tardaram muito a aparecer. Evidências anômalas acumularam-se e Daniel Bernoulli foi o principal porta-voz da insatisfação com este princípio. Se o princípio da expectância matemática é adequado, como justificar a existência de seguros? Em 1738, Bernoulli publicou um ensaio no Commentarii Academiae Scientiarum Imperialis Petropolitanae (Autos da Academia Imperial de Ciências de São Petersburgo), apontando que um mercador prudente pode segurar seu navio contra perdas no mar mesmo sabendo que estará aumentando a riqueza esperada da companhia de seguros às custas da sua. Para alguém que mantivesse o princípio da expectância matemática em seus processos de decisão, seria tão absurdo fazer um seguro quanto jogar dinheiro fora. Como qualquer um concordaria que não é insanidade fazer um seguro, então o comportamento do mercador seria uma violação flagrante do princípio da expectância matemática.

Neste mesmo artigo, Bernoulli citou o hoje famoso Paradoxo de São Petersburgo, para enfatizar que "homens prudentes" não obedecem invariavelmente ao princípio da expectância matemática. O paradoxo foi publicado originalmente por Nicholas Bernoulli, seu primo, em 1731. O paradoxo pode ser apresentado do seguinte modo: suponha que uma moeda é jogada repetidamente até que a primeira "cara" apareça. O jogo paga 2^{n-1} reais se a primeira cara aparecer na n-ésima jogada. Qual o preço que um indivíduo pagaria para entrar neste jogo?

Se o indivíduo se baseasse no princípio da expectância matemática, ele estaria disposto a pagar, no máximo, o valor da esperança matemática. Como o valor da esperança matemática é

$$
\begin{aligned}
\mathbb{E}(L) &= \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot 2^{n-1} \\
&= \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 4 + \ldots \\
&= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots = +\infty, \qquad (2.1)
\end{aligned}
$$

ele estaria disposto a pagar qualquer preço para entrar neste jogo. Não importa o quão rico fosse o indivíduo, ele estaria disposto a entregar toda a sua riqueza para poder participar deste jogo. O que certamente destoaria do comportamento observável no mundo real, onde a maioria das pessoas não estariam dispostas a pagar mais do que uns poucos dólares para participar deste jogo.

A solução que Bernoulli propôs para o Paradoxo de São Petersburgo é considerada o marco inicial da teoria da utilidade esperada (EU). Bernoulli argumentou que o valor que uma pessoa atribui a sua riqueza não é o próprio valor monetário desta, mas sim seu "valor moral" ou utilidade:

> "[...] a determinação do valor de um item não pode ser baseado em seu preço, mas sim na utilidade que ele fornece. O preço de um item depende somente do próprio item e é igual para todo mundo; a utilidade, contudo, depende das circunstâncias particulares do indivíduo que faz a estimativa."

Bernoulli postulou o que mais tarde seria conhecido como a lei da utilidade marginal decrescente, que implica que à medida que a riqueza aumenta, decresce a utilidade adicional devido ao aumento da riqueza. Em termos matemáticos, esta lei diz que a utilidade em função do dinheiro ou da riqueza é uma função côncava.

Bernoulli foi além e supôs que a utilidade é igual ao logaritmo (em qualquer base) do resultado em termos monetários. Ou seja, $u(x) = \log_B x$ em que x é o resultado e B é uma base qualquer ($B > 0$ e $B \neq 1$). O cálculo da utilidade esperada é semelhante ao cálculo do valor esperado. Assim, a utilidade esperada de uma loteria é

$$U(L) = \sum_i p_i u(x_i). \qquad (2.2)$$

Bernoulli afirmou que os indivíduos procuram maximizar a "esperança moral" ou, equivalentemente, a utilidade esperada dos resultados.

Aplicando o conceito de utilidade esperada ao Paradoxo de São Petersburgo e supondo que a utilidade de qualquer resultado é igual ao logaritmo na base 10 do resultado, obtemos:

$$U(L) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot \log_{10}(2^{n-1}) = 0,30103. \qquad (2.3)$$

Isto significa que, sendo X o preço máximo que o indivíduo estaria disposto a pagar para entrar no jogo, temos

$$U(X) = \log_{10} X = 0,30103 \quad X = 2, \tag{2.4}$$

isto é, o indivíduo estaria disposto a pagar, no máximo, 2 reais.

De fato, podemos mostrar que independentemente da base que utilizássemos para o logaritmo, o indivíduo estaria disposto a pagar apenas 2 reais.

Demonstração. Suponha que

$$U(L) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot \log_B \left(2^{n-1}\right). \tag{2.5}$$

Então,

$$\begin{aligned}
U(X) = \log_B X &= \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot \log_B \left(2^{n-1}\right) \\
X &= B^{\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot \log_B \left(2^{n-1}\right)} \\
X &= B^{\left(\frac{1}{2}\log_B 1 + \frac{1}{4}\log_B 2 + \frac{1}{8}\log_B 4 + \ldots\right)} \\
&= B^{\left(\log_B 1 + \log_B 2^{\frac{1}{4}} + \log_B 4^{\frac{1}{8}} + \ldots\right)} \\
&= B^{\left[\log_B \left(1 \cdot 2^{\frac{1}{4}} \cdot 4^{\frac{1}{8}} \ldots\right)\right]} \\
&= B^{\log_B \left[\prod_{n=1}^{\infty} \left(2^{(n-1)}\right)^{\left(\frac{1}{2}\right)^n}\right]} \\
&= \prod_{n=1}^{\infty} \left(2^{(n-1)}\right)^{\left(\frac{1}{2}\right)^n} \\
&= \prod_{n=1}^{\infty} 2^{\frac{(n-1)}{2^n}} \\
\log X &= \log 2 \underbrace{\sum_{n=1}^{\infty} \frac{n-1}{2^n}}_{\to 1} \\
\log X &= \log 2
\end{aligned}$$

$$X = 2. \tag{2.6}$$

■

Os créditos do desenvolvimento inicial da EU não devem ser atribuídos unicamente a Bernoulli. Em 1731, sete anos antes da publicação histórica de Bernoulli, Gabriel Cramer forneceu uma solução semelhante para o paradoxo através de uma carta enviada ao primo de Daniel Bernoulli, Nicholas Bernoulli. Em um pós-escrito ao artigo de 1738, Daniel Bernoulli reconheceu o trabalho de Cramer, citando várias partes.

Com a teoria da utilidade esperada, a subjetividade foi definitivamente introduzida à teoria da decisão. Para efetuar cálculos utilizando o princípio da expectância matemática, não era necessário fazer qualquer tipo de avaliação subjetiva, bastava multiplicar as probabilidades pelos possíveis resultados. Com a EU, a avaliação subjetiva dos tomadores de decisão passou a ter um papel fundamental. Os possíveis resultados e as probabilidades passaram a não ser mais suficientes para determinar a decisão pois "a utilidade ... depende das circunstâncias específicas de quem faz a estimativa ... Não há razão para supor que os riscos estimados por cada indivíduo devam ser considerados de mesmo valor".

Na década de 1940, a teoria da utilidade esperada já era bicentenária. Às vezes, alguém fazia referência à teoria de Bernoulli, sugerindo que a maximização da utilidade esperada seria um meio adequado de representar as preferências dos indivíduos em condições de incerteza. Porém, a sugestão se encerrava em uma fraqueza: não havia razão para supor que as escolhas dos indivíduos seriam suportadas pela teoria da utilidade esperada. Por que especificamente a utilidade esperada seria a medida relevante para representar a tomada de decisões? Por que não utilizar a variância, a amplitude, a curtose ou outras características de uma função para determinar as preferências?

John Von Neumann e Oskar Morgenstern, em sua obra seminal publicada em 1944, "Theory of games and economic behavior", forneceram a "resposta", elaborando as bases axiomáticas para a teoria da utilidade esperada. Eles mostraram que a maximização da utilidade esperada é logicamente equivalente à hipótese de que o comportamento de escolha satisfaz algumas restrições sob a forma de axiomas. Assim, se estes axiomas são satisfeitos, então é possível construir uma função utilidade esperada que represente as preferências de um indivíduo. A relevância deste resultado é que se estes axiomas são plausíveis, então a hipótese da utilidade esperada também é. E, portanto, pode ser aplicada para modelar o comportamento dos tomadores de decisão.

O primeiro passo da representação é definir um conjunto de resultados possíveis ou conjunto de prêmios de uma situação de escolha. Assim, chamaremos este conjunto de X e definiremos $X = [0, M] \subseteq \mathbb{R}$ como um intervalo compacto onde $M > 0$. Cada elemento de X pode ser considerado uma quantidade não-negativa de dinheiro até um valor limite M, que é o resultado mais alto possível.

As medidas de probabilidade podem ser representadas por suas funções distribuições acumuladas. Denotaremos por $D(X)$ o conjunto de todas funções distribuições acumuladas sobre X. Formalmente, este conjunto é formado por todas funções contínuas à direita e não-decrescentes $F\colon X \to \mathbb{R}$ tal que $F(0) \geq 0$ e $F(M) = 1$. Cada função $F \in D(X)$ determina uma única medida de probabilidade P em X, através da equação

$$P([0, X]) = F(x) \text{ para todo } x \in X. \tag{2.7}$$

Assim, se um indivíduo se defronta com uma situação de risco representada pela função distribuição acumulada F, $F(x)$ é a probabilidade de receber uma quantia menor ou igual a x.

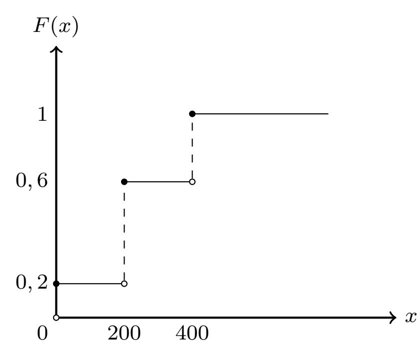
O conjunto de todos elementos de $D(X)$ com imagem finita, que correspondem às medidas de probabilidade com suporte finito, é denotado por $D^0(X)$. Se F tem imagem finita, então F tem um conjunto finito de pontos descontínuos $\{x_1, x_2, \ldots, x_n\}$ tal que $x_1 < x_2 < \ldots < x_n$. Assumindo que $F(x_0) = 0$, denotaremos o "salto" de F em x_i, $i = 1, \ldots, n$, por p_i. Portanto, $\sum_i p_i = 1$. Destes "saltos", obtemos uma lista $L = (p_1, \ldots, p_n)$ com $p_i \geq 0$ para todo i em que p_i é a probabilidade do resultado (ou do prêmio) i ocorrer. Chamaremos esta lista de uma distribuição de probabilidade simples ou, como mais usualmente denominada no contexto da teoria da decisão, de uma loteria simples.

Exemplo 2.1.1. *Seja a função distribuição acumulada F tal que*

$$F(x) = \begin{cases} 0, & \textit{se } x < 0 \\ 0,2, & \textit{se } 0 \leq x < 200 \\ 0,6, & \textit{se } 200 \leq x < 400 \\ 1, & \textit{se } x \geq 400 \end{cases} \tag{2.8}$$

Esta função está representada no gráfico abaixo. Note que embora esta função não seja contínua, ela é contínua à direita. Note também que ela possui imagem finita.

Figura 2.1 – Função distribuição acumulada



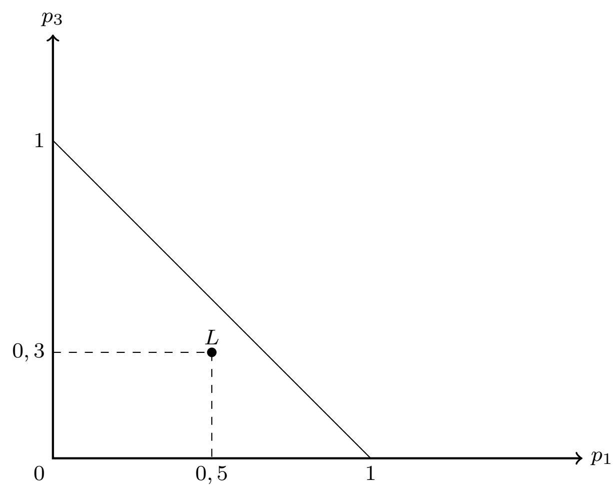
A partir de cada ponto descontínuo x_i em F, obtemos uma probabilidade p_i associada ao prêmio x_i tal que $p_i = F(x_i) - F(x_0)$. No nosso exemplo, $\{x_1, x_2, x_3\} = \{0, 200, 400\}$ e $\{p_1, p_2, p_3\} = \{0,2; 0,4; 0,4\}$. Portanto, a partir da função F dada, é possível obter uma loteria simples $L = \{0,2; 0,4; 0,4\}$, referente aos prêmios 0, 200 e 400.

Para efetuar nossas análises, iremos supor que as preferências são monotônicas - ou seja, "mais dinheiro é preferível a menos". Isto é também chamado de monotonicidade das preferências.

Assim, por exemplo, um tomador de decisão preferirá uma loteria simples que forneça \$15 com probabilidade 1 à outra que forneça \$10 com probabilidade 1.

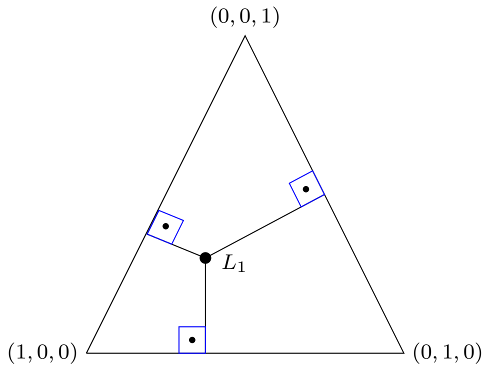
Quando uma loteria simples tem três resultados possíveis ($n = 3$), podemos representá-la em um triângulo de Marschak-Machina ou diagrama em triângulo, ordenando os três resultados possíveis x_1, x_2, x_3 de tal forma que $x_3 > x_2 > x_1$. Dadas as respectivas probabilidades p_1, p_2, p_3, sabemos que $p_1 + p_2 + p_3 = 1$. Portanto, definindo duas probabilidades, a terceira já está completamente definida. No eixo das abscissas, representamos a probabilidade do resultado mais baixo, p_1, e no eixo das ordenadas, a probabilidade do resultado mais alto, p_3. A probabilidade do resultado mediano, dois, é sempre $p_2 = 1 - p_1 - p_3$.

Figura 2.2 – Triângulo de Marschak-Machina



A representação de loterias através do triângulo de Marschak-Machina não é a única possibilidade disponível na literatura econômica. As loterias podem também ser representadas como um ponto no simplex dimensional $(n-1)$, $\Delta = \{p \in \mathbb{R}^n_+ : p_1 + \ldots + p_n = 1\}$. Quando os resultados possíveis são três, podemos utilizar um triângulo equilátero para representar as loterias. Os triângulos equiláteros apresentam uma propriedade muito útil para efetuar a representação: a soma das perpendiculares de qualquer ponto até os três lados do triângulo é igual a altura do triângulo. Assim, estabelecendo um triângulo equilátero com altura 1, temos a propriedade geométrica de que a probabilidade p_n do resultado n de uma loteria associada a algum ponto deste simplex é igual ao comprimento desta perpendicular, que inicia no ponto em questão e termina no lado oposto ao vértice n. No triângulo equilátero abaixo, representamos a loteria $L_1 = (0,5; 0,2; 0,3)$.

Figura 2.3 – Simplex dimensional



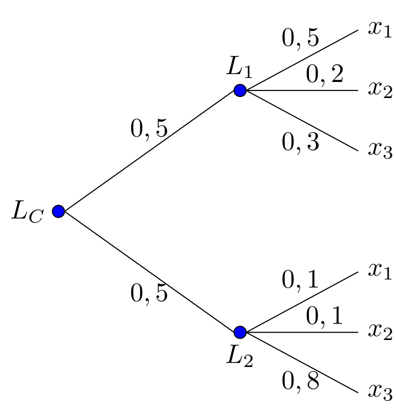
Assim, o comprimento da perpendicular que inicia no ponto L_1 e termina no lado oposto ao vértice 1 é a probabilidade do resultado 1. Neste caso, como o comprimento é 0,5, ele indica que a probabilidade do resultado 1 é 0,5. Da mesma forma, o comprimento das outras perpendiculares mostram que a probabilidade do resultado 2 é 0,3 e a probabilidade do resultado 3 é 0,2.

Além da loteria simples, podemos definir a loteria composta. Dadas k loterias simples $L_k = (p_1^k, \ldots, p_n^k), k = 1, \ldots, K$ e probabilidades $\alpha_k > 0$ com $\sum_k \alpha_k = 1$, a loteria composta $(L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$ é a alternativa de risco que fornece a loteria simples L_k com probabilidade α_k para $k = 1, \ldots, K$.

Assim, uma loteria composta é uma loteria em que os resultados não são prêmios monetários mas, loterias simples.

Exemplo 2.1.2. *Seja a loteria composta* $L_C = (L_1, L_2; 0,5; 0,5)$*, referente às seguintes loterias simples:* $L_1 = (0,5; 0,2; 0,3)$ *e* $L_2 = (0,1; 0,1; 0,8)$*. Então, a loteria composta* L_C *é tal que fornece a loteria simples* L_1 *com probabilidade 0,5 e a loteria simples* L_2 *com probabilidade 0,5. A representação dessa loteria é apresentada na Figura 2.4.*

Figura 2.4 – Representação de uma loteria composta



Repare que a probabilidade de obtermos x_1 *com esta loteria composta é a probabilidade de* x_1 *acabar sendo sorteado por meio da loteria* L_1 *mais a probabilidade de* x_1 *acabar sendo sorteado* *por meio da loteria simples L_2, ou seja,* $0,25 + 0,05$*, que é igual a* $0,3$*. Da mesma maneira, a probabilidade de obtermos x_2 é* $0,15$ *e a probabilidade de obtermos x_3 é* $0,55$*.*

Von Neumann e Morgenstern perceberam que era necessário um instrumental teórico capaz de lidar tanto com loterias simples como com loterias compostas. Porém, uma teoria que lidasse diretamente com loterias compostas geraria uma série de complicações que seriam mais difíceis de tratar. Por outro lado, a loteria composta é um conceito importante - uma parte significativa dos fenômenos econômicos do mundo real não correspondem a loterias simples. A solução que Von Neumann e Morgenstern apresentaram, bastante profícua, foi o axioma do consequencialismo.

Toda loteria composta pode ser reduzida a uma loteria simples. Porém, a princípio, isto não significa que elas sejam intercambiáveis; isto é, que elas sejam equivalentes para o tomador de decisão. O papel do axioma do consequencialismo é exatamente impor a equivalência entre a loteria composta e a sua reduzida. Formalmente, temos:

Axioma 2.1.1 (Consequencialismo)**.** *Se L é a loteria reduzida da loteria composta* $(L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$*, então* $L \sim (L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$*.*

O axioma do consequencialismo afirma que somente a probabilidade sobre os resultados finais é de relevância para o tomador de decisão. Não importa se as loterias são apresentadas em vários estágios ou não (i.e., se são ou não loterias compostas), desde que as probabilidades sobre os resultados finais sejam as mesmas, o tomador de decisão será indiferente entre elas.

Seguindo o axioma consequencialista, assumiremos que somente as loterias reduzidas sobre os resultados finais são de relevância para o tomador de decisão. Note que toda loteria simples é também a reduzida de si mesma. Assim, iremos definir o conjunto de loterias simples e, posteriormente, os outros axiomas serão estabelecidos em relação a este conjunto.

Definimos o conjunto de alternativas, denotado por $\mathscr{L}$, como o conjunto de todas loterias simples (ou distribuições de probabilidade simples) sobre o conjunto de resultados X. A partir do conjunto de loterias simples, podemos definir uma relação de preferência. As preferências do tomador de decisão serão formalizadas pela relação binária $\succeq$, subconjunto do produto cartesiano $\mathscr{L} \times \mathscr{L}$. A relação de preferência$\succeq$ é chamada de relação de preferência fraca. Se $L \succeq L'$, então lemos "L é fracamente preferível a L'". A partir de $\succeq$, definiremos duas outras relações: (i) relação de preferência estrita; e (ii) relação de indiferença.

Para gerar instrumentos de análise das escolhas dos indivíduos, é necessário impor algum tipo de consistência sobre as suas preferências, de forma que possibilite o tratamento matemático. Von Neumann e Morgenstern impuseram uma consistência através da suposição da racionalidade ou ordenabilidade das preferências.

A representação das preferências por uma função real requer uma suposição de continuidade. Von Neumann e Morgenstern utilizaram o axioma da continuidade arquimediana.

Axioma 2.1.2 (Continuidade Arquimediana)**.** *A relação de preferências $\succeq$ no espaço das loterias simples $\mathscr{L}$ é contínua se para qualquer $L, L', L'' \in \mathscr{L}$ e $L \succeq L' \succeq L''$, existe um $\alpha, \beta \in (0,1)$ tal que $\alpha L + (1-\alpha)L'' \succeq L' \succeq \beta L + (1-\beta)L''$. A continuidade implica que pequenas mudanças marginais nas probabilidades não alteram o ordenamento entre as loterias.*

Além da racionalidade e da continuidade, a axiomatização de Von Neumann e Morgenstern inclui o axioma da independência, que desempenha um papel central na teoria da utilidade esperada. Curiosamente, Von Neumann e Morgenstern assumiram apenas implicitamente este axioma, sem fazer referência direta a ele.

Axioma 2.1.3 (Independência)**.** *A relação de preferências $\succeq$ no espaço de loterias simples $\mathscr{L}$ satisfaz o axioma da independência se para todo $L, L', L'' \in \mathscr{L}$ e $\alpha \in (0,1)$, temos*

$$L \succeq L' \iff \alpha L + (1-\alpha)L'' \succeq \alpha L' + (1-\alpha)L''.$$

O axioma da independência afirma que se um indivíduo considera a loteria L preferível a L', então uma loteria composta que forneça L com probabilidade α e L'' com probabilidade $(1-\alpha)$ é preferível a uma loteria composta que forneça L' com probabilidade α e L'' com probabilidade $(1-\alpha)$. Em outras palavras, se misturarmos a loteria L com uma loteria L'' e a loteria L' com a loteria L'', L'' entrando sempre com a mesma probabilidade, então o ordenamento das duas misturas não depende (é independente) da loteria L''. O ordenamento das duas misturas depende apenas do ordenamento entre L e L'.

Com o estabelecimento do axioma da independência, completamos a lista dos axiomas necessários e suficientes para a existência de uma representação de preferências através de uma função utilidade com a forma de utilidade esperada. Esta forma, como vimos, foi pioneiramente proposta por Bernoulli no século XVIII. Abaixo, temos a definição formal de forma de utilidade esperada.

Uma função utilidade $U\colon \mathscr{L} \to \mathbb{R}$ tem a forma de utilidade esperada $U(L) = \sum_i p_i u(x_i)$ se e somente se é linear nas probabilidades; isto é, se e somente se satisfaz a propriedade

$$U\left(\sum_{k=1}^{K} \alpha_k L_k\right) = \sum_{k=1}^{K} \alpha_k U(L_k). \tag{2.9}$$

para qualquer K loterias $L_k \in \mathscr{L}$, $k = 1, \ldots, K$ e probabilidades $(\alpha_1, \ldots, \alpha_K) \geq 0$, com $\sum_k \alpha_k = 1$. Uma função utilidade $U\colon \mathscr{L} \to \mathbb{R}$ com a forma de utilidade esperada é chamada de função utilidade esperada de Von Neumann-Morgenstern (função utilidade v.N-M). A função utilidade u (sobre as consequências) é chamada de função utilidade de Bernoulli.

Veremos agora, o resultado mais importante da teoria da decisão sob incerteza, o chamado teorema da utilidade esperada ou teorema de Von Neumann-Morgenstern. Este teorema afirma que se as preferências são racionais, contínuas e satisfazem o axioma da independência, então elas são representáveis por uma função utilidade com a forma de utilidade esperada.

Teorema 2.1.1 (Teorema de Von Neumann-Morgenstern)**.** *Se a relação de preferências $\succeq$ é racional e satisfaz os axiomas da continuidade e da independência, então existe uma função utilidade com a forma de utilidade esperada que representa $\succeq$, uma função $U\colon \mathscr{L} \to \mathbb{R}$ tal que:*

$$L \succeq L' \iff \sum_{i=1}^{n} p_i u(x_i) \geq \sum_{i=1}^{n} p_i' u(x_i). \tag{2.10}$$

para qualquer $L = (p_1, \ldots, p_n)$ *e* $L' = (p_1', \ldots, p_n')$ *tal que* $L, L' \in \mathscr{L}$*. Em outras palavras,* $\succeq$ *possui uma representação de utilidade esperada. Cada resultado possível tem um nível de utilidade correspondente e o valor de cada loteria simples (ou distribuição de probabilidade simples) é medido pela utilidade esperada que é designada a cada uma.*

O teorema da utilidade esperada foi provado por Von Neumann e Morgenstern na segunda edição (1947) de "Theory of Games and Economic Behavior". Mais tarde, Herstein e Milnor (1953) simplificaram a prova original. Várias variantes do teorema da utilidade esperada foram posteriormente provadas.

Demonstração. Abaixo apresentamos uma prova do teorema da utilidade esperada, baseada em Gollier (2001).

Digamos que L_1 e L_2 são, respectivamente, a loteria menos preferida e a loteria mais preferida em $\mathscr{L}$, obtidas pela resolução do problema de minimização e maximização de $U(L)$ em $\mathscr{L}$. Assim, para qualquer $L \in \mathscr{L}$, temos $L_2 \succeq L \succeq L_1$. Pelo axioma da continuidade arquimediana, existem escalares $\alpha^a, \alpha^b \in [0,1]$ tais que $L^a \sim \alpha^a L_2 + (1-\alpha^a)L_1$ e $L^b \sim \alpha^b L_2 + (1-\alpha^b)L_1$. Suponha que $\alpha^a \geq \alpha^b$. Note que $L^a \succeq L^b \iff \alpha^a \geq \alpha^b$. Seja $\gamma = \dfrac{\alpha^a - \alpha^b}{1-\alpha^b}$ tal que $\gamma \in [0,1]$, então:

$$\begin{aligned} \alpha^a L_2 + (1-\alpha^a)L_1 &\sim \gamma L_2 + (1-\gamma)[\alpha^b L_2 + (1-\alpha^b)L_1] \\ &\succeq \gamma[\alpha^b L_2 + (1-\alpha^b)L_1] + (1-\gamma)[\alpha^b L_2 + (1-\alpha^b)L_1] \\ &\sim \alpha^b L_2 + (1-\alpha^b)L_1. \end{aligned} \tag{2.11}$$

Assim, $U(L) = \alpha$, em que α é tal que $L \sim \alpha L_2 + (1-\alpha)L_1$. Logo, $U(L^a) = \alpha^a$ e $U(L^b) = \alpha^b$. Assim, $U(L^a) \geq U(L^b) \iff L^a \succeq L^b$.

Falta ainda provar que $U[\beta L^a + (1-\beta)L^b] = \beta\alpha^a + (1-\beta)\alpha^b$ ou, equivalentemente, que

$$\beta L^a + (1-\beta)L^b \sim [\beta\alpha^a + (1-\beta)\alpha^b]L_2 + [\beta(1-\alpha^a) + (1-\beta)(1-\alpha^b)]L_1. \tag{2.12}$$

Utilizando o axioma da independência duas vezes, temos:

$$\begin{aligned} \beta L^a + (1-\beta)L^b &\sim \beta[\alpha^a L_2 + (1-\alpha^a)L_1] + (1-\beta)L^b \\ &\sim \beta[\alpha^a L_2 + (1-\alpha^a)L_1] + (1-\beta)[\alpha^b L_2 + (1-\alpha^b)L_1] \\ &\sim [\beta\alpha^a + (1-\beta)\alpha^b]L_2 + [\beta(1-\alpha^a) + (1-\beta)(1-\alpha^a)]L_1. \end{aligned} \tag{2.13}$$

■

Um aspecto que gerou muita polêmica à cerca da teoria de Von Neumann-Morgenstern é que ela implica em uma representação de preferências através de um indexador cardinal. Esta cardinalidade foi motivo de descontentamento por parte de muitos economistas, que entenderam que significaria a volta da "utilidade sensitiva", já anteriormente banida pelo surgimento da abordagem ordinal-operacionalista. Assim, a "controvérsia da mensurabilidade" foi reaberta; de um lado, estavam aqueles que sustentavam a possibilidade de uma interpretação operacional da cardinalidade; de outro, aqueles que acreditavam que a aceitação do indexador de Von Neumann-Morgenstern requereria fundamentalmente a volta da "psicologia de sensação".

Porém, o debate à cerca da mensurabilidade acabou por demonstrar que a cardinalidade da função utilidade de Von Neumann-Morgenstern poderia ser interpretada em termos puramente operacionais. A cardinalidade da função utilidade de Von Neumann-Morgenstern teria a finalidade última apenas de ordenar as preferências sobre as loterias.

Baumol (1958) forneceu um exemplo elucidativo. Suponha que queiramos comprar dois pedaços de tecido que sejam compridos o suficiente para cobrir completamente uma mesa de nossa casa. Gostaríamos de predizer quais dois pedaços de tecidos cobrirão nossa mesa sem precisar testá-los sobre a mesa e, caso seja necessário, voltar à loja para trocá-los. Neste caso, um indexador pode nos ajudar. Se a nossa mesa tem 3m de comprimento, então podemos comprar dois pedaços de tecidos com $1,5$m e $2,25$m de comprimento, que cobrirão com facilidade a nossa mesa. Como a combinação de tecidos é indexada por $1,5 + 2,25 = 3,75$, um número maior do que 3, então seu comprimento é maior do que da mesa e, portanto, pode ser utilizada para cobri-la por completo. O indexador de comprimento foi elaborado de forma que possibilite que as combinações de objetos possam ser ordenadas pela simples soma dos comprimentos individuais dos objetos, sem a necessidade de medir diretamente as combinações. Este indexador é cardinal pois não serve apenas para ordenar os comprimentos de objetos, mas também para ordenar o comprimento das combinações de objetos.

Este mesmo raciocínio pode ser aplicado à função utilidade de Von Neumann-Morgenstern. Queremos atribuir números para os possíveis resultados (através da função utilidade de Bernoulli) de maneira que quando combinados através da função utilidade de Von Neumann-Morgenstern, possamos predizer o comportamento de escolha dos indivíduos sem a necessidade de conhecer toda sua ordenação de preferências em relação às loterias. E, para atribuir números, não é necessário a utilização de nenhum tipo de "psicologia de sensação".

O que se poderia questionar é se é efetivamente possível atribuir números para os possíveis resultados de forma que a teoria da utilidade esperada nos permita efetuar a predição do comportamento de um indivíduo. Mas isto nós já estabelecemos - se as tomadas de decisão de um indivíduo forem compatíveis com os axiomas da EU, então podemos fazer esta atribuição. Neste caso, basta observar algumas decisões de escolha do indivíduo em relação a loterias, derivar a função utilidade de Bernoulli destas observações e utilizar a função utilidade de Von Neumann-Morgenstern para predizer qualquer escolha do indivíduo em relação a qualquer loteria.

Portanto, a indexação cardinal oferecida pela função utilidade de Von Neumann-Morgenstern é compatível com uma interpretação operacional da utilidade, como no caso sem incerteza. A diferença é que no caso sem incerteza, a função utilidade era um indexador ordinal. Assim, a função utilidade era única sobre transformações monotônicas crescentes. No caso com incerteza, isto não é verdade, transformações monotônicas crescentes podem não preservar a forma de utilidade esperada da função utilidade. Isto acontece porque a forma de utilidade esperada é uma propriedade cardinal de uma função utilidade. Assim, a preservação somente é garantida através de transformações lineares crescentes, uma classe particular das transformações monotônicas crescentes.

Podemos, então, afirmar a seguinte proposição à cerca da unicidade da função utilidade esperada.

Proposição 2.1.1. *Suponha que $U\colon \mathscr{L} \to \mathbb{R}$ é uma função utilidade esperada de Von Neumann-Morgenstern para a relação de preferências $\succeq$ em $\mathscr{L}$. Então $V\colon \mathscr{L} \to \mathbb{R}$ é outra função utilidade esperada de Von Neumann-Morgenstern para $\succeq$ se e somente se existem escalares $a > 0$ e $b > 0$ tal que $V(L) = a + bU(L)$ para todo $L \in \mathscr{L}$. Uma função utilidade esperada de Von Neumann-Morgenstern é única sobre (exceto por) transformações lineares crescentes (positivas).*

Ou seja, se efetuarmos uma transformação linear crescente em uma função utilidade de Von Neumann-Morgenstern, ela continuará sendo uma função com forma de utilidade esperada, representando as mesmas preferências. Além disso, se duas funções utilidade de Von Neumann-Morgenstern representam as mesmas preferências, então elas são transformações lineares crescentes entre si. Sumarizando, podemos afirmar que uma função utilidade esperada de Von Neumann-Morgenstern é única sobre transformações lineares crescentes.

Existem duas propriedades particulares das transformações lineares que têm grande importância para a teoria da utilidade esperada:

- se o incremento de uma função utilidade é sempre positivo, ele será sempre positivo também para todas transformações lineares positivas desta função.
- se o incremento de uma função utilidade decresce (cresce) com o aumento dos prêmios, então o incremento das transformações lineares positivas desta função também será decrescente (crescente).

A primeira propriedade acima afirma, em termos matemáticos, que o sinal da primeira derivada é invariante a transformações lineares positivas. A segunda propriedade nos diz, em termos matemáticos, que o sinal da segunda derivada é invariante a transformações lineares positivas. A primeira propriedade é compartilhada com a classe mais abrangente das transformações monotônicas crescentes. Porém, a segunda propriedade, em geral, não é verdadeira para transformações monotônicas crescentes (exceto para o caso particular de transformações lineares crescentes). Verifiquemos matematicamente estas propriedades. Seja U uma função utilidade de Von Neumann-Morgenstern, V uma transformação linear crescente de U e u uma função utilidade de Bernoulli. Então, com $U(L) = \sum_i p_i u(x_i)$, temos:

$$\frac{\partial U(L)}{\partial x_i} = p_i \frac{\partial u(x_i)}{\partial x_i}. \tag{2.14}$$

Por outro lado, $V(L) = a + bU(L)$ implica:

$$\begin{aligned}\frac{\partial V(L)}{\partial x_i} &= b\frac{\partial U(L)}{\partial u(x_i)}\frac{\partial u(x_i)}{\partial x_i} \\ &= bp_i\frac{\partial u(x_i)}{\partial x_i}. \end{aligned} \tag{2.15}$$

Assim,

$$\begin{aligned}\frac{\partial V(L)}{\partial x_i} &= b\frac{\partial U(L)}{\partial x_i} \\ \frac{\partial^2 V(L)}{\partial x_i^2} &= b\frac{\partial^2 U(L)}{\partial x_i^2}. \end{aligned} \tag{2.16}$$

Como $b > 0$, então o sinal da primeira derivada de V em relação ao prêmio x_i é igual ao sinal da primeira derivada de U em relação ao prêmio x_i. Podemos observar que o sinal da derivada segunda também é invariante. A invariância do sinal da derivada primeira tem um papel evidente, que é assegurar que a transformação não altere a ordenação das loterias. A importância da invariância do sinal da segunda derivada está relacionada com os conceitos de utilidade marginal crescente e decrescente, que, como veremos, desempenham um papel fundamental na teoria da utilidade esperada. Se o sinal da segunda derivada é invariante à determinada transformação, então a utilidade marginal crescente ou decrescente é preservada.

Vimos até aqui que quando uma função utilidade de Von Neumann-Morgenstern é submetida a uma transformação linear crescente, o resultado é outra função utilidade de Von Neumann-Morgenstern que representa as mesmas preferências. Porém, e se a transformação linear crescente for efetuada na função utilidade de Bernoulli, a forma de utilidade esperada da função utilidade de Von Neumann-Morgenstern será preservada? E esta continuará representando as mesmas preferências? Vamos verificar nossas indagações. Seja U a função utilidade de Von Neumann-Morgenstern referente à função utilidade de Bernoulli u e V referente à função utilidade de Bernoulli v, onde v é uma transformação linear crescente de u. Portanto, seja $U(L) = \sum_i p_i u(x_i)$ e $v(x_i) = a + bu(x_i)$. Assim, temos:

$$\begin{aligned} V(L) &= \sum_i p_i[a + bu(x_i)] \\ &= \sum_i p_i a + \sum_i p_i bu(x_i) \\ &= a + b\sum_i p_i u(x_i) \\ &= a + bU(L). \end{aligned} \tag{2.17}$$

Portanto, a nova função utilidade V é uma transformação linear crescente de U. Assim, efetuar uma transformação linear crescente sobre a função utilidade de Bernoulli equivale a efetuar uma transformação linear crescente sobre a função utilidade de Von Neumann-Morgenstern. O contrário também é verdadeiro. Assim, transformações lineares crescentes sobre a função utilidade de Von Neumann-Morgenstern ou sobre a função utilidade de Bernoulli são fundamentalmente equivalentes.

Há ainda um aspecto importante que deve ser observado. Dado uma função utilidade com a forma de utilidade esperada de Von Neumann-Morgenstern U, vimos que somente transformações lineares crescentes de U podem gerar funções utilidade de Von Neumann-Morgenstern que representem as mesmas preferências. Porém, isto não significa que uma função obtida por uma transformação monotônica crescente qualquer de U, digamos, V, não represente as mesmas preferências que U. Mesmo que a transformação monotônica não seja linear, V representa as mesmas preferências que U. O que acontece é que no caso de uma transformação monotônica crescente não-linear, apesar da ordenação das loterias não ser alterada, a forma de utilidade esperada não é preservada. Vejamos um exemplo. Sejam x_1 e x_2 os resultados possíveis, U uma função utilidade de Von Neumann-Morgenstern e u uma função utilidade de Bernoulli, temos

$$U(L) = p_1 u(x_1) + p_2 u(x_2). \tag{2.18}$$

Efetuando uma transformação monotônica crescente não-linear sobre U, digamos, $V(L) = e^{U(L)}$, temos:

$$V(L) = e^{[p_1 u(x_1) + p_2 u(x_2)]}. \tag{2.19}$$

Note que V não é Von Neumann-Morgenstern, isto é, não tem a forma de utilidade esperada. Porém, V representa as mesmas preferências que U, pois, dadas duas loterias $L_1, L_2 \in \mathscr{L}$, temos

$$L_1 \succeq L_2 \iff U(L_1) \geq U(L_2) \iff e^{U(L_1)} \geq e^{U(L_2)} \iff V(L_1) \geq V(L_2). \tag{2.20}$$

Assim, como no caso sem incerteza, uma transformação monotônica crescente qualquer não altera o ordenamento das preferências sobre as loterias. Portanto, os mesmos axiomas que fornecem a base para o uso da função U, também fornecem base para o uso da função V.

Exemplo 2.1.3. *Um indivíduo tem renda de* $\$12,00$*. Este indivíduo tem a possibilidade de investir em um ativo sem risco que dá um retorno unitário de* $\$16,00$ *com probabilidade* $0,5$*, e retorno zero com probabilidade* $0,5$*. O preço unitário do ativo é* $\$3,00$*. sua função de utilidade Von Neumann-Morgenstern é* $u(x) = \sqrt{x}$*. Se adquirisse o ativo, qual seria sua utilidade esperada?*

$$\begin{aligned}
\mathbb{E}\left[u(x)\right] &= \frac{1}{2}5 \times (\sqrt{12+16-3}) + \frac{1}{2} \times (\sqrt{12+0-3}) \\
&= \frac{1}{2} \times (\sqrt{25}) + \frac{1}{2} \times (\sqrt{9}) \\
&= \frac{5}{2} + \frac{3}{2}
\end{aligned}$$

$$= 4 \tag{2.21}$$

O valor esperado dessa loteria seria

$$\mathbb{E}[x] = \frac{1}{2}25 + \frac{1}{2}9 = 17 \tag{2.22}$$

A utilidade do valor esperado seria $u\left(\mathbb{E}[x]\right) = u(17) = \sqrt{17} = 4,123.$

2.2 Representação Gráfica e Propriedades da Utilidade Esperada

Se os axiomas da EU são satisfeitos pelo tomador de decisão, seu comportamento de escolha deverá ser particularmente compatível com estas propriedades. Assim, uma discussão à cerca destas propriedades é útil tanto para uma melhor compreensão teórica-matemática dos postulados, bem como para perceber alguns desdobramentos do comportamento dos tomadores de decisão demandado pela teoria.

As quatro propriedades que discutiremos são as seguintes:

1. Linearidade nas probabilidades
2. Separabilidade aditiva
3. Propriedade da razão comum
4. Propriedade da consequência comum

Estas propriedades são tais que a linearidade nas probabilidades (1) implica a separabilidade aditiva (2), que por sua vez implica a propriedade da razão comum (3) e a propriedade da consequência comum (4).

2.2.1 Linearidade nas Probabilidades

A linearidade nas probabilidades é decorrência direta do axioma da independência. Suponha três possíveis resultados x_1, x_2 e x_3 tais que $x_3 > x_2 > x_1$ com probabilidades p_1, p_2 e p_3. Como vimos, definindo $p_2 = 1 - p_1 - p_3$, podemos representar todas as loterias simples no plano (p_1, p_3). Considerando um nível fixo de utilidade esperada $\overline{U}$, podemos isolar p_3 na função utilidade de Von Neumann-Morgenstern. Assim, obtemos a equação da curva de indiferença referente ao nível de utilidade esperada $\overline{U}$.

$$\begin{aligned}
\overline{U} &= p_1 u(x_1) + p_2 u(x_2) + p_2 u(x_3) \\
&= p_1 u(x_1) + (1 - p_1 - p_3) u(x_2) + p_3 u(x_3) \\
&= p_1 u(x_1) + u(x_2) - p_1 u(x_2) - p_3 u(x_2) + p_3 u(x_3).
\end{aligned} \tag{2.23}$$

Resolvendo para p_3, temos:

$$\begin{aligned} p_3u(x_2) - p_3u(x_2) &= [u(x_1) - u(x_2)]p_1 + u(x_2) - \overline{U} \\ p_3 &= \left[\frac{u(x_1) - u(x_2)}{u(x_2) - u(x_3)}\right] p_1 + \frac{u(x_2) - \overline{U}}{u(x_2) - u(x_3)}. \end{aligned} \tag{2.24}$$

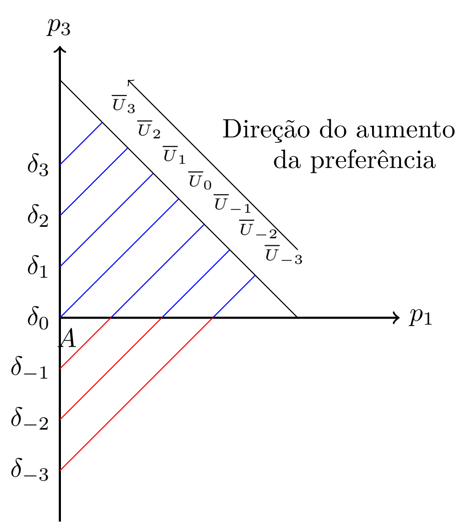
Como todas utilidades são constantes, a equação acima é linear. A curva de indiferença é tanto mais inclinada quanto maior for o "excesso de utilidade" que o prêmio x_2 fornece em relação ao prêmio x_1 e quanto menor for o "excesso de utilidade" que o prêmio x_3 oferece em relação a x_2. Seja e_{21} o "excesso de utilidade" de x_2 em relação a x_1, e_{32} o "excesso de utilidade" de x_3 em relação à x_2 e δ o intercepto da curva de indiferença, temos:

$$\begin{aligned} p_3 &= \frac{e_{21}}{e_{32}} p_1 + \delta \\ &= \phi p_1 + \delta. \end{aligned} \tag{2.25}$$

Na Figura 2.5, as curvas de indiferença são representadas em um triângulo de Marschak-Machina. Cada curva de indiferença é referente a um nível fixo de utilidade esperada. No ponto A, tanto p_1 com p_3 são iguais a zero e, portanto, $p_2 = 1$. Deste modo, $U = p_1u(x_1) + p_2u(x_2) + p_3u(x_3)$ implica $\overline{U}_0 = u(x_2)$. As curvas de indiferença referentes a níveis de utilidade esperada abaixo de $\overline{U}_0$ têm interceptos negativos, mas estes interceptos são "virtuais", já que estão fora do triângulo que demarca a área relevante de análise (as probabilidades atribuídas a p_3 são menores do que zero).

Note que as curvas de indiferença são linhas retas, devido à linearidade nas probabilidades. A inclinação das curvas de indiferença é positiva e independente do nível de utilidade $\overline{U}$. De fato, todas as curvas de indiferença são paralelas, apresentando o mesmo coeficiente de inclinação ϕ.

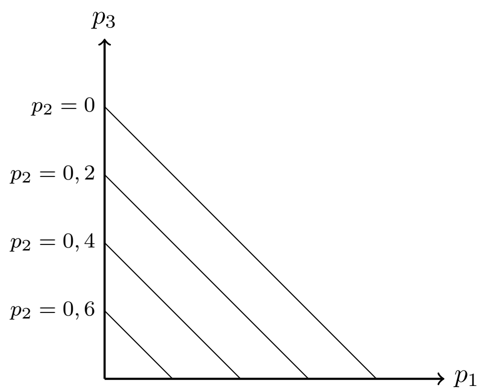
Figura 2.5 – Curvas de indiferença



Uma limitação da representação através do diagrama em triângulo é que ele não permite a visualização direta da probabilidade do resultado 2. Porém, um modo prático de trabalhar com esta limitação é utilizar o que chamaremos de linhas de iso-p_2, como na Figura 2.6. Sob uma linha de iso-p_2, a probabilidade do prêmio 2 se mantém constante. A linha de iso-p_2 é o lugar geométrico de todas as combinações possíveis de p_1 e p_3 para um dado p_2 fixo.

Podemos observar na Figura 2.6 que a linha de iso-p_2 referente a $p_2 = 0$ coincide com a hipotenusa do triângulo. Qualquer loteria sobre esta linha fornece o resultado 2 com probabilidade zero. Note que as linhas de iso-p_2 são combinações convexas entre p_1 e p_3. Assim, por exemplo, sobre a linha referente a $p_2 = 0,2$, sempre temos $p_1 + p_3 = 0,8$. Na medida em que nos deslocamos para a esquerda ou para baixo no espaço cartesiano (p_1, p_3), a probabilidade de p_2 aumenta.

Figura 2.6 – Linhas de iso-p_2



2.2.2 Separabilidade Aditiva

A propriedade da separabilidade aditiva implica que a contribuição de um prêmio e de sua probabilidade para a utilidade esperada da loteria é independente dos outros componentes da loteria. Se um tomador de decisão, por exemplo, se defrontar com a loteria $L = (0,1; 0,2; 0,7)$ referente aos prêmios $\{x_1, x_2, x_3\} = \{1, 50, 200\}$, então a contribuição do prêmio de \$1 com probabilidade $0,1$ para a utilidade esperada da loteria L não depende dos prêmios x_2 e x_3, nem de suas respectivas probabilidades. Esta propriedade exclui a possibilidade de "desapontamento antecipado".

Vejamos um exemplo em que poderia haver "desapontamento antecipado". Seja $\{x_1, x_2, x_3\} = \{\text{não fazer nada}, \text{ler um livro sobre limosines}, \text{ganhar uma limosine}\}$, suponha que o tomador de decisão seja um entusiasta por limosines, de maneira que suas preferências sejam $x_3 \succ x_2 \succ x_1$. Ele pode escolher entre duas loterias referentes aos prêmios x_1, x_2, x_3. A primeira loteria é $L_1 = (0; 0,01; 0,99)$ e a segunda, $L_2 = (0,01; 0; 0,99)$. Ou seja, a primeira loteria fornece "ler um livro sobre limosines" com probabilidade $0,01$ e "ganhar uma limosine" com probabilidade $0,99$; a segunda, fornece "não fazer nada" com probabilidade $0,01$ e "ganhar uma limosine" com probabilidade $0,99$. Neste caso, o axioma da independência implica que a loteria L_1 deve ser escolhida. Por quê?

Seja $L_A = (0; 1; 0)$ e $L_B = (1; 0; 0)$, temos $L_A \succ L_B$ pois $x_2 \succ x_1$. Seja $L_C = (0; 0; 1)$, $L_A \succ L_B$ implica, pelo axioma da independência, $0,01L_A + 0,99L_C \succ 0,01L_B + 0,99L_C$. Como

$L_1 = 0,01L_A + 0,99L_C$ e $L_2 = 0,01L_B + 0,99L_C$, então $L_1 \succ L_2$.

O problema surge devido à possibilidade de o tomador de decisão não ganhar a limosine. Caso isto realmente acontecesse, ele poderia ficar desapontado - afinal de contas, havia 99% de probabilidade de ganhá-la, e não ganhou. Ler um livro sobre limosines após o desapontamento com a "perda" da limosine, que estava praticamente "certa", talvez seja algo bastante indesejável, que trouxesse repetida frustração com a "perda". Assim, o tomador de decisão pode antecipar este possível "desapontamento" e escolher L_2, evitando os possíveis aborrecimentos com a leitura do livro, mesmo que, a princípio, ele preferisse ler o livro a não fazer nada.

2.2.3 Propriedade da Razão Comum

A propriedade da razão comum implica que a preferência entre duas loterias não é afetada se todas as probabilidades em ambas as loterias forem multiplicadas por uma constante $t \in (0,1)$ e a probabilidade remanescente $(1-t)$ for designada para uma consequência comum.

Suponha, por exemplo, que temos $L_1 = (0,2;0;0,8)$ e $L_2 = (0,6;0;0,4)$ referentes a x_1, x_2, x_3 quaisquer, de maneira que $x_3 \succ x_2 \succ x_1$ e $L_1 \succ L_2$. A princípio, estas loterias sorteiam apenas entre x_1 e x_3, pois a probabilidade de x_2 é zero. Se as probabilidades forem multiplicadas por $t = 0,5$ e se a probabilidade remanescente $(1-0,5) = 0,5$ for realocada para a consequência comum x_2, então teremos $L_{t_1} = (0,1;0,5;0,4)$ e $L_{t_2} = (0,3;0,5;0,2)$. Neste caso, a propriedade da razão comum implica $L_{t_1} \succ L_{t_2}$. Por quê?

Seja $L_1 = (0,2;0;0,8)$, $L_2 = (0,6;0;0,4)$, $L_B = (0;1;0)$ e $L_1 \succ L_2$, então, pelo axioma da independência, temos $0,5L_1 + 0,5L_B \succ 0,5L_2 + 0,5L_B$. Como $L_{t_1} = 0,5L_1 + 0,5L_B$ e $L_{t_2} = 0,5L_2 + 0,5L_B$, então $L_{t_1} \succ L_{t_2}$.

2.2.4 Propriedade da Consequência Comum

A propriedade da consequência comum afirma que, dadas duas loterias com uma consequência comum de mesma probabilidade, a substituição desta consequência por outra consequência comum de mesma probabilidade não influencia a preferência entre as loterias.

Assim, suponha, por exemplo, que temos duas loterias $L_1 = \left(\frac{4}{10}, \frac{2}{10}, \frac{4}{10}, 0\right)$ e $L_2 = \left(\frac{6}{10}, \frac{2}{10}, \frac{2}{10}, 0\right)$ referentes a $\{x_1, x_2, x_3, x_4\} = \{1, 50, 200, 70\}$ tal que $L_1 \succ L_2$. Então, se substituirmos x_2 por x_4, de maneira que as loterias fiquem $L_{01} = \left(\frac{4}{10}, 0, \frac{4}{10}, \frac{2}{10}\right)$ e $L_{01} = \left(\frac{6}{10}, 0, \frac{2}{10}, \frac{2}{10}\right)$, L_{01} deve ser preferível a L_{02}. Por quê?

Seja $L_A = (0,1,0,0)$, $L_{1B} = \left(\frac{5}{10}, 0, \frac{5}{10}, 0\right)$ e $L_{2B} = \left(\frac{15}{20}, 0, \frac{5}{20}, 0\right)$. Como $L_1 = \frac{2}{10}L_A + \frac{8}{10}L_{1B}$, $L_2 = \frac{2}{10}L_A + \frac{8}{10}L_{2B}$ e $L_1 \succ L_2$, então $\frac{2}{10}L_A + \frac{8}{10}L_{1B} \succ \frac{2}{10}L_A + \frac{8}{10}L_{2B}$. Pelo axioma da independência, $L_{1B} \succ L_{2B}$. Seja $L_C = (0,0,0,1)$. Pelo axioma da independência, $\frac{2}{10}L_C + \frac{8}{10}L_{1B} \succ \frac{2}{10}L_C + \frac{8}{10}L_{2B}$. Como $\frac{2}{10}L_C + \frac{8}{10}L_{1B} = L_{01}$ e $\frac{2}{10}L_C + \frac{8}{10}L_{2B} = L_{02}$, então $L_{01} \succ L_{02}$.

2.3 Comportamento frente ao Risco

Usualmente, quando tratamos de teorias que envolvam decisões sob o contexto de incerteza, classificamos os tomadores de decisão em três tipos básicos: avessos ao risco, neutros ao risco e propensos ao risco. Neste aspecto, a EU tem um grande atrativo - ela consegue lidar com o conceito de aversão ao risco de uma maneira ágil e bastante simples, associando-o à concavidade da função utilidade de Bernoulli. Bernoulli, em seu artigo de 1738, prontamente supôs uma função logarítmica côncava, capaz de suavizar a progressão dos payoffs estabelecidos pelo Paradoxo de São Petersburgo. Posteriormente, a EU foi estendida de forma a permitir a representação de uma classe mais ampla de preferências que, além da aversão ao risco, incluísse a neutralidade e a propensão ao risco.

Antes de analisarmos as definições de atitudes frente ao risco no contexto da teoria da utilidade esperada, veremos as definições em um contexto geral, que não presumem uma representação de preferências através da EU. Nestas definições, consideraremos EL uma loteria degenerada que fornece o valor esperado da loteria L com certeza. Uma loteria é degenerada quando atribui probabilidade 1 para algum prêmio e 0 para os outros. Por exemplo, a loteria $D = (0, 1, 0, 0)$ é uma loteria degenerada.

Definição 2.3.1 (Aversão ao risco)**.** *Um tomador de decisão é avesso ao risco (ou exibe aversão ao risco) se para qualquer $L \in \mathscr{L}$, $E_L \succ L$. Ou seja, um tomador de decisão é avesso ao risco se, para qualquer loteria L, a loteria degenerada E_L que fornece o mesmo valor esperado que L é fracamente preferível a L. Além disso, dizemos que um tomador de decisão é estritamente avesso ao risco se para qualquer $L \in \mathscr{L}$ não-degenerada, temos $E_L \succ L$. Isto é, um indivíduo é estritamente avesso ao risco se, para qualquer loteria L não-degenerada, a loteria degenerada E_L que fornece o mesmo valor esperado que L é estritamente preferível a L.*

Definição 2.3.2 (Neutralidade ao risco)**.** *Um tomador de decisão é neutro ao risco (ou exibe neutralidade ao risco) se para qualquer $L \in \mathscr{L}$, $E_L \sim L$. Ou seja, um tomador de decisão é neutro ao risco se, para qualquer loteria L, a loteria degenerada E_L que fornece o mesmo valor esperado que L é indiferente a L.*

Definição 2.3.3 (Propensão ao risco)**.** *Um tomador de decisão é propenso ao risco (ou exibe propensão ao risco) se para qualquer $L \in \mathscr{L}$, $L \succeq E_L$. Ou seja, um tomador de decisão é propenso ao risco se qualquer loteria L é fracamente preferível a uma loteria degenerada E_L que fornece o mesmo valor esperado que L. Além disso, dizemos que um tomador de decisão é estritamente propenso ao risco se para qualquer $L \in \mathscr{L}$ não-degenerada, temos $L \succ E_L$. Isto é, um indivíduo é estritamente propenso ao risco se qualquer loteria L não-degenerada é estritamente preferível à loteria degenerada E_L que fornece o mesmo valor esperado que L.*

Note que todo tomador de decisão estritamente avesso ao risco é também avesso ao risco e que todo tomador de decisão estritamente propenso ao risco é também propenso ao risco. Além disso, observe que todo indivíduo neutro ao risco é também, ao mesmo tempo, avesso e propenso ao risco. Ou seja,

$$\text{Aversão ao risco estrita} \;\Rightarrow\; \text{Aversão ao risco} \tag{2.26}$$

$$\text{Propensão ao risco estrita} \;\Rightarrow\; \text{Propensão ao risco} \tag{2.27}$$

$$\text{Neutralidade ao risco} \iff \text{Aversão ao risco e propensão ao risco} \tag{2.28}$$

Porém, note que o contrário não é necessariamente verdadeiro. Quando um indivíduo é neutro ao risco, apesar de ser também avesso e propenso ao risco, ele não é estritamente avesso ou estritamente propenso ao risco.

Exemplo 2.3.1. *Vejamos um exemplo para aplicar as definições de atitudes frente ao risco. Suponha que um tomador de decisão se defronte com as seguintes loterias, referentes a* $\{x_1, x_2, x_3\} = \{0, 40, 100\}$*:*

$$L_1 = (0,6; 0, 0,4) \;\Rightarrow\; E_{L_1} = 40 \tag{2.29}$$

$$E_{L_1} = (0; 1; 0) \;\Rightarrow\; E(E_{L_1}) = 40 \tag{2.30}$$

Note que L_1 *é uma loteria não-degenerada e que* E_{L_1} *é uma loteria degenerada que fornece com certeza o valor esperado de* L_1*. Assim, ambas loterias apresentam o mesmo valor esperado, mas* L_1 *é mais arriscada. Neste caso, conhecendo apenas a atitude do indivíduo frente ao risco, já podemos obter a sua ordenação. Assim, se o tomador de decisão é:*

$$\textit{Estritamente avesso ao risco} \;\Rightarrow\; E_{L_1} \succ L_1 \tag{2.31}$$

$$\textit{Avesso ao risco} \;\Rightarrow\; E_{L_1} \succeq L_1 \tag{2.32}$$

$$\textit{Neutro ao risco} \;\Rightarrow\; E_{L_1} \sim L_1 \tag{2.33}$$

$$\textit{Propensão ao risco} \;\Rightarrow\; L_1 \succeq E_{L_1} \tag{2.34}$$

$$\textit{Estritamente propenso ao risco} \;\Rightarrow\; L_1 \succ E_{L_1} \tag{2.35}$$

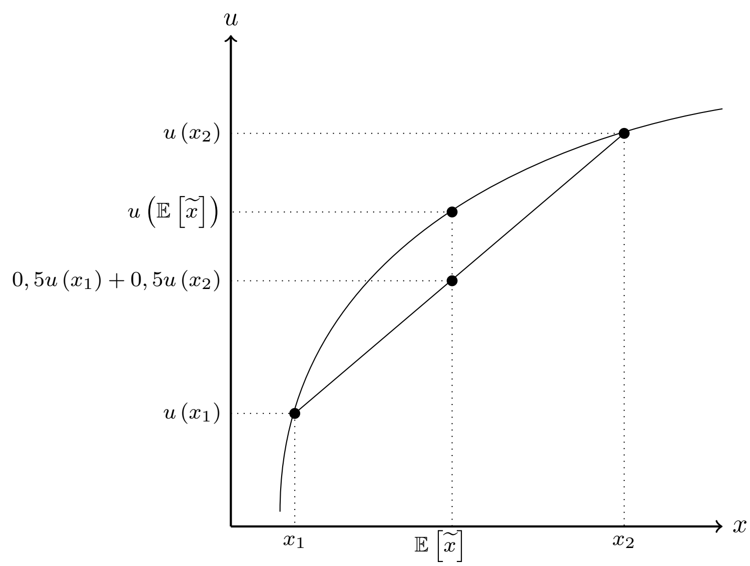
2.4 Função Utilidade de Bernoulli e Atitudes frente ao Risco

No contexto da teoria da utilidade esperada, a aversão ao risco é uma propriedade estabelecida a partir da função utilidade de Bernoulli. Segue direto da definição de aversão ao risco que um tomador de decisão é avesso ao risco se e somente se

$$\mathbb{E}[u(\widetilde{x})] \leq u\left(\mathbb{E}[\widetilde{x}]\right). \tag{2.36}$$

A partir de agora, toda variável com um til em cima, tal como $\widetilde{x}$, é uma variável aleatória. Podemos pensar $\widetilde{x}$ como uma variável aleatória distribuída de acordo com alguma loteria simples L. A desigualdade acima é chamada de desigualdade de Jensen. Ela nos informa que se o tomador de decisão é avesso ao risco, então a utilidade esperada de uma loteria é menor ou igual à utilidade do valor esperado da loteria. Se a desigualdade de Jensen é satisfeita, então, pela própria definição de concavidade, a função utilidade de Bernoulli é côncava. Portanto, sob a teoria da utilidade esperada, a aversão ao risco equivale à concavidade da função utilidade de Bernoulli. Na figura 2.7 representamos uma função utilidade (estritamente) côncava, referente a um indivíduo (estritamente) avesso ao risco.

Figura 2.7 – Aversão estrita ao risco



Se este indivíduo se defrontar com a loteria $L = (0,5; 0,5)$ referente a x_1 e x_2 quaisquer, então a utilidade do valor esperado, $u(\mathbb{E}[L])$ ou $u(\mathbb{E}[\widetilde{x}])$ é superior a utilidade esperada da loteria $U(L) = 0,5u(x_1) + 0,5u(x_2)$ ou $\mathbb{E}[u(\widetilde{x})]$. Em termos intuitivos, o elemento crucial na EU que dá origem ao comportamento de aversão ao risco é a utilidade marginal decrescente. Se a função utilidade é tal que os acréscimos de utilidade caem sistematicamente à medida que x aumenta, então ela representa as preferências de um tomador de decisão (estritamente) avesso ao risco.

Exemplo 2.4.1. *Suponha que a função de utilidade do tipo Von Neumann-Morgenstern de um agente é dada por $u(M) = \ln(M)$. Calculando a segunda derivada com relação a M, temos que:*

$$\frac{du(M)}{dM} = \frac{1}{M} \tag{2.37}$$

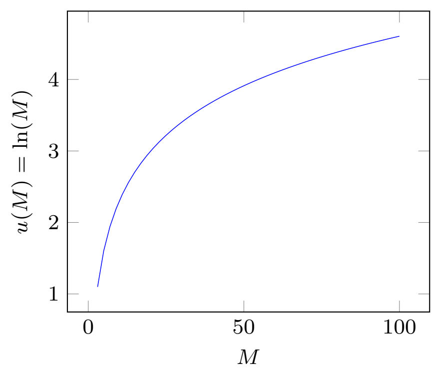
$$\frac{d^2u(M)}{dM^2} = -\frac{1}{M^2} < 0 \tag{2.38}$$

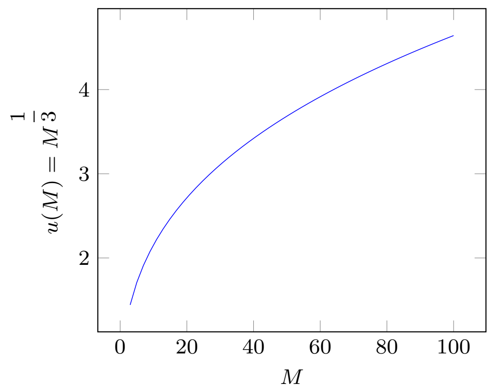
Agora, suponha que a função de utilidade do tipo Von Neumann-Morgenstern de um agente é dada por $u(M) = M^{\frac{1}{3}}$. Calculando a segunda derivada com relação a M, temos que:

$$\frac{du(M)}{dM} = \frac{1}{3}M^{-\frac{2}{3}} \tag{2.39}$$

$$\frac{d^2u(M)}{dM^2} = -\frac{2}{9}M^{-\frac{5}{3}} < 0 \tag{2.40}$$

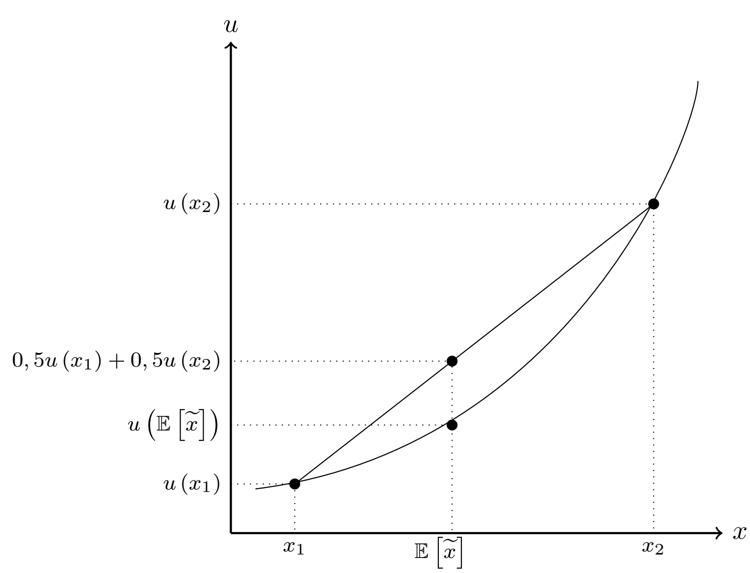
Figura 2.8 – Função de utilidade Von Neumann-Morgenstern





Por outro lado, a propensão ao risco equivale à convexidade da função utilidade de Bernoulli. Na figura 2.9, temos uma função utilidade (estritamente) convexa que representa as preferências de um tomador de decisão (estritamente) propenso ao risco. Neste caso, a utilidade do valor esperado da loteria é inferior a utilidade esperada da loteria. Portanto, quando um indivíduo descrito pela Figura 2.9 se defronta com duas loterias de mesmo valor esperado, ele prefere a loteria mais arriscada.

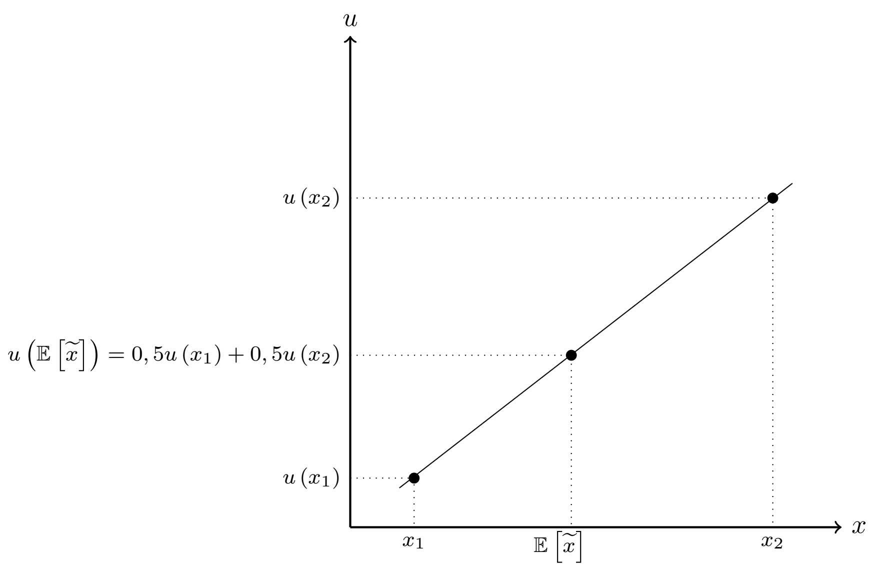
Figura 2.9 – Propensão estrita ao risco



Digamos que $L = (0,5; 0,5)$, $x_1 = 0$ e $x_2 = 4.000.000$, de maneira que $\mathbb{E}[\widetilde{x}] = 2.000.000$. Por que alguém abriria mão de ganhar \$2.000.000 para participar de uma loteria em que há a possibilidade de ganhar \$4.000.000 com probabilidade 0,5 e de não ganhar nada com probabilidade 0? Uma possível explicação é que alguém poderia valorizar tanto os \$4.000.000 em relação aos \$2.000.000, que se disporia a incorrer no risco para tentar os \$4.000.000. Porém, na verdade, não precisamos estabelecer este tipo de justificação. Podemos interpretar o fato de maneira puramente operacional - seja qual for a razão que leve o indivíduo a incorrer no risco, o que importa é que ele prefere incorrer no risco e, portanto, precisamos de uma função utilidade capaz de representar esta sua propensão a incorrer no risco. A característica requerida é exatamente a convexidade estrita. Assim, neste caso, a utilidade marginal é crescente.

Finalmente, se um indivíduo for neutro ao risco, a sua função utilidade de Bernoulli é tanto convexa quanto côncava e, portanto, linear, como mostra a Figura 2.10. Neste caso, o indivíduo é indiferente entre a loteria $L = (0,5; 0,5)$ referente a x_1 e x_2 e a loteria que fornece o valor esperado de L com certeza. Podemos ver na Figura 2.10, que o nível de utilidade da loteria $L = (0,5; 0,5)$ é igual ao nível de utilidade do valor esperado $\mathbb{E}[\widetilde{x}]$.

Figura 2.10 – Neutralidade ao risco



Assim, podemos sumarizar a relação entre as atitudes frente ao risco e o formato das funções utilidade de Bernoulli do seguinte modo:

$$\text{Aversão ao risco} \;\Leftrightarrow\; u''(x) \leq 0 \;\forall x \text{ (UMg decrescente)} \;\Leftrightarrow\; U \text{ convexa} \tag{2.41}$$

$$\text{Propensão ao risco} \;\Leftrightarrow\; u''(x) \geq 0 \;\forall x \text{ (UMg crescente)} \;\Leftrightarrow\; U \text{ côncava} \tag{2.42}$$

$$\text{Neutralidade ao risco} \;\Leftrightarrow\; u''(x) = 0 \;\forall x \text{ (UMg constante)} \;\Leftrightarrow\; U \text{ linear} \tag{2.43}$$

Exemplo 2.4.2. *Um indivíduo tem renda de \$12,00. Este indivíduo tem a possibilidade de investir* *em um ativo sem risco que dá um retorno unitário de* $\$16,00$ *com probabilidade* $0,5$*, e retorno zero com probabilidade* $0,5$*. O preço unitário do ativo é* $\$3,00$*. sua função de utilidade Von Neumann-Morgenstern é* $u(x) = \sqrt{x}$*. Se adquirisse o ativo, qual seria sua utilidade esperada?*

$$\begin{aligned} \mathbb{E}\left[u(x)\right] &= \frac{1}{2}5 \times (\sqrt{12+16-3}) + \frac{1}{2} \times (\sqrt{12+0-3}) \\ &= 4 \end{aligned} \tag{2.44}$$

O valor esperado dessa loteria seria

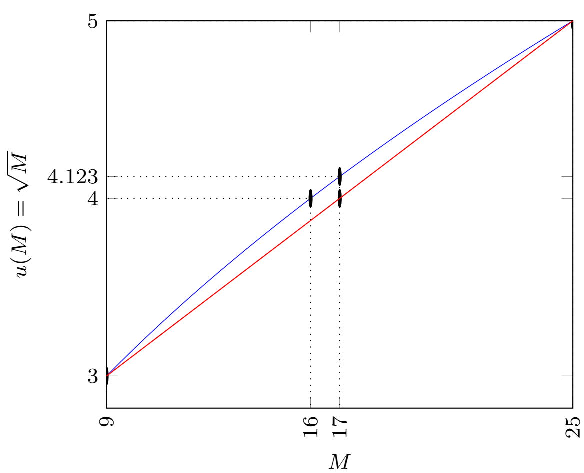
$$\mathbb{E}[x] = \frac{1}{2}25 + \frac{1}{2}9 = 17 \tag{2.45}$$

A utilidade do valor esperado seria

$$u\left(\mathbb{E}[x]\right) = u(17) = \sqrt{17} = 4,123 \tag{2.46}$$

Graficamente, temos a seguinte situação:

Figura 2.11 – Função de utilidade Von Neumann-Morgenstern $u(M) = \sqrt{M}$



2.5 Medidas de Aversão ao Risco

Uma questão importante que surge é como medir o grau de aversão ao risco de diferentes tomadores de decisão. Esta questão foi fundamentalmente tratada em Pratt (1964). Nesse artigo, o autor introduziu algumas medidas de aversão ao risco que, mais tarde, ficaram conhecidas como os coeficientes de aversão ao risco de Arrow-Pratt. Os coeficientes levaram o nome também de Kennett Arrow porque quando John Pratt publicou seu artigo, Arrow já tinha esboçado algumas análises sobre a função que Pratt tomou como medida de aversão ao risco.

O artigo de Pratt, "Risk aversion in the small and in the large", foi publicado na edição de janeiro-abril de 1964 da Econometrica e deu muitas contribuições importantes para a teoria da utilidade esperada - além das medidas, introduziu teoremas e análises sobre a aversão ao risco. Apesar do tempo passado desde a sua publicação, o artigo continua uma referência importante no campo de medidas de aversão ao risco.

2.5.1 Prêmio de Risco e Equivalente-Certeza

Como vimos, um tomador de decisão avesso ao risco se dispõe a reduzir em certa quantia o valor esperado de uma loteria em troca de uma redução no risco. Assim, poderíamos fazer a seguinte pergunta: quanto um indivíduo avesso ao risco estaria disposto a pagar pela eliminação completa de um risco? Esta questão nos remete ao conceito de prêmio de risco.

O prêmio de risco (π) é definido como a quantia máxima que um indivíduo está disposto a pagar para evitar determinado risco, digamos, $\widetilde{g}$. O prêmio de risco de um indivíduo é tal que deixa o indivíduo indiferente entre receber um risco $\widetilde{g}$ e receber uma quantia certa $\mathbb{E}[\widetilde{g}] - \pi$. Assim, o prêmio de risco pode ser calculado através da seguinte equação:

$$\mathbb{E}[u(x + \widetilde{g})] = u\left[x + \mathbb{E}[\widetilde{g}] - \pi(x, u, \widetilde{g})\right]. \tag{2.47}$$

em que x são os ativos ou a riqueza do indivíduo. A equação acima afirma que a utilidade esperada da riqueza mais o risco $\widetilde{g}$ deve ser igual a utilidade da soma: [riqueza do indivíduo] + [valor esperado do risco $\widetilde{g}$] [prêmio de risco para se livrar de $\widetilde{g}$]. Assim, o prêmio de risco é uma função dos parâmetros que aparecem na equação acima: $\pi = \pi(x, u, \widetilde{g})$.

Para o prêmio de risco ser um indicador adequado de aversão ao risco, o prêmio de risco deve aumentar com o aumento da aversão ao risco. Assim, um tomador de decisão mais avesso ao risco deve estar disposto a pagar mais para evitar um risco dado.

Seja $\widetilde{z}$ um risco puro, com $\mathbb{E}[\widetilde{z}] = 0$. Com isso, seja a seguinte proposição:

Proposição 2.5.1. *O tomador de decisão 2 é mais avesso ao risco do que o tomador de decisão 1 se e somente se o tomador de decisão 2 está sempre disposto a pagar um valor igual ou maior do que o indivíduo 1 para evitar determinado risco; isto é, se e somente se $\pi(x, u_2, \widetilde{z}) \leq \pi(x, u_1, \widetilde{z})$ para quaisquer x e $\widetilde{z}$.*

Um conceito importante relacionado ao prêmio de risco é o equivalente-certeza. O equivalente-certeza ou equivalente-certo $C(x, u, \widetilde{g})$ é quantia em dinheiro que deixa o indivíduo indiferente entre a loteria representada pela variável aleatória $\widetilde{g} = \mu + \widetilde{z}$ com $\mathbb{E}[\widetilde{z}] = 0$ e a quantia $C(x, u, \widetilde{g})$. Isto é,

$$u\left(x + C(x, u, \widetilde{g})\right) = u(x + \widetilde{g}). \tag{2.48}$$

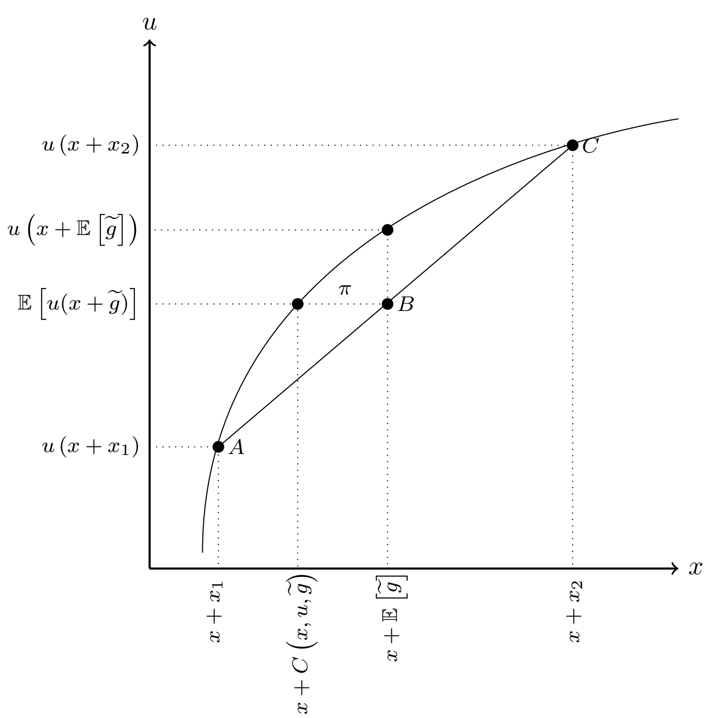
Pode-se mostrar que se a função utilidade u é contínua, então cada loteria tem pelo menos um equivalente-certeza. Além disso, se u é estritamente crescente, cada loteria tem no máximo um equivalente-certeza. Alternativamente, o equivalente-certeza de uma loteria representada pela variável aleatória $\widetilde{g}$ é igual ao valor esperado da loteria menos o prêmio de risco.

Portanto, como um aumento na aversão ao risco faz com que o prêmio de risco π aumente, podemos verificar que um aumento na aversão ao risco reduz o equivalente-certeza. Assim, podemos escrever a proposição abaixo.

Proposição 2.5.2. *O tomador de decisão 2 é mais avesso ao risco do que o tomador de decisão 1 se e somente se* $C(x, u_1, \widetilde{g}) \leq C(x, u_2, \widetilde{g})$ *para qualquer* x *e* $\widetilde{g}$.

Na Figura 2.12, representamos os conceitos de equivalente-certeza e prêmio de risco. A Figura mostra o gráfico da função utilidade de Bernoulli de um indivíduo estritamente avesso ao risco que se defronta com uma loteria L (ou um risco $\widetilde{g}$) que fornece x_1 com probabilidade $0,5$ e x_2 com probabilidade $0,5$. O valor esperado desta loteria é $\mathbb{E}[\widetilde{g}]$. Mas como o indivíduo é estritamente avesso ao risco, ele está disposto a pagar um prêmio de risco π para não incorrer no risco.

Figura 2.12 – Equivalente-certeza e prêmio de risco



Note que a utilidade esperada do indivíduo com a loteria é igual a utilidade referente ao ponto médio B da combinação convexa dos pontos A e C. Assim, o nosso problema é achar a quantia certa em dinheiro que fornece a mesma utilidade esperada que este ponto médio. Este é um problema simples - a quantia certa em dinheiro que fornece a utilidade esperada $\mathbb{E}[u(x + \widetilde{g})]$ é exatamente $x + C(x, u, \widetilde{g})$, valor que pode ser obtido através da função utilidade de Bernoulli, utilizando (2.48).

Portanto, como $x + \mathbb{E}[\widetilde{g}] > \mathbb{E}[x + C(x, u, \widetilde{g})]$, o indivíduo representado no gráfico está disposto a abrir mão de parte do valor esperado de sua riqueza para se livrar do risco. Para qualquer riqueza certa acima de $x + C(x, u, \widetilde{g})$, o indivíduo prefere a riqueza certa. Por outro lado, para qualquer riqueza certa abaixo de $x + C(x, u, \widetilde{g})$, o indivíduo prefere que sua riqueza seja modificada pela loteria L. Portanto, o indivíduo está disposto a abrir mão de até $\pi(x, u, \widetilde{g}) = x + \mathbb{E}[\widetilde{g}] - [x + C(x, u, \widetilde{g})] = \mathbb{E}[\widetilde{g}] - C(x, u, \widetilde{g})$ para se livrar do risco. Assim, π é o prêmio de risco, o valor máximo que o indivíduo está disposto a pagar para se livrar do risco.

Exemplo 2.5.1. *Um consumidor tem uma função de utilidade de Von Neumann-Morgenstern representada por $u(x) = \log_2(x)$. Ele possui uma riqueza inicial de \$ 128 e participará de uma loteria que pagará \$ 384 com probabilidade $\frac{1}{2}$ e \$ 0 com probabilidade $\frac{1}{2}$. O menor valor que o consumidor estaria disposto a receber em troca do bilhete é de 2^β. Qual o valor de β?*

Dada sua renda inicial de $w_0 = 128$, se o indivíduo participa da loteria, terá o seguinte payoff:

$$
\begin{aligned}
\$384+\$128 &= \$512 \text{ com probabilidade } \frac{1}{2}\\
\$0+\$128 &= \$128 \text{ com probabilidade } \frac{1}{2}
\end{aligned}
$$

A utilidade esperada da riqueza é, portanto,

$$
\begin{aligned}
\mathbb{E}[u(x)] &= p_1 u(c_1) + p_2 u(c_2)\\
&= \left(\frac{1}{2}\right)\log_2(512) + \left(\frac{1}{2}\right)\log_2(128)\\
&= \left(\frac{1}{2}\right)9 + \left(\frac{1}{2}\right)7\\
&= 8
\end{aligned}
\tag{2.49}
$$

Para encontrar o equivalente-certeza basta igualar o valor da utilidade esperada de Von Neumann-Morgenstern à função de utilidade do consumidor, da seguinte forma:

$$
\begin{aligned}
\log_2(x_{EC}) &= \mathbb{E}[u(x)]\\
\log_2(x_{EC}) &= 8\\
x_{EC} &= 2^8
\end{aligned}
\tag{2.50}
$$

Exemplo 2.5.2. *Um indivíduo tem riqueza não nula e sua função de utilidade de Bernoulli tem a forma funcional $u(x) = k - \frac{a}{x}$, em que a e k são constantes positivas e $x > \frac{a}{k}$. Este indivíduo é convidado a participar de uma loteria que triplica sua riqueza com probabilidade p e a reduz à terça parte com probabilidade $(1-p)$. Qual deve ser o valor mínimo de p para que o indivíduo aceite participar da loteria?*

Dado w_0, o indivíduo tem a seguinte loteria:

$$\begin{aligned} 3w_0 \quad & \textit{com probbailidade } p \\ \frac{w_0}{3} \quad & \textit{com probbailidade } 1-p \end{aligned}$$

A função de utilidade em uma situação sem risco é:

$$u(w) = K - \frac{a}{w_0} \tag{2.51}$$

A função de utilidade em uma situação com risco é:

$$\mathbb{E}[u(w)] = p\left[K - \frac{a}{3w_0}\right] + (1-p)\left[K - \frac{3a}{w_0}\right] \tag{2.52}$$

A condição de não-arbitragem é $u(w) = \mathbb{E}[u(w)]$. *Logo,*

$$\begin{aligned} K - \frac{a}{w_0} &= p\left[K - \frac{a}{3w_0}\right] + (1-p)\left[K - \frac{3a}{w_0}\right] \\ K - \frac{a}{w_0} &= pK - \frac{pa}{3w_0} + K - \frac{3a}{w_0} - pK + \frac{p3a}{w_0} \\ -1 &= -\frac{p}{3} - 3 + 3p \\ p &= 0,75 \end{aligned} \tag{2.53}$$

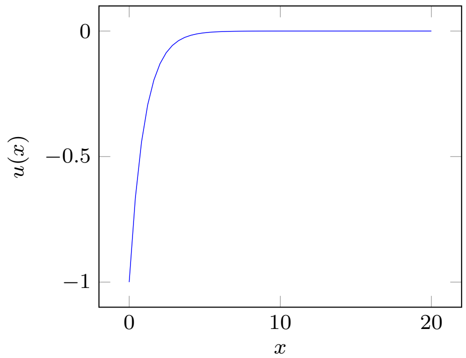
2.5.2 Medida de Aversão ao Risco de Arrow-Pratt

A nossa questão agora é desenvolver uma maneira de medir o grau de aversão ao risco de um tomador de decisão. Em um primeiro momento, poderia parecer adequado medir a curvatura de u através de sua segunda derivada u'', já que, quanto mais côncava for a função utilidade de Bernoulli, mais avesso ao risco é o tomador de decisão.

Porém, a segunda derivada da função utilidade de Bernoulli não é uma medida apropriada, pois não é invariante a transformações lineares positivas da função utilidade. Quando fazemos uma transformação linear positiva em uma função utilidade, ela continua representando as mesmas preferências. Assim, uma medida apropriada de aversão ao risco não deve ser sensível a transformações lineares positivas.

Imagine um indivíduo maximizador de utilidade esperada com função utilidade de Bernoulli $u(x) = -e^{-x}$. Na Figura 2.13 temos o gráfico desta função. Na medida em que $x \to \infty$, tanto $u'(x) \to 0$ como $u''(x) \to 0$. Assim, enquanto x vai crescendo, esta função vai se aproximando cada vez mais da assíntota $u = 0$, parecendo graficamente cada vez menos côncava.

Figura 2.13 – Gráfico da função $u(x) = -e^{-x}$



Porém, o comportamento frente ao risco representado por u é o mesmo para todo o x, já que

$$u(k+x) = -e^{-k-x} \quad \Longrightarrow \quad u(k+x) = e^{-k}(-e^{-x}). \tag{2.54}$$

Fazendo $b = e^{-k}$, temos:

$$u(k+x) = b(-e^{-x}). \tag{2.55}$$

Como $u(x) = -e^{-x}$, temos:

$$u(k+x) = bu(x). \tag{2.56}$$

Portanto, $u(k+x)$ é uma transformação linear positiva de $u(x)$. Assim, qualquer que seja k, não haverá mudança do comportamento frente ao risco. Deste modo, apesar da aparência do gráfico, a função $u(x) = -e^{-x}$ está tão longe de implicar comportamento neutro ao risco em $x = \infty$ ou em $x = 0$.

Uma característica de $u''(x)$ que tem significado em relação à atitude frente ao risco é o seu sinal. Um sinal negativo significa aversão ao risco e um sinal positivo significa propensão ao risco. Porém, a magnitude absoluta de $u''(x)$ não tem qualquer significado para a teoria da utilidade esperada. Apesar disso, podemos obter uma medida de aversão ao risco, normalizando $u''(x)$, dividindo pela derivada primeira $u'(x)$. Assim, obtemos a medida conhecida por coeficiente de aversão ao risco (absoluto) de Arrow-Pratt.

Definição 2.5.1 (Coeficiente de Aversão ao Risco Absoluto de Arrow-Pratt)**.** *Dado uma função utilidade u (duas vezes continuamente diferenciável, côncava e estritamente crescente)*

$$r(x) = -\frac{u''(x)}{u'(x)}, \tag{2.57}$$

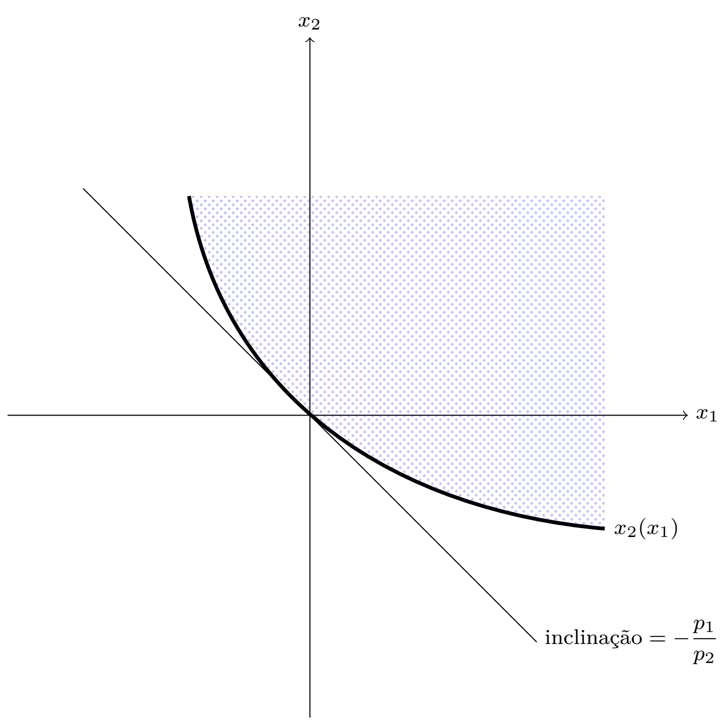
é o coeficiente de aversão ao risco (absoluto) de Arrow-Pratt.

Note que a medida de aversão ao risco absoluto de Arrow-Pratt caracteriza completamente o comportamento do tomador de decisão. A função $r(x)$ apresenta toda informação necessária para prever o comportamento de escolha do tomador de decisão. Observe que a função utilidade de Bernoulli u pode ser recuperada de $r(x)$, integrando duas vezes. As constantes de integração são irrelevantes, já que transformações lineares positivas não alteram a representação de preferências.

Mas qual é a justificativa para o uso da função $r(x)$ como uma medida de aversão ao risco? Para justificar a medida de Arrow-Pratt, tomemos uma aposta com dois resultados possíveis x_1 e x_2 com probabilidades p_1 e p_2.

Para efetuar a análise, utilizaremos o conceito de conjunto de aceitação. O conjunto de aceitação de um tomador de decisão é o conjunto de todas as apostas ou loterias que o tomador de decisão aceitaria em um nível de riqueza x. Na Figura 2.14 representamos o conjunto de aceitação. Note que no ponto de origem do gráfico, a riqueza do indivíduo permanece inalterada, já que, neste caso, $x_1 = 0$ e $x_2 = 0$ e, portanto, $x + x_1$ ou $x + x_2$ são iguais a x.

Figura 2.14 – Conjunto de aceitação



A fronteira deste conjunto - a curva mais escura - é o conjunto (ou curva) de loterias indiferentes e pode ser dado pela função implícita $x_2(x_1)$, como mostra a Figura 2.14. Se o tomador de decisão for avesso ao risco, então o conjunto de aceitação será um conjunto convexo.

Se u é contínua e estritamente crescente, então $x_2(x_1)$ deve satisfazer a seguinte identidade:

$$u(x) \equiv p_1 u(x + x_1) + p_2 u(x + x_2). \tag{2.58}$$

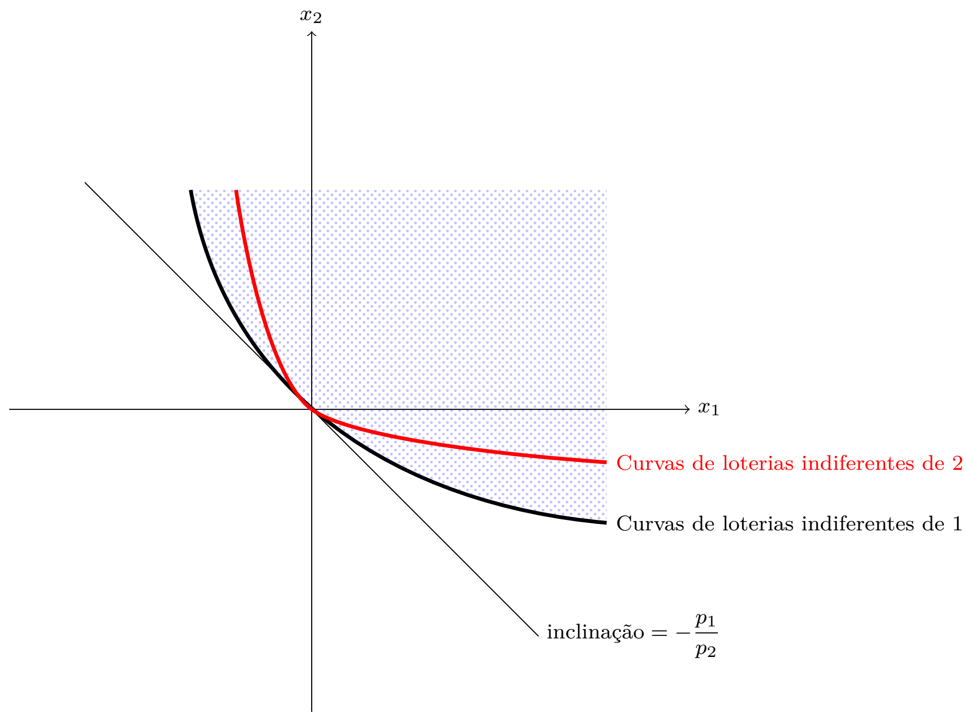
A identidade acima é a curva de loterias indiferentes para uma dada riqueza inicial x. Note que a identidade requer que $x_2(x_1)$ seja o valor que mantenha o nível de utilidade esperada constante na medida em que x_1 varia. Isto é, $x_2(x_1)$ deve ser tal que mantenha o nível de utilidade $u(x)$ referente à riqueza inicial x.

Podemos então, diferenciar a identidade acima com respeito a x_1 e avaliá-la em $x_1 = 0$. Assim, obteremos a inclinação da fronteira do conjunto de aceitação em $x_1 = 0$:

$$\begin{aligned} p_1 u'(x + x_1) + p_2 u'(x + x_2(x_1)) x_2'(x_1) &= 0 \\ p_1 u'(x) + p_2 u'(x) x_2'(0) &= 0 \\ x_2'(0) &= -\frac{p_1}{p_2}. \end{aligned} \tag{2.59}$$

Agora, vamos supor dois tomadores de decisão com probabilidades idênticas para os dois resultados possíveis. É natural afirmar que o tomador de decisão 1 é mais avesso ao risco do que o tomador de decisão 2, ao nível de riqueza x, se o conjunto de aceitação do tomador de decisão 1 está contido no conjunto de aceitação do tomador de decisão 2. Observe a Figura 2.15 - qualquer loteria que faz parte do conjunto de aceitação do tomador de decisão 1 também faz parte do conjunto de aceitação do tomador de decisão 2. Assim, o tomador de decisão 2 aceitará qualquer aposta que o tomador de decisão 1 aceita. Se nos limitarmos às apostas pequenas (apostas com resultados tendendo a zero), podemos efetuar uma análise mais útil ainda.

Figura 2.15 – Comparação de aversão ao risco através do conjunto de aceitação



Podemos perceber que o indivíduo 1 é localmente mais avesso ao risco do que o indivíduo 2 se o conjunto de aceitação do indivíduo 1 é "mais curvado" do que o conjunto de aceitação do indivíduo 2, na vizinhança do ponto $(0, 0)$. De fato, podemos observar que quanto "mais curvado" for o conjunto de aceitação de um indivíduo, maior será o valor que o indivíduo requererá para o resultado positivo (note que da forma como o problema foi montado, sempre haverá um possível resultado negativo e outro positivo) para contrabalançar o risco de ocorrer o resultado negativo.

Assim, podemos obter a curvatura do conjunto de aceitação na vizinhança de $(0, 0)$ diferenciando $p_1u'(x+x_1)+p_2u'(x+x_2(x_1))x_2'(x_1) = 0$ com respeito a x_1 e avaliando a derivação resultante em $x_1 = 0$. Note que $x_2(x_1) = 0$ quando $x_1 = 0$. Assim,

$$
\begin{aligned}
p_1u''(x + x_1) + p_2u''(x + x_2(x_1))x_2'(x_1)^2 + p_2u'(x + x_2(x_1))x_2''(x_1) &= 0 \\
p_1u''(x) + p_2u''(x)x_2'(0)^2 + p_2u'(x)x_2''(0) &= 0 \\
p_1u''(x) + p_2u''(x)\left(\frac{p_1^2}{p_2^2}\right) + p_2u'(x)x_2''(0) &= 0 \quad \left[x_2'(0) = -\frac{p_1}{p_2}\right] \\
p_2u'(x)x_2''(0) &= -p_1u''(x) - p_2u''(x)\left(\frac{p_1^2}{p_2^2}\right) \\
p_2u'(x)x_2''(0) &= -u''(x)\left(p + \frac{p_1^2}{p_2^2}\right) \\
p_2u'(x)x_2''(0) &= -u''(x)\left(\frac{p_1 - p_1^2 + p_1^2}{p_2}\right)
\end{aligned}
$$

$$x_2''(0) = \frac{p_1}{p_2^2}\left[-\frac{u''(x)}{u'(x)}\right]. \tag{2.60}$$

Da equação acima, podemos perceber que a curvatura do conjunto de aceitação é proporcional à medida de aversão ao risco absoluto de Arrow-Pratt. Assim, temos:

$$x_2''(0) = \frac{p_1}{p_2^2} r(x). \tag{2.61}$$

Fazendo $A = \frac{p_1}{p_2^2}$, temos que A é constante, pois p_1 e p_2 são constantes. Logo,

$$x_2''(0) = Ar(x). \tag{2.62}$$

Portanto, a curvatura da fronteira do conjunto de aceitação em $(0,0)$ é diretamente proporcional ao coeficiente de aversão ao risco absoluto de Arrow-Pratt. Assim, podemos fazer a seguinte esquematização: *o indivíduo* 1 *é localmente mais avesso ao risco do que o indivíduo* 2 $\Longleftrightarrow$ *o conjunto de aceitação do indivíduo* 1 *é mais curvado do que o conjunto de aceitação do indivíduo* 2 *na vizinhança de* $(0,0) \Longleftrightarrow r_1(x) \geq r_2(x)$.

Assim, o indivíduo 2 aceitará mais pequenas apostas do que o indivíduo 1 se e somente se o indivíduo 1 tem um maior coeficiente de aversão ao risco absoluto de Arrow-Pratt. Portanto, a medida de Arrow-Pratt está justificada.

Acima, tratamos de conceitos relacionados à aversão ao risco absoluto; ou seja, trabalhamos com conceitos desenvolvidos para lidar com loterias que apresentam resultados que são ganhos ou perdas absolutas em relação ao nível corrente de riqueza.

Agora, iremos tratar de loterias cujos ganhos ou perdas são valores percentuais (relativos) em relação ao nível corrente de riqueza. Por exemplo, digamos que um indivíduo com riqueza inicial x enfrente um risco de ganhar ou perder 20% de sua riqueza. Portanto, neste caso, o valor em termos absolutos que o indivíduo irá ganhar ou perder, dependerá de seu nível de riqueza inicial, já que a aposta é feita em termos relativos à sua riqueza inicial. Se o indivíduo possuir \$1000, ele poderá ganhar ou perder \$200, terminando com \$1200 ou \$800. Se o indivíduo possuir \$3000, ele poderá ganhar ou perder \$600, ficando com \$3600 ou \$2400.

Definição 2.5.2 (Coeficiente de Aversão ao Risco Relativo de Arrow-Pratt)**.** *Dado uma função utilidade u (duas vezes continuamente diferenciável, côncava e estritamente crescente)*

$$r(x) = -x\frac{u''(x)}{u'(x)}, \tag{2.63}$$

é o coeficiente de aversão ao risco (relativo) de Arrow-Pratt.

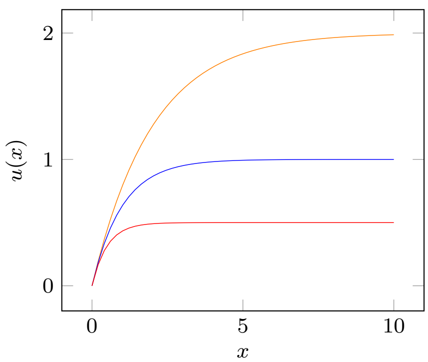
Exemplo 2.5.3. *Considere um indivíduo cuja função de utilidade do tipo Bernoulli é dada por* $u(x) = \sqrt{x}$. *O coeficiente de aversão relativo ao risco será dado por*

$$r(x) = -x\frac{u''(x)}{u'(x)} = -x\frac{-\frac{1}{4}x^{-\frac{3}{2}}}{\frac{1}{2}x^{-\frac{1}{2}}} = \frac{1}{2} \tag{2.64}$$

Exemplo 2.5.4. *Considere um indivíduo cuja função de utilidade do tipo Bernoulli é dada por* $u(w) = \dfrac{1-\exp(-aw)}{a}$. *O coeficiente de aversão relativo ao risco será dado por*

$$r(x) = -w\frac{u''(w)}{u'(w)} = -w\frac{-a\exp(-aw)}{\exp(-aw)} = aw \tag{2.65}$$

Figura 2.16 – Gráfico da função $u(x) = -e^{-x}$



Referências Bibliográficas

[Arrow1971] Arrow, K. J. (1971). The Theory of Risk Aversion. *Essays in the Theory of Risk-Bearing*, pages 90–120.

[Debreu1959] Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Yale University Press.

[Diamond and Rothschild2014] Diamond, P. and Rothschild, M. (2014). *Uncertainty in Economics: Readings and Exercises*. Academic Press.

[Dreze and Modigliani1975] Dreze, J. H. and Modigliani, F. (1975). Consumption Decisions under Uncertainty. In *Stochastic Optimization Models in Finance*, pages 459–486. Elsevier.

[Ebert and Wiesen2014] Ebert, S. and Wiesen, D. (2014). Joint Measurement of Risk Aversion, Prudence, and Temperance. *Journal of Risk and Uncertainty*, 48(3):231–252.

[Eeckhoudt and Gollier2005] Eeckhoudt, L. and Gollier, C. (2005). The Impact of Prudence on Optimal Prevention. *Economic Theory*, 26(4):989–994.

[Eeckhoudt et al.1996] Eeckhoudt, L., Gollier, C., and Schlesinger, H. (1996). Changes in Background Risk and Risk Taking Behavior. *Econometrica*, 64(3):683–689.

[Eeckhoudt and Schlesinger2006] Eeckhoudt, L. and Schlesinger, H. (2006). Putting Risk in Its Proper Place. *American Economic Review*, 96(1):280–289.

[Gilboa2009] Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge university press.

[Gilboa and Schmeidler1995] Gilboa, I. and Schmeidler, D. (1995). Case-Based Decision Theory. *Quarterly Journal of Economics*, 110(3):605–639.

[Hirshleifer and Riley1992] Hirshleifer, J. and Riley, J. G. (1992). *The Analytics of Uncertainty and Information*. Cambridge University Press.

[Jehle2001] Jehle, G. A. (2001). *Advanced Microeconomic Theory*. Pearson Education India.

[Kreps2018] Kreps, D. (2018). *Notes on the Theory of Choice*. Routledge.

[Machina et al.1988] Machina, M. J. et al. (1988). Choice under Uncertainty: Problems Solved and Unsolved: Response. *Journal of Economic Perspectives*, 2(2):181–183.

[Von Neumann et al.2007] Von Neumann, J., Morgenstern, O., and Kuhn, H. W. (2007). *Theory of Games and Economic Behavior*. Princeton University Press.

Capítulo 3

Teoria da Firma

Contents

3.1 Introdução

A atividade econômica não envolve apenas o consumo, mas também a produção e o comércio. A produção deve ser interpretada de forma muito ampla, para incluir a produção de bens físicos - como arroz ou automóveis - e serviços - como assistência médica ou serviços financeiros. Uma empresa pode ser caracterizada por muitos fatores e aspectos, como setores, escala de produção, propriedades, estruturas organizacionais, etc. Mas quais são as características mais importantes para estudarmos o comportamento do produtor ao fazer escolhas? Para compreender as características mais importantes no estudo do comportamento do produtor e nas escolhas da moderna teoria do produtor, supõe-se que a principal característica de uma empresa seja o conjunto de produção. A característica do produtor, juntamente com a suposição de comportamento, são blocos de construção em qualquer modelo da teoria do produtor. O conjunto de produção representa o conjunto de todos os planos de produção tecnologicamente viáveis. A suposição de comportamento expressa o princípio orientador que o produtor usa para fazer escolhas. É geralmente assumido que o produtor procura identificar e selecionar uma produção que seja mais lucrativa. Apresentaremos primeiro um quadro geral de tecnologia de produção. Por si só, o framework não descreve como as escolhas de produção são feitas. Apenas especifica a característica básica de uma empresa que define quais escolhas podem ser feitas; não especifica quais escolhas devem ser feitas. Em seguida, discutiremos quais escolhas devem ser feitas com base nas suposições de comportamento das empresas. Um pressuposto de comportamento básico sobre os produtores é a maximização do lucro. Depois disso, descreveremos as possibilidades de produção em termos físicos, que são reformulados em termos econômicos - usando funções de custo.

3.2 Tecnologia de Produção

A produção é o processo de transformar insumos em produtos. Normalmente, os insumos consistem em mão de obra, equipamento de capital, matérias-primas e bens intermediários adquiridos de outras empresas. Os produtos consistem em produtos acabados ou serviço, ou bens intermediários a serem vendidos a outras empresas. Muitas vezes, métodos alternativos estão disponíveis para produzir o mesmo produto, usando diferentes combinações de insumos. Uma empresa produz produtos de várias combinações de insumos. Para estudar as escolhas das firmas, precisamos de uma maneira conveniente de resumir as possibilidades de produção da empresa, ou seja, quais combinações de insumos e produtos são tecnologicamente viáveis.

3.2.1 Mensurando Insumos e Produtos

Geralmente é mais satisfatório pensar nos insumos e produtos como sendo medidas em termos de fluxos: uma certa quantidade de insumos por período é usada para produzir uma certa quantidade de produtos por unidade no período em algum local. É uma boa ideia incluir explicitamente as dimensões de hora e local em uma especificação de insumos e produtos[1]. O nível de detalhe que usaremos na especificação de insumos e produtos dependerá do problema em questão, mas devemos

[1] Shepahrd (1953) destaca a importância desse aspecto em seu livro.

estar cientes do fato de que um determinado insumo ou produto pode ser especificado em detalhes arbitrariamente precisos. No entanto, quando discutimos as escolhas tecnológicas em abstrato, como fazemos neste capítulo, é comum omitir as dimensões de tempo e de localização.

3.2.2 Especificação da Tecnologia

A realidade fundamental que as empresas devem enfrentar nesse processo é a viabilidade tecnológica. O estado da tecnologia determina e restringe o que é possível ao combinar insumos para produzir produtos, e há várias maneiras de representar essa restrição. A maneira mais geral é pensar na firma como possuindo uma possibilidade de produção. Suponha que a empresa tenha L bens possíveis para servir como insumos e/ou produtos. Se uma empresa usa unidades y_j^i de um bem j como insumo e produz y_j^o do bem como produto, então o produto líquido do bem j é dada por $y_j = y_j^o - y_j^i$.

Um plano de produção é simplesmente uma lista de produtos líquidos de vários bens. Podemos representar um plano de produção por um vetor y em $\mathbb{R}^L$ em que y_j é negativo se o bem j serve como um insumo líquido e positivo se o bem j servir como produto líquido. O conjunto de todos os planos de produção tecnologicamente viáveis é chamado de conjunto de possibilidades de produção da empresa e será denotado por Y, um subconjunto de $\mathbb{R}^L$. O conjunto Y deve descrever todos os padrões de insumos e produtos tecnologicamente viáveis. Nos dá uma descrição completa das possibilidades tecnológicas que a empresa enfrenta.

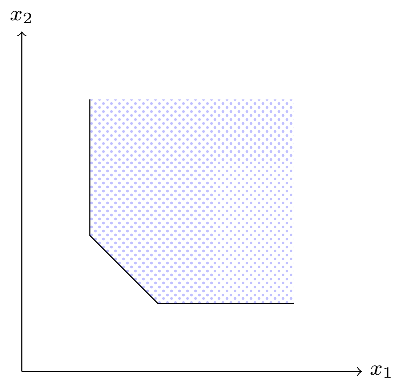
Quando estudamos o comportamento de uma empresa em determinados ambientes econômicos, podemos querer distinguir entre planos de produção que são "imediatamente viáveis"e aqueles que são "eventualmente" viáveis. Geralmente, assumimos que tais restrições podem ser descritas por algum vetor z em $\mathbb{R}^L$. As possibilidades de produção restritas ou de curto prazo serão denotadas por $Y(z)$; isso consiste em todas as cestas de produtos líquidos compatíveis com o nível de restrição z. A seguir estão alguns exemplos de tais restrições.

Exemplo 3.2.1 (Conjunto de Requerimento de Insumos)**.** *Suponha que uma empresa produza apenas um produto. Nesse caso, escrevemos a cesta de produto líquido como $(y, -x)$, em que x é um vetor de insumos que podem produzir y unidades de produto. Podemos então definir um caso especial de um conjunto de possibilidades de produção restrito, o conjunto de requerimento de insumos:*

$$V(y) = \{x \in \mathbb{R}^L_+ \colon (y, -x) \text{ } \textit{está em } Y\} \tag{3.1}$$

O conjunto de requerimento de insumos é o conjunto de todas as cestas de insumos que produzem pelo menos y unidades de produto.

Figura 3.1 – Conjunto de requerimento de insumos convexo

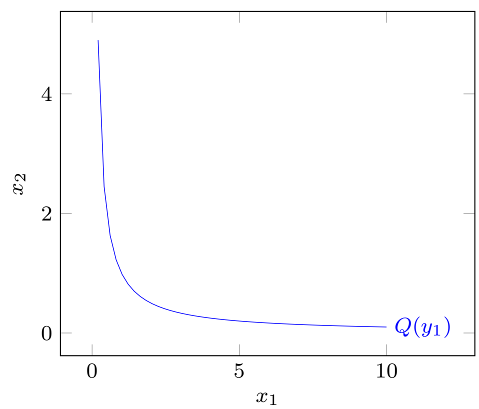


Exemplo 3.2.2 (Isoquantas)**.** *No caso acima, podemos também definir uma isoquanta:*

$$\mathcal{Q}(y) = \{x \in \mathbb{R}^n_+ : x \text{ está em } V(y) \text{ e não está em } V(y') \text{ para } y' > y\} \tag{3.2}$$

A isoquanta fornece todos as cestas de insumos que produzem exatamente y unidades de produto.

Figura 3.2 – Isoquanta



Exemplo 3.2.3 (Conjunto de possibilidades de produção no curto prazo)**.** *Suponha que uma empresa produza algum produto a partir do trabalho e de algum tipo de máquina a que nos referiremos como "capital". Os planos de produção então se parecem com* $(y, -l, -k)$ *em que* y *é o nível de produção,* l *a quantidade do insumo trabalho e* k *a quantidade do insumo capital. Nós imaginamos que o trabalho pode ser variado imediatamente, mas que o capital é fixo no nível* $\overline{k}$ *no curto prazo. Então*

$$Y(\overline{k}) = \{(y, -l, -k) \in Y : k = \overline{k}\} \tag{3.3}$$

é o conjunto de possibilidades de produção de curto prazo.

Exemplo 3.2.4 (Função de produção)**.** *Se a empresa tiver apenas um produto, podemos definir a seguinte função de produção função de produção:*

$$f(x) = \{y \in \mathbb{R} \colon y \text{ é o produto máximo associado a } -x \in Y\} \tag{3.4}$$

Exemplo 3.2.5 (Função de transformação)**.** *Um plano de produção y em Y é (tecnologicamente) eficiente se não houver y' em Y tal que $y' \geqq y$ e $y' \neq y$; ou seja, um plano de produção é eficiente se não houver maneira de produzir mais produto com os mesmos insumos ou produzir o mesmo produto com menos insumos. Muitas vezes assumimos que podemos descrever o conjunto de planos de produção tecnologicamente eficientes por uma função de transformação $T \colon \mathbb{R}^L \to \mathbb{R}$ em que $T(y) = 0$ se e somente se y é eficiente. Assim como uma função de produção seleciona o produto escalar máximo como uma função dos insumos, a função de transformação seleciona os vetores máximos de produtos líquidos.*

Exemplo 3.2.6 (Função de produção Cobb-Douglas)**.** *Seja α um parâmetro tal que $0 < \alpha < 1$. Então a tecnologia de produção Cobb-Douglas é definida da seguinte maneira:*

$$\begin{aligned}
Y &= \{(y, -x_1, -x_2) \in \mathbb{R}^3 \colon y \leqq x_1^\alpha x_2^{1-\alpha}\} \\
V(y) &= \{(x_1, x_2) \in \mathbb{R}^2_+ \colon y \leqq x_1^\alpha x_2^{1-\alpha}\} \\
Q(y) &= \{(x_1, x_2) \in \mathbb{R}^2_+ \colon y = x_1^\alpha x_2^{1-\alpha}\} \\
Y(z) &= \{(y, -x_1, -x_2) \in \mathbb{R}^3 \colon y \leqq x_1^\alpha x_2^{1-\alpha}, x_2 = z\} \\
T(y, x_1, x_2) &= y - x_1^\alpha x_2^{1-\alpha} \\
f(x_1, x_2) &= x_1^\alpha x_2^{1-\alpha}
\end{aligned}$$

Exemplo 3.2.7 (Função de produção Leontief)**.** *Seja $a > 0$ e $b > 0$ parâmetros. Então a tecnologia de produção Leontief é definida da seguinte maneira:*

$$\begin{aligned}
Y &= \{(y, -x_1, -x_2) \in \mathbb{R}^3 \colon y \leqq \min\{ax_1, bx_2\}\} \\
V(y) &= \{(x_1, x_2) \in \mathbb{R}^2_+ \colon y \leqq \min\{ax_1, bx_2\}\} \\
Q(y) &= \{(x_1, x_2) \in \mathbb{R}^2_+ \colon y = \min\{ax_1, bx_2\}\} \\
T(y, x_1, x_2) &= y - \min\{ax_1, bx_2\} \\
f(x_1, x_2) &= \min\{ax_1, bx_2\}
\end{aligned}$$

3.2.3 Propriedades Comuns dos Conjuntos de Produção

Embora os conjuntos de possibilidades de produção de diferentes processos possam diferir amplamente na estrutura, muitas tecnologias compartilham certas propriedades gerais. Se for possível supor que essas propriedades são satisfeitas, podem ser obtidos resultados teóricos especiais. Algumas propriedades importantes são definidas abaixo.

1. Possibilidade de inação: $0 \in Y$

   Possibilidade de inação significa que nenhuma ação na produção é um plano de produção possível.

2. Livre descarte ou monotonicidade

   Se $y \in Y$ implica que $y' \in Y$ para todo $y' \leqq y$, então o conjunto Y é dito satisfazer a propriedade de livre descarte ou monotonicidade. O livre descarte implica que os bens (insumos ou produtos) podem ser descartados. Essa propriedade significa que, se $y \in Y$, Y inclui todos os vetores no ortante negativo traduzidos em y, ou seja, existem apenas insumos, mas não produtos. Um requisito mais fraco é que nós apenas assumamos que o requisito de insumo é monotônico: se x está em $V(y)$ e $x' \geq x$, então x' está em $V(y)$. A monotonicidade de $V(y)$ significa que, se x é uma maneira viável de produzir y unidades de produto e x' é um vetor de insumo com pelo menos tanto de cada insumo, então x' deve ser uma maneira viável de produzir y.

3. Irreversibilidade: $Y \cap \{-Y\} = \{0\}$

   Irreversibilidade significa que um plano de produção não é reversível, a menos que seja um plano de não-ação.

4. Convexidade

   Y é convexo se sempre que y e y' estão em Y, a média ponderada $ty + (1-t)y$ também está em Y para qualquer t com $0 \leqq t \leqq 1$.

   A convexidade de Y significa que, se todos os bens são divisíveis, é razoável assumir que dois planos de produção y e y' podem ser reduzidos e combinados. No entanto, deve-se notar que a convexidade do conjunto de produção é uma hipótese forte. Por exemplo, a convexidade do conjunto de produção exclui os custos iniciais e outros tipos de retorno à escala. Isso será discutido em maior detalhe em breve.

5. Convexidade estrita

   Y é estritamente convexo se sempre que y e y' estão em Y, a média ponderada $ty + (1-t)y$ também está em Y para qualquer t com $0 < t < 1$.

   Como mostraremos, a convexidade estrita de Y pode garantir que o plano de produção que maximiza o lucro seja único, desde que exista.

   Proposição 3.2.1. *$V(y)$ (o conjunto de requerimento de produção) é um conjunto convexo se e somente se a função de produção $f(x)$ é uma função quase côncava.*

3.2.4 Retornos Marginais Decrescentes

Explicação baseada em Armen Alchian e William Allen. O exemplo mais tradicional para ilustrar a lei dos rendimentos marginais decrescentes sempre foi o da ocupação de terras adicionais.

Foi assim com David Ricardo, foi assim com von Thunen e também com John Bates Clark. Quando os agricultores ou proprietários de terra vão ocupando a fronteira agrícola (aquelas terras que ainda não foram exploradas), as terras que são ocupadas primeiro são as melhores dentre as disponíveis. São, portanto, as terras mais férteis ou mais produtivas. Para a mesma quantidade de fatores de produção, essas terras mais produtivas produzem mais produto do que as terras menos férteis. Quando a fronteira se expande, já tendo sido exploradas as terras boas, são as terras piores que passam a ser ocupadas. O produto marginal proveniente da utilização da parcela adicional da terra menos fértil é menor do que o produto marginal da parcela ocupada imediatamente antes. Com a mesma quantidade de fatores, por exemplo, capital e trabalho, a quantidade de produto é menor, pois a terra é menos fértil. É a isso que chamamos de Lei dos Rendimentos Marginais Decrescentes.

A melhor maneira de entender o seu significado é imaginar o que aconteceria se ela não valesse. Assim sendo, suponha que você possui um lote de terra de 100 acres e que você usa certo montante de fatores para produzir uma quantidade específica de produto. O proprietário de terras vizinhas à sua lhe concede gratuitamente o direito de uso, por um ano, de um lote adjacente ao seu com a mesma medida e com a mesma qualidade, um lote de terra perfeitamente igual ao seu em todos os aspectos economicamente relevantes. É evidente que, para você usar esse lote adicional de terra para produzir, você precisaria aplicar recursos, fatores de produção, e os aplicaria nas mesmas quantidades que você aplica no seu atual lote, para produzir a mesma quantidade. Se a lei dos rendimentos marginais decrescentes não fosse válida, você não ganharia coisa alguma com a utilização desse lote marginal que você já não ganharia se aplicasse os recursos adicionais no seu próprio lote, mesmo que o lote adicional lhe seja cedido de graça. Você simplesmente aplicaria os recursos adicionais ao seu próprio lote para produzir mais do que você já produzia. Em outras palavras, com um vaso de planta você poderia alimentar o mundo.

Não é isso obviamente que observamos. Como entender então os conceitos de retornos marginais constantes e crescentes? Enquanto houver lotes de terra tão bons quanto o seu atual lote, esses lotes serão explorados pelo mercado. Alguém que não possua o direito de propriedade sobre um desses lotes adicionais de igual qualidade e que, de seu ponto de vista privado, julgue ser economicamente recomendável adquirir o lote e plantar, isto é, caso o retorno líquido de aquisição e produção seja maior do que o seu custo privado de oportunidade de não fazê-lo, poderá adquirir finalmente o lote adicional e produzir. Eventualmente todos os lotes de igual qualidade (e produtividade), estarão esgotados e o mercado deverá buscar lotes de qualidade inferior. Enquanto isso não ocorre, o que se observa são retornos constantes de escala. Observaríamos retornos crescentes de escala se se descobrissem novos lotes de terra de qualidade superior às já utilizadas. Esses lotes adicionais terão evidentemente produtividade maior. É por isso que fronteiras agrícolas crescem tão rapidamente quando deixadas sob a mão benéfica do livre mercado e da livre transação de direitos de propriedade. Uma vez exploradas, novas terras terão naturalmente qualidade inferior e menor produtividade, configurando-se, assim, retornos marginais decrescentes.

O esgotamento das terras de igual produtividade e o consequente uso de terras de menor qualidade tem o seu correlato também no fator de produção trabalho. Um marceneiro trabalha 8 horas por dia. Se trabalhar uma hora a mais, seu produto certamente será maior. Se trabalhar uma segunda hora adicional, então seu produto será ainda maior, mas o acréscimo de produto é menor do que o acréscimo proveniente da hora adicional anterior. Isso se dá em razão do esgotamento tanto físico quanto do ânimo. Variemos um pouco. Na primeira hora de trabalho, no início do dia, todo seu esforço produz uma determinada quantidade de produto e é bem possível que a mera contemplação do produto da primeira hora e a expectativa de usufruto futuro de seu retorno contribua para que na segunda hora de trabalho o cansaço físico seja superado pelo ânimo e pelo mergulho mental no trabalho. Isso é equivalente à descoberta de terras mais produtivas. Essas primeiras horas, portanto, podem apresentar retornos crescentes de escala. Após certo número de horas, o cansaço fala mais alto, o esgotamento mental reduz o ânimo e o produto incremental se reduz. Não há mais onde buscar forças. Entre esses dois estágios, não há perda de generalidade em supor que por um breve intervalo de tempo o trabalho apresente retornos constantes de escala.

O princípio da produtividade marginal decrescente ou simplesmente lei dos rendimentos decrescentes é, portanto, uma lei geral. Se em um mercado encontram-se retornos crescentes de escala, é porque se enfrenta uma situação análoga à da fronteira agrícola com terras mais férteis ainda sendo descobertas e certamente há de chegar o momento em que o esgotamento de todas as oportunidades não exploradas levará aos retornos decrescentes. Assim, os recursos disponíveis serão naturalmente alocados para a produção em unidades adicionais de terras de melhor qualidade até o ponto em que os retornos marginais de um dado montante de recursos sejam igualados entre si em todos os seus usos possíveis.

Observe que, no exemplo em que você atuava como o proprietário de terras a quem foi cedido um lote adicional de igual qualidade, a sua decisão de desprezar a terra incremental se dá ao fato de que a transação de cessão do lote incremental pode exigir que se incorra em custos de transação que você não está disposto a aceitar. Existem custos de instalação e sunk costs que não são recuperáveis. Para outro potencial proprietário, porém, pode ser que valha a pena incorrer em tais custos, se o custo de oportunidade assim o recomendar. Logo, existe um elemento na ideia de retornos de escala que transcende o mero conjunto de possibilidades de produção ou tecnologia. Não é apenas a tecnologia em si que apresenta retornos de escala crescentes, constantes ou decrescentes. A fonte dessas propriedades é a junção da estrutura tecnológica com as condições econômicas em que ela é utilizada. Se os custos de transação, até mesmo para a simples cessão gratuita da terra incremental, forem infinitos para aquele específico lote incremental naquela região (por exemplo, numa sociedade teocrática em que o proprietário doador fosse um excomungado e você proibido de transacionar com ele, sob pena de morte na fogueira), então aquele lote de igual qualidade, jamais será utilizado. Nem mesmo será utilizado se apresentasse maior produtividade e você pudesse simplesmente permutar a sua terra pela do vizinho. Há, assim um elemento de natureza institucional (os direitos de propriedade, os custos de transação) e um de natureza econômica (os interesses alheios e custos de oportunidade na aquisição e utilização das terras marginais) inerentes ao conceito de retornos de escala. O caso das fronteiras agrícolas é o melhor exemplo. Ainda que as terras da fronteira agrícola sejam as mais férteis, jamais veremos retornos crescentes de escala se a aquisição e utilização dessas terras sofressem de severas restrições institucionais. Pode ser também que os custos de oportunidades dos potenciais proprietários agrícolas na fronteira sejam tão altos que as terras sequer sejam ocupadas. Eis aí um aspecto do conceito que é simplesmente ignorado nos livros.

O princípio da produtividade marginal decrescente de um fator de produção requer que todos os demais fatores se mantenham constantes. Suponha que a produção de tábuas de madeira de um bem dependa de capital e trabalho, sendo que o capital é medido pelo valor monetário dos instrumentos ou ferramentas de trabalho utilizados durante a semana, como por exemplo, serras elétricas, machados etc., ao passo que o trabalho é medido pelo número de trabalhadores durante a semana. Um trabalhador adicional contribui positivamente para o nível de produto, mantendo-se constante o valor das ferramentas. O segundo trabalhador adicional também contribui positivamente, mas como o valor das ferramentas é constante, isso só será possível se as ferramentas forem menos elaboradas. Assim, uma serra elétrica, que só pode ser usada por um trabalhador, terá que dar lugar a dois machados, o que fará com que o aumento de produto do segundo trabalhador seja menor do que o aumento do produto decorrente da adição do trabalhador anterior. Alternativamente, se os instrumentos forem de igual qualidade, o segundo trabalhador será menos produtivo do que o primeiro, em razão de todos os trabalhadores de igual qualidade do primeiro já terem sido empregados.

O princípio da produtividade marginal decrescente é uma lei empírica que adquiriu ares de generalidade com o passar do tempo. Já em 1815 Thomas Malthus (Thomas Malthus: Nature and Progress of Rent) e Sir Edward West (Edward West: The Application of Capital to Land, Hollander ed., 1815) afirmavam que se doses adicionais de capital e trabalho fossem aplicadas a um lote de terra, o produto cresceria por incrementos decrescentes. Na verdade, antes deles, em 1768, Turgot (Turgot: Oeuvres, 1844) já afirmara que sucessivas aplicações de trabalho a um lote de terra levariam a incrementos decrescentes do produto. O problema é que, para Malthus e West, capital e trabalho eram variados simultaneamente para uma mesma quantidade de terra. Não perceberam eles que o incremento decrescente do produto poderia decorrer de incrementos de cada fator separadamente, ou seja, ceteris paribus. Foi somente nos anos 1840 que von Thünen (Johann von Thünen: Der Isolierte Staat; Zweiter Teil, pp. 507 e 557559) quebrou a dose combinada de capital e trabalho em doses separadas, proclamando a lei da produtividade marginal decrescente. Por um desses infelizes acidentes históricos ou mais pela insistência dos economistas britânicos de se isolarem dos pensadores europeus do continente, a obra de von Thünen não teve a divulgação merecida. Foi o economista norte-americano John Bates Clark quem em 1888 deu vida nova ao princípio da produtividade marginal decrescente em um encontro da American Economic Association, tendo culminado seus estudos com sua obra clássica de 1899, The Distribution of Wealth.

3.2.5 Retornos de Escala

Suponha que estamos usando algum vetor de insumos x para produzir algum produto y e decidimos "escalar" todas os insumos para cima ou para baixo em alguma quantidade $t \geqq 0$. O que acontecerá com o nível de produto? As noções de retorno à escala podem ser usadas para responder a essa pergunta. Retorno à escala se refere a como o produto varia quando os insumos são todos variados na mesma proporção, de modo que considerem processos de produção de longo prazo. Existem três possibilidades: a tecnologia exibe (i) retornos constantes de escala; (ii) retornos decrescentes de escala e (iii) retornos crescentes de escala. Formalmente, nós temos

Proposição 3.2.2 (Retorno global de escala)**.** *Uma função de produção $f(x)$ exibe*

- *retornos constantes de escala se* $f(tx) = tf(x)$ *para todo* $t \geqq 0$.
- *retornos decrescentes de escala se* $f(tx) < tf(x)$ *para todo* $t > 1$.
- *retornos crescentes de escala se* $f(tx) > tf(x)$ *para todo* $t > 1$.

Retornos constantes de escala significa que a duplicação de insumos duplica exatamente os produtos, o que geralmente é uma suposição razoável a ser feita sobre tecnologias. Retornos de escala decrescentes significa que o dobro de insumos está diminuindo a produção. Retornos de escala crescentes significa que a duplicação de insumos está mais do que dobrando os resultados. Note que uma tecnologia tem retornos constantes de escala se, e somente se, sua função de produção é homogênea de grau 1.

Pode-se notar que os vários tipos de retornos à escala definidos acima são de natureza global. Pode muito bem acontecer que uma tecnologia apresente retornos crescentes de escala para alguns valores de x e retornos decrescentes de escala para outros valores. Assim, em muitas circunstâncias, uma medida local de retorno à escala é útil. Para definir localmente, retorna à escala, primeiro definimos a elasticidade da escala.

A elasticidade da escala mede o aumento percentual da produção devido a um aumento de um por cento em todos os insumos - isto é, devido a um aumento na escala de operações.

Seja $y = f(x)$ a função de produção. Seja t um escalar positivo e considere a função $y(t) = f(tx)$. Se $t = 1$, temos a escala atual de operação; se $t > 1$, estamos aumentando todos os insumos por t; se $t < 1$, estamos redimensionando para baixo todos os insumos por t. A elasticidade da escala é dada por

$$e(x) = \frac{\dfrac{dy(t)}{y(t)}}{\dfrac{dt}{t}}, \tag{3.5}$$

avaliada em $t = 1$.

Rearranjando esta expressão, temos:

$$e(x) = \frac{dy(t)}{dt}\frac{t}{y}\bigg|_{t=1} = \frac{df(tx)}{dt}\frac{t}{f(tx)}\bigg|_{t=1}. \tag{3.6}$$

Note que devemos avaliar a expressão em $t = 1$ para calcular a elasticidade da escala no ponto x. Assim, temos os seguintes retornos locais para escala: diz-se que uma função de produção $f(x)$ exibe retornos localmente crescentes, constantes ou decrescentes, à medida que $e(x)$ é maior, igual ou menor que 1.

3.2.6 Taxa Marginal de Substituição Técnica

Suponha que a tecnologia seja resumida por uma função de produção e que estamos produzindo em um determinado ponto $y^* = f(x_1^*, x_2^*)$. Suponha que queremos aumentar uma pequena quantidade do insumo 1 e diminuir alguma quantidade do insumo 2, de modo a manter um nível constante de produto. Como podemos determinar essa taxa marginal de substituição técnica (MRTS) entre esses dois fatores? O caminho é o mesmo que para derivar a taxa marginal de substituição de uma curva de indiferença. Diferenciando a função de produção quando o produto se mantém constante, temos:

$$\frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 = 0, \tag{3.7}$$

que pode ser resolvido para:

$$\frac{dx_2}{dx_1} = -\frac{\frac{\partial f}{\partial x_1}}{\frac{\partial f}{\partial x_2}} \equiv -\frac{PMg_{x_1}}{PMg_{x_2}}. \tag{3.8}$$

Exemplo 3.2.8. *Seja a função de produção* $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$. *Assim,*

$$\frac{\partial f(x)}{\partial x_1} = \alpha x_1^{\alpha-1} x_2^{1-\alpha} = \alpha \left(\frac{x_2}{x_1}\right)^{1-\alpha} \tag{3.9}$$

$$\frac{\partial f(x)}{\partial x_2} = (1-\alpha) x_1^{\alpha} x_2^{-\alpha} = (1-\alpha) \left(\frac{x_1}{x_2}\right)^{\alpha} \tag{3.10}$$

Segue que

$$\frac{dx_2}{dx_1} = -\frac{\alpha}{1-\alpha} \frac{x_2}{x_1} \tag{3.11}$$

3.2.7 Elasticidade de Substituição

A taxa marginal de substituição técnica mede a inclinação de uma isoquanta. A elasticidade da substituição mede a curvatura de uma isoquanta. Mais especificamente, a elasticidade de substituição mede a variação percentual na razão de fatores dividida pela variação percentual na taxa marginal de substituição técnica, com a produção sendo mantida fixa. Se denotarmos por $\Delta \left(\frac{x_2}{x_1}\right)$ a mudança na razão dos fatores e por ΔMRTS a mudança na taxa técnica de substituição, podemos expressar isso como

$$\sigma = \frac{\frac{\Delta\left(\frac{x_2}{x_1}\right)}{\frac{x_2}{x_1}}}{\frac{\Delta \text{MRTS}}{\text{MRTS}}}. \tag{3.12}$$

Essa é uma medida relativamente natural da curvatura: ela pergunta como a proporção dos insumos muda à medida que a inclinação da isoquanta muda. Se uma pequena mudança na inclinação nos dá uma grande mudança na taxa dos insumos, a isoquanta é relativamente plana, o que significa que a elasticidade da substituição é grande.

Em termos gerais, pode ser utilizada a seguinte expressão:

$$\sigma_{ij}(x) = \frac{x_i f_i(x) + x_j f_j(x)}{f_j^2(x) f_{ii}(x) + 2 f_i(x) f_j(x) f_{ij}(x) + f_i^2(x) f_{jj}(x)} \frac{f_i(x) f_j(x)}{x_i x_j}, \tag{3.13}$$

em que $f_i(x) = \dfrac{\partial f(x)}{\partial x_i}$.

Na prática, pensamos na mudança percentual como sendo muito pequena e tomamos o limite dessa expressão quando $\Delta \to 0$. Assim, a expressão para σ torna-se

$$\sigma = \frac{\text{MRTS}}{\dfrac{x_2}{x_1}} \frac{d\left(\dfrac{x_2}{x_1}\right)}{d\text{MRTS}} = \frac{d\ln\left(\dfrac{x_2}{x_1}\right)}{d\ln|\text{MRTS}|}. \tag{3.14}$$

Exemplo 3.2.9. *Nós vimos que* $MRTS = -\dfrac{\alpha}{1-\alpha}\dfrac{x_2}{x_1}$.

Segue que

$$\begin{aligned}
\frac{x_2}{x_1} &= -\frac{1-\alpha}{\alpha} MRTS \\
\ln\left(\frac{x_2}{x_1}\right) &= \ln\left(-\frac{1-\alpha}{\alpha}\right) + \ln|MRTS| \\
\sigma &= \frac{d\ln\left(\dfrac{x_2}{x_1}\right)}{d\ln|MRTS|} \\
\sigma &= 1
\end{aligned} \tag{3.15}$$

Exemplo 3.2.10. *Seja a função de produção CES dada por* $y = (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}}$. *Assim,*

$$\frac{\partial y}{\partial x_1} = (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}-1} x_1^{\rho-1} \tag{3.16}$$

$$\frac{\partial y}{\partial x_2} = (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}-1} x_2^{\rho-1} \tag{3.17}$$

Segue que

$$\begin{aligned}
MRTS &= -\left(\frac{x_1}{x_2}\right)^{\rho-1} \\
\frac{x_2}{x_1} &= \frac{1}{|MRTS|^{\dfrac{1}{\rho-1}}} \\
\ln\left(\frac{x_2}{x_1}\right) &= \frac{1}{\rho-1}\ln|MRTS| \\
\sigma &= \frac{d\ln\left(\dfrac{x_2}{x_1}\right)}{d\ln|MRTS|} \\
\sigma &= \frac{1}{\rho-1} \qquad (3.18)
\end{aligned}$$

A elasticidade da substituição indica a facilidade com que as empresas podem mudar seu mix de insumos à medida que os preços relativos dos insumos mudam. A elasticidade de substituição é determinada pela função de produção.

Valores mais altos de σ indicam que a empresa pode substituir mais facilmente os insumos à medida que os preços relativos dos insumos mudam. Por exemplo, uma firma com uma alta elasticidade de substituição responderia a um aumento na taxa de câmbio substituindo rapidamente o capital pelo trabalho (por exemplo, substituindo os trabalhadores por máquinas). Deve ficar claro, então, que uma empresa com uma tecnologia de produção caracterizada por uma alta elasticidade de substituição também deve ter uma elevada elasticidade da demanda de mão-de-obra em relação aos salários. Uma empresa que possa substituir facilmente entre trabalho e capital responderia a uma mudança nos salários, mudando rapidamente entre trabalho e capital e mudando rapidamente sua demanda de trabalho.

3.2.8 Função de Produção Homotética

O conceito de homotetia foi introduzido na literatura econômica por Shephard[2]. Essencialmente, uma função de produção é homotética se seu mapa de isoquanta coincide com o mapa de isoquanta de uma função homogênea de grau 1. Funções de produção homotéticas são muito comuns em pesquisas econômicas aplicadas, mormente por apresentarem propriedades geométricas simples e com significados econômicos imediatos.

Definição 3.2.1. *Uma função de produção $f(x)$ é dita homotética se $f(x)$ for uma transformação crescente de uma função homogênea de grau 1, ou seja, se $f(x)$ pode ser escrita como $f(x) = \varphi(h(x))$, em que $\varphi'(\cdot) > 0$ ($\varphi\colon \mathbb{R} \to \mathbb{R}$ é estritamente crescente) e $h(x)$ é homogênea de grau 1.*

Na geometria, uma homotetia é uma transformação afim $\mathscr{H}\colon \mathbb{R}^\ell \to \mathbb{R}^\ell$ determinada por um ponto $a = (a_1, \ldots, a_\ell)$, chamado centro, e um escalar não nulo λ, chamado raio, e dada por $\mathscr{H}(y) = a + \lambda(y - a)$. O que ela faz é deslocar rigidamente um objeto pelo deslocamento da origem $0 = (0, \ldots, 0)$ do sistema de coordenadas até o ponto $a = (a_1, \ldots, a_m)$. Com esse movimento, o

[2] Seção baseada em Shephard, R. (1953). Cost and Production Functions. Princeton University Press.

objeto se desloca também. Chegado a esse ponto, o objeto é reescalado pelo fator λ. Será uma dilatação se $\lambda > 1$; uma contração, se $0 < \lambda < 1$.

O conceito de homotetia que usamos em Economia é o caso particular em que $a = 0$, isto é, a é o vetor nulo, o que significa simplesmente que o componente de deslocamento não existe, apenas o de reescalagem. Neste caso, $\mathscr{H}(y) = \lambda y$. Ao tomarmos um plano de produção $y = (-x, q)$, em que $q = f(x)$, então a transformação homotética de y nos dá $\mathscr{H}(y) = \lambda y$, ou seja,

$$\begin{aligned} \mathscr{H}(-x, q) &= \mathscr{H}(-x, f(x)) \\ &= \lambda(-x, f(x)) \\ &= (-\lambda x, \lambda f(x)) \\ &= (-\lambda x, \lambda q). \end{aligned} \tag{3.19}$$

Isso implica que, se o vetor de insumos x produz a quantidade q de produto, então λx produz λq. No que concerne à geometria do mapa de isoquanta, qualquer transformação estritamente crescente da homotetia apresentará o mesmo mapa da homotetia subjacente. É por isso que em Economia a homotetia é definida do modo que comumente é. Essa definição, porém, esconde sob o mesmo manto duas coisas distintas: a homotetia propriamente dita e a configuração geométrica do mapa de isoquanta. Quando $f(x)$ é uma função de produção homotética, ela possui, como dissemos, o mesmo mapa de isoquanta da função homogênea de grau 1 que lhe é subjacente. De fato, a taxa técnica de substituição entre os fatores i e j da função f em x, denotada por $TTS^f_{ij}(x)$, satisfaz:

$$\begin{aligned} TTS^f_{ij}(x) &= -\frac{PMg^f_i(x)}{PMg^f_j(x)} \\ &= -\frac{\dfrac{\partial f(x)}{\partial x_i}}{\dfrac{\partial f(x)}{\partial x_j}}. \end{aligned} \tag{3.20}$$

Pela regra da cadeia,

$$-\frac{\dfrac{\partial f(x)}{\partial x_i}}{\dfrac{\partial f(x)}{\partial x_j}} = -\frac{\varphi'(h(x))\dfrac{\partial h(x)}{\partial x_i}}{\varphi'(h(x))\dfrac{\partial h(x)}{\partial x_j}}. \tag{3.21}$$

Como por hipótese $\varphi'(h(x)) > 0$, temos:

$$-\frac{\dfrac{\partial f(x)}{\partial x_i}}{\dfrac{\partial f(x)}{\partial x_j}} = -\frac{\dfrac{\partial h(x)}{\partial x_i}}{\dfrac{\partial h(x)}{\partial x_j}}. \tag{3.22}$$

Substituindo na expressão acima, $TTS_{ij}^f(x) = -\dfrac{\frac{\partial h(x)}{\partial x_i}}{\frac{\partial h(x)}{\partial x_j}}$. Porém, por definição, $\dfrac{\partial h(x)}{\partial x_i} = PMg_i^h(x)$, $\forall\ i = 1, \ldots, m$. Portanto,

$$-\frac{\frac{\partial h(x)}{\partial x_i}}{\frac{\partial h(x)}{\partial x_j}} = -\frac{PMg_i^h(x)}{PMg_j^h(x)}. \tag{3.23}$$

Logo,

$$TTS_{ij}^f(x) = TTS_{ij}^h(x). \tag{3.24}$$

Assim, as taxas técnicas de substituição da função homotética em cada ponto coincidem com as taxas técnicas de substituição da função homogênea de grau 1 que lhe subjaz. Como uma curva é totalmente determinada pelas suas inclinações em cada ponto, concluímos que as curvas que representam as isoquantas são exatamente as mesmas. Entretanto, os níveis não são os mesmos.

Outra propriedade geométrica importante da função de produção homotética é que, ao longo de linhas retas que partem da origem, as taxas técnicas de substitução são constantes. Na verdade, a intuição econômica por trás da homotetia é que é suficiente que as taxas técnicas de substituição sejam constantes ao longo de retas que saem da origem, ou seja, os caminhos de expansão são retas, ao longo das quais a combinação relativa ótima de dois quaisquer fatores é a mesma para qualquer nível de produto:

$$\frac{\frac{\partial f(tx)}{\partial x_i}}{\frac{\partial f(tx)}{\partial x_j}} = \frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}}, \quad t > 0, \quad \forall\ i \neq j. \tag{3.25}$$

Exemplo 3.2.11. *Uma condição técnica importante para se evitar a divisão por zero é que os produtos marginais sejam estritamente positivos. Alguns autores afirmam que a condição acima é equivalente ao conceito de homotetia. Essa equivalência, porém, é falsa. Considere a seguinte função de produção:*

$$f(K, L) = \begin{cases} (L - K)^2 & \textit{se } K \leq L \\ (K - L)^3 & \textit{se } K > L \end{cases} \tag{3.26}$$

A reta $D = \{(K, L) \in \mathbb{R}_+^2 : K = L\}$ divide o quadrante $\mathbb{R}_+^2$ em dois cones. Pode-se mostrar que f é continuamente diferenciável (inclusive em pontos de D) e que satisfaz a condição acima. Com efeito, na região dada pelo cone $\{(K, L) \in \mathbb{R}_+^2 : K < L\}$, temos:

$$\begin{aligned} f(tK,tL) &= (tL - tK)^2 \\ &= t^2(L-K)^2 \\ &= t^2 f(K,L), \end{aligned} \tag{3.27}$$

de modo que, para qualquer $t > 0$:

$$\begin{aligned} \frac{\dfrac{\partial f(tK,tL)}{\partial K}}{\dfrac{\partial f(tK,tL)}{\partial L}} &= \frac{t^2 \dfrac{\partial f(K,L)}{\partial K}}{t^2 \dfrac{\partial f(K,L)}{\partial L}} \\ &= \frac{\dfrac{\partial f(K,L)}{\partial K}}{\dfrac{\partial f(K,L)}{\partial L}}. \end{aligned} \tag{3.28}$$

Similarmente, na região dada pelo cone $\{(K,L) \in \mathbb{R}^2_+ : K > L\}$, temos:

$$\begin{aligned} f(tK,tL) &= (tK - tL)^3 \\ &= t^3(K-L)^3 \\ &= t^3 f(K,L), \end{aligned} \tag{3.29}$$

de modo que, para qualquer $t > 0$:

$$\begin{aligned} \frac{\dfrac{\partial f(tK,tL)}{\partial K}}{\dfrac{\partial f(tK,tL)}{\partial L}} &= \frac{t^3 \dfrac{\partial f(K,L)}{\partial K}}{t^3 \dfrac{\partial f(K,L)}{\partial L}} \\ &= \frac{\dfrac{\partial f(K,L)}{\partial K}}{\dfrac{\partial f(K,L)}{\partial L}}. \end{aligned} \tag{3.30}$$

Na região fronteiriça $D = \{(K,L) \in \mathbb{R}^2_+ : K = L\}$ o mesmo ocorre, devido à diferenciabilidade de f em D. Mas $f(tK,tL)$ cresce à taxa t^2 em um cone e à taxa t^3 em outro, não podendo ser, por conseguinte, homotética.

Exemplo 3.2.12. *Seja $f(K,L) = KL$ uma função de produção. Ora, podemos reescrever $f(K,L)$ como $f(K,L) = (\sqrt{KL})^2$. Então, $f(K,L) = \varphi(h(K,L))$, em que $h(K,L) = \sqrt{KL}$ é homogênea de grau 1 e $\varphi(\theta) = \theta^2$ é estritamente crescente para $\theta > 0$. Logo, $f(K,L) = KL$ é homotética.*

Para mostrarmos a validade desse resultado para uma função de produção homotética com dois insumos, digamos $f(K,L)$, fixe uma reta pela origem dada por $L = \alpha K$, em que $\alpha > 0$ é constante. Vamos mostrar que, ao longo dessa linha reta (relação capital/trabalho constante), a

TTS é constante. Sem perda de generalidade, podemos supor que a função de produção é homogênea de grau 1, pois o mapa de isoquanta é o mesmo da função homogênea de grau 1 subjacente à definição de homotetia de $f(K, L)$. Neste caso:

$$\begin{aligned} TTS_{KL}(K, L) &= TTS_{KL}(K, \alpha K) \\ &= -\frac{\dfrac{\partial f(K, \alpha K)}{\partial K}}{\dfrac{\partial f(K, \alpha K)}{\partial L}}. \end{aligned} \tag{3.31}$$

Como f é homogênea de grau 1, os produtos marginais são homgêneos de grau zero, de modo que (dividindo cada argumento da função por K)

$$\frac{\partial f(K, \alpha K)}{\partial K} = \frac{\partial f(1, \alpha)}{\partial K} \tag{3.32}$$

$$\frac{\partial f(K, \alpha K)}{\partial L} = \frac{\partial f(1, \alpha)}{\partial L}. \tag{3.33}$$

Portanto,

$$\begin{aligned} TTS_{KL} &= -\frac{\dfrac{\partial f(K, \alpha K)}{\partial K}}{\dfrac{\partial f(K, \alpha K)}{\partial L}} \\ &= -\frac{\dfrac{\partial f(1, \alpha)}{\partial K}}{\dfrac{\partial f(1, \alpha)}{\partial L}}, \end{aligned} \tag{3.34}$$

que é constante, pois não depende de K e L, apenas de α. Podemos dizer algo mais. Pelo teorema da exaustão do produto temos que

$$\frac{\partial f(K, \alpha K)}{\partial K} K + \frac{\partial f(K, \alpha K)}{\partial L} \alpha K = f(K, \alpha K). \tag{3.35}$$

Dividindo por K em ambos os lados da igualdade e lembrando que, se f é homogênea de grau 1, então os produtos marginais são funções homogêneas de grau zero, temos que

$$\frac{\partial f(1, \alpha)}{\partial K} + \frac{\partial f(1, \alpha)}{\partial L} \alpha = f(1, \alpha). \tag{3.36}$$

Isolando a $TTS(K, \alpha K)$, encontramos:

$$
\begin{aligned}
TTS_{KL}(K,\alpha K) &= -\frac{\dfrac{\partial f(1,\alpha)}{\partial K}}{\dfrac{\partial f(1,\alpha)}{\partial L}} \\
&= \alpha - \frac{f(1,\alpha)}{\dfrac{\partial f(1,\alpha)}{\partial L}} \\
&= \alpha - \alpha \frac{\dfrac{f(1,\alpha)}{\alpha}}{\dfrac{\partial f(1,\alpha)}{\partial L}}
\end{aligned} \tag{3.37}
$$

Temos que $\varepsilon_L = \dfrac{\dfrac{\partial f(1,\alpha)}{\partial L}}{\dfrac{f(1,\alpha)}{\alpha}}$. *Assim,*

$$
\begin{aligned}
TTS_{KL}(K,\alpha K) &= \alpha - \alpha \frac{1}{\varepsilon_L(1,\alpha)} \\
&= -\alpha \left(\frac{1-\varepsilon_L(1,\alpha)}{\varepsilon_L(1,\alpha)} \right).
\end{aligned} \tag{3.38}
$$

Sabemos que $\varepsilon_K(1,\alpha)+\varepsilon_L(1,\alpha) = e(K,L)$ *é a elasticidade de escala. Como supusemos, sem perda de generalidade, que a função de produção é homogênea de grau 1, então* $e(K,L) = 1$, *de modo que* $\varepsilon_K(1,\alpha)+\varepsilon_L(1,\alpha) = 1$. *Temos, portanto, para funções de produção homogêneas de grau 1:*

$$
TTS_{KL}(K,\alpha K) = -\alpha \frac{\varepsilon_K(1,\alpha)}{\varepsilon_L(1,\alpha)}. \tag{3.39}
$$

Quando a função de produção é homotética, então:

$$
TTS^f_{KL}(K,\alpha K) = -\alpha \frac{\varepsilon^h_K(1,\alpha)}{\varepsilon^h_L(1,\alpha)}, \tag{3.40}
$$

em que h é a função homogênea de grau 1 subjacente à função de produção homotética. Assim, para funções homotéticas de capital e trabalho, a TTS *em cada ponto ao longo de uma linha reta pela origem é inteiramente determinada pela relação capital-trabalho e pelas elasticidades-produto da combinação de uma unidade de capital e de trabalho. O que faz variar a* TTS *no caso das funções homotéticas de capital e trabalho é apenas a relação capital-trabalho.*

3.3 Curto Prazo versus Longo Prazo

Para entendermos o que são longo-prazo e curto-prazo, é preciso, primeiro, entender o que são custos fixos e como estes diferem dos custos históricos ou sunk costs.

O custo histórico é um custo do qual não se pode escapar. Por exemplo, assim que você compra um carro, você já não consegue vendê-lo pelo mesmo preço que pagou, caso o queira fazer meia hora depois de comprá-lo. Essa perda é inevitável. Assim, no momento em que, tendo comprado o carro meia hora antes, vem-lhe à cabeça a ideia de vendê-lo, o seu cálculo sobre os custos e benefícios da decisão de vender o carro meia hora depois de comprá-lo não deve incorporar essa perda, pois ela sequer é custo mais. Portanto, é preciso excluir do cálculo dos custos aqueles eventos que se configurem em meros custos históricos. A única função dos custos históricos é dar-lhe a conhecer como serão os custos envolvidos em decisões de natureza semelhante no futuro. Mais nada. Disse que os custos históricos não são verdadeiramente custos. De fato, como o custo histórico não pode ser recuperado, não existe qualquer opção alternativa ao evento que o subjaz e de cujo valor se possa desistir quando de sua ocorrência. Mesmo assim, por tradição e conveniência, usa-se o termo custo.

Vamos ver agora por que o conceito de custo fixo é mais complexo do que pode nos parecer à primeira vista. Não é verdade que alguns insumos estão literalmente fixos, ou melhor, há por trás dessa condição toda a economia relevante para o seu entendimento. O que acontece é que é muito mais caro variar esses insumos em um dado intervalo de tempo do que variar outros. É preciso entender que, se fosse mais barato ajustar o nível de utilização de um fator às novas condições de mercado do que mantê-lo fixo, então a firma certamente optaria pelo ajustamento.

Suponha, por exemplo, que você possui uma empresa, uma churrascaria, cuja administração é basicamente semanal. Toda segunda-feira de manhã você contrata a compra de picanha de um açougue, com entregas de carne programadas para quarta-feira e sábado. Imagine agora que na terça-feira você descobre que outro açougue vende a mesma picanha a um preço unitário muito mais baixo. Será muito caro você desistir do contrato já assinado, pois o açougue já incorreu em custos elevados para providenciar do abatedouro a carne encomendada. A carne já foi inspecionada e sua qualidade verificada. Se você quiser desistir do contrato assinado com o intuito de contratar o outro açougue, terá que incorrer em custos advocatícios elevados, pois o advogado terá que resolver todo o problema em um único dia, a fim de que você possa contratar o outro açougue em tempo hábil para atender a sua demanda no dia seguinte. Além disso, o açougue concorrente também tem seus próprios contratos para serem honrados. Para atender ao seu pedido em cima da hora, ele teria que desfazer-se de outros contratos e os custos disso seriam cobrados de você, pelo menos nesta sua primeira contratação. Observe quão importante é, para o cálculo correto dos custos e benefícios econômicos, que os direitos de propriedade estejam bem definidos. Os custos do ajuste imediato superam em muito os benefícios oriundos da compra de carne mais barata do outro açougue. Mantendo o insumo fixo, você perde somente o valor monetário correspondente à economia que faria caso comprasse a carne mais barata. Variando-o, você perde o valor monetário correspondente aos custos advocatícios e de rescisão de contratos. Poderíamos ainda acrescentar a estes últimos os custos de iniciar uma nova relação contratual com outro açougue, muitos dos quais de natureza informacional. Ao final, não vale a pena ajustar o fator de produção picanha nesse momento, pois seu ajuste seria muito mais caro do que a alternativa de mantê-lo fixo. Note que você tem a escolha de evitar esse custo de ajuste, caso não valha a pena. Se o ajuste valer a pena, você ajusta e, assim, não evita o custo, pois os benefícios serão maiores. Mantê-los fixos é uma decisão econômica, não uma restrição técnica, ou melhor, não inteiramente uma restrição técnica.

Verduras são outro fator de produção. A compra de verduras pela sua churrascaria é feita diariamente. Toda manhã você compra verduras frescas na feira. Qualquer variação nos preços das verduras (ou mesmo um aumento da demanda de verduras por parte dos clientes da churrascaria) pode ser levado em conta já no dia seguinte a fim de ajustar a quantidade de verduras em seu processo de produção às condições de mercado. Essa flexibilidade maior permite que o ajuste quase imediato desse fator de produção às condições de mercado valha a pena.

Certamente existem outros fatores de produção, como horas de trabalho dos garçons por semana, o capital, etc. Os ajustes nas quantidades desses fatores podem levar dois dias, quiçá três, alguns mais de uma semana e outros até apenas algumas horas. Cada fator de produção terá um tempo de ajuste necessário para que tal ajuste realmente valha a pena frente à alternativa de mantê-lo fixo.

A segunda coisa que deve ser mencionada para entendermos o que são longo-prazo e curto-prazo é o tempo de produção, ou seja, o papel do tempo de administração ou gerência que mencionei acima. Imagine que a tomada de decisão da firma, a escolha de seu plano de produção, desde a decisão sobre quais fatores usar, quanto de cada e quanto produzir, tomem um tempo considerado normal ou padrão na indústria a que sua firma pertence. Por exemplo, para um restaurante, uma semana, uma quinzena ou um mês. Para um produtor de navios, pode ser um par de anos. Tudo depende da indústria. Indústria é o conjunto de firmas que produzem certo bem mais ou menos homogêneo com tecnologias mais ou menos homogêneas.

Assim, entre um insumo estar absolutamente fixo e estar absolutamente variável existe, em verdade, um espectro contínuo de ajustabilidade. Eis a palavra-chave! Se você, para a sua churrascaria, demarcar uma linha de corte, por exemplo, três dias, então, e somente então, haverá sentido em dizer que um insumo é fixo enquanto outro é variável. Se o ajuste de um insumo pode ser feito em até três dias, pode-se considerar que o insumo é variável. Se forem necessários mais de três dias para que o ajuste seja feito e seja vantajoso, pode-se dizer que ele está fixo durante o período de comércio semanal da churrascaria. É evidente que o período de uma semana que venho usando no exemplo é fictício. Por favor, atente para a ideia. A linha de corte que separa o curto-prazo do longo-prazo é a resposta à pergunta: "Qual a menor extensão de tempo para que todos os fatores possam ser ajustados livremente a mudanças inesperadas de mercado?"

Tudo depende, então, do tipo de comércio ou produção, de quão razoável é ou não considerar tal ou qual linha de corte. Vemos, assim, que a classificação entre fatores fixos e variáveise, por conseguinte, entre custos fixos e variáveis é uma classificação bipolar e tem caráter meramente didático.

Uma característica importante do custo fixo é ele ser evitável, diferentemente dos custos históricos (sunk costs). Assim, o custo fixo não é um custo histórico. O fato de ser evitável sugere que o custo fixo não é tão fixo assim. Por essa razão, existem nomes alternativos que melhor descrevem a natureza econômica do custo fixo e do custo histórico. Há três linhas gerais (que denotarei por A, B e C) de nomes para esses custos identificáveis na literatura. Na falta de termos apropriados, farei a seguinte distinção: os custos variáveis (tal como aprendemos nos livros) continuarei a chamar de custos variáveis; já os demais custos chamarei de custos invariáveis, os quais, por natureza, podem, por sua vez, ser subdivididos em dois tipos: inevitáveis e evitáveis. Armen Alchian chama-os de custos invariantes. É importante conhecer essa distinção para não se incorrer em confusão semântica, pois os livros trazem nomes distintos para a mesma ideia.

Assim, enquanto um autor chama os custos invariáveis inevitáveis de custos fixos e os custos invariáveis evitáveis de custos quase-fixos (por exemplo, Hal Varian), outros os chamam, respectivamente, de sunk costs e custos fixos (por exemplo, Armen Alchian).

Como decorrência dessa simplificação, o próprio conceito de curto-prazo e longo-prazo revela também seu caráter simplificador e didático. Em verdade, existem vários prazos, e não apenas curto e longo. Deve ficar claro que curto prazo e longo prazo são correlatos semânticos da classificação simplificadora dos insumos em fixos (isto é, invariantes evitáveis) e variáveis. São, portanto, dois extremos didáticos de um espectro contínuo de prazos.

Na indústria a que a churrascaria pertence, digamos que o tempo de gerência normal é de uma semana e que todas as decisões, de todas as firmas na indústria, para a semana são tomadas na segunda-feira. Assim, qualquer ajuste a novas condições de mercado que não possa ser feito até o término da semana terá como elemento causador um conjunto de insumos que permanecerão fixos nesse período porque é mais barato que assim fiquem. Ao restringir a análise dessa decisão microeconômica ao próprio tempo de produção e tendo estabelecido a linha de corte, você deverá considerar, então, que a firma atua no curto-prazo. Se, ao contrário, seu horizonte de análise é mais longo e se todos os seus fatores de produção podem ser ajustados a novas condições de mercado, então você deverá considerar que a firma atua no longo-prazo. Cada indústria tem um tempo de produção “normal” e uma linha de corte “normal”. Uso “normal” entre aspas para enfatizar o sentido que Marshall dava ao termo: o sentido estatístico.

Vejamos os problemas. Em primeiro lugar, já sabendo que curto-prazo e longo-prazo são apenas dois extremos didáticos de um espectro contínuo de prazos, é fundamental entender que curto-prazo e longo-prazo são específicos da indústria. Cada indústria tem um curto-prazo e um longo-prazo padrão ou médio. Além disso, a linha de corte temporal pode depender da natureza mesma da variação das condições de mercado. Para certas mudanças, o ajuste pode ser mais fácil; para outras, não. A especificidade da dicotomia é mais complexa do que parece.

Em segundo lugar, dentro da mesma indústria, os tempos de produção de firmas diferentes não são concomitantes, as suas janelas de tempo não se sobrepõem perfeitamente. Não existe razão para supor que assim seja. Daí a importância de interpretar as noções de forma estatística, ou seja, distribucional. Na média, com certa variância, ajustes que tenham de ser feitos dentro de uma extensão de tempo ocorrem no curto-prazo e, para além disso, no longo-prazo.

Vê-se agora quão problemática é a análise macroeconômica ao agregar todas as distintas indústrias num único super-agente chamado setor produtivo e tirar conclusões de política econômica a partir da divisão entre curto-prazo e longo-prazo mediante a simples suposição de algum fator fixo numa função de produção agregada. Os resultados teóricos, embora coerentes com a sua fundamentação microeconômica no universo da modelagem matemática ou da análise lógica, têm uma grande chance de não serem coerentes com o fato de que, nesse ínterim, os agentes no nível microeconômico já tenham feito mudanças estruturais de preferências e tecnologias de tal magnitude e de tal qualidade que a realidade dos efeitos já não há de condizer com as hipóteses subjacentes do modelo macroeconômico sob o guante do ceteris paribus.

Essa crítica é grave e deve ser encarada de frente. Quem diz que a agregação macroeconômica é tão-somente a soma algébrica de quantidades microeconômicas não compreende muito bem o que diz. Hayek alertava para esses problemas da agregação. Os austríacos vão ao extremo de rejeitar a agregação como construto possível de uma análise econômica séria. Mises é um deles. Na teoria microeconômica mainstream, aceita-se alguma agregação. Harberger, por exemplo, amparado por Hicks e Pigou, é pioneiro na análise de bem-estar em mercados, quando, supondo ausência de efeito-renda, usa o fato de ser possível monetizar os excedentes de consumidores e produtores para somá-los algebricamente e assim mensurar as perdas de eficiência provenientes de distorções arbitrárias de preços. Mas ele sempre aplicou esses princípios dentro da indústria, no mercado de um bem, não para o agregado de todas elas. Ele, porém, nunca abusou das limitações da Microeconomia, porque ele as compreende. Sua justificativa sempre foi a mesma, a de que pelo menos assim se pode fazer alguma análise.

3.4 Maximização do Lucro

Uma hipótese básica sobre o comportamento individual da empresa na teoria da firma é que ela sempre escolherá um plano de produção mais lucrativo do conjunto de produção. Vamos derivar as funções de demanda por insumo e de oferta do produto, considerando um modelo de comportamento de maximização de lucro, juntamente com uma descrição das restrições de produção subjacentes.

O lucro econômico é definido como sendo a diferença entre a receita que uma empresa recebe e os custos que ela incorre. É importante entender que todos os custos (explícitos e implícitos) devem ser incluídos no cálculo do lucro. Tanto as receitas como os custos de uma empresa dependem das ações tomadas pela empresa. Podemos escrever a receita como uma função do nível de operações de algumas n ações, $R(a_1, \ldots, a_n)$, e custos como uma função desses mesmos n níveis de atividade, $C(a_1, \ldots, a_n)$, em que as ações podem estar em termos de nível de emprego de insumos ou nível de produção ou preços de produtos se a empresa tiver um poder de mercado para estabelecer os preços.

Uma suposição básica da maioria das análises econômicas do comportamento das empresas é que uma empresa atua de modo a maximizar seus lucros; isto é, uma empresa escolhe as ações $(a_1, \ldots, a_n)$ de modo a maximizar $R(a_1, \ldots, a_n) - C(a_1, \ldots, a_n)$. O problema de maximização do lucro que a empresa enfrenta pode então ser escrito como

$$\max_{a_1,\ldots,a_n} R(a_1, \ldots, a_n) - C(a_1, \ldots, a_n). \tag{3.41}$$

As condições de primeira ordem para ações ótimas interiores, $a^* = (a_1^*, \ldots, a_n^*)$, são caracterizadas pelas condições

$$\frac{\partial R(a^*)}{\partial a_i} = \frac{\partial C(a^*)}{\partial a_i}, \quad i = 1, \ldots, n, \tag{3.42}$$

A intuição por trás dessas condições é clara: se a receita marginal for maior do que o custo marginal, a firma pagaria para aumentar o nível da atividade; se a receita marginal fosse menor do que o custo marginal, a firma pagaria para diminuir o nível da atividade. Em geral, a receita é composta de duas partes: quanto uma empresa vende de vários produtos vezes o preço de cada produto. Os custos também são compostos de duas partes: quanto uma empresa usa de cada insumo vezes o preço de cada insumo.

O problema de maximização do lucro da empresa, portanto, reduz-se ao problema de determinar que preços deseja cobrar por seus produtos ou pagar por seus insumos e quais níveis de produtos e insumos deseja usar. Ao determinar sua política ótima, a empresa enfrenta dois tipos de restrições: restrições tecnológicas que são especificadas por conjuntos de produção e restrições de mercado que dizem respeito ao efeito de ações de outros agentes na empresa. Supõe-se que as empresas descritas no restante deste capítulo exibem o tipo mais simples de comportamento de mercado, a saber, o comportamento de tomador de preços. Cada firma assumirá os preços como dados. Assim, a empresa estará preocupada apenas em determinar os níveis de de produtos e insumos que maximizam seu lucro. Tal empresa é muitas vezes referida como uma empresa competitiva.

3.4.1 Escolha Ótima da Firma

Seja p um vetor de preços de insumos e produtos da empresa. O problema de maximização do lucro da empresa pode ser escrito como

$$\begin{aligned} \pi(p) &= \max py \\ \text{sujeito a} \quad & y \in Y. \end{aligned} \tag{3.43}$$

Observe que, como os produtos são mensurados como números positivos e os insumos são mensurados como números negativos, a função objetivo para esse problema são os lucros: receitas menos custos. A função $\pi(p)$, que nos dá o lucro máximo em função dos preços, é chamada de função lucro da firma.

1. Problema de maximização de curto prazo. Nesse caso, podemos definir a função lucro de curto prazo, também conhecida como a função lucro restrito:

$$\begin{aligned} \pi(p, z) &= \max py \\ \text{sujeito a} \quad & y \in Y(z). \end{aligned} \tag{3.44}$$

2. Se a empresa produz apenas um produto, a função lucro pode ser escrita como

$$\pi(p, w) = \max pf(x) - wx. \tag{3.45}$$

em que p é o preço (escalar) do produto, w é o vetor dos preços dos fatores e os dados são medidos pelo vetor (não negativo) $x = (x_1, \dots, x_n)$.

O valor de y que resolve o problema do lucro não é, em geral, único. Quando existe um plano de produção exclusivo, o plano de produção é chamado de função de produção líquida ou função de oferta líquida, a parte de insumo correspondente é chamada de função de demanda por insumo e o vetor de produto correspondente é chamado de função oferta do produtor. Veremos a partir da seguinte proposta que a convexidade rigorosa do conjunto de produção garantirá a unicidade do plano de produção ideal.

3.4.2 Maximização do Lucro: Condições de Primeira Ordem

O comportamento de maximização de lucro pode ser caracterizado por cálculo quando a tecnologia pode ser descrita por uma função de produção diferenciável. Por exemplo, as condições de primeira ordem para o problema de maximização do lucro, no caso de um único produto, com solução interior são

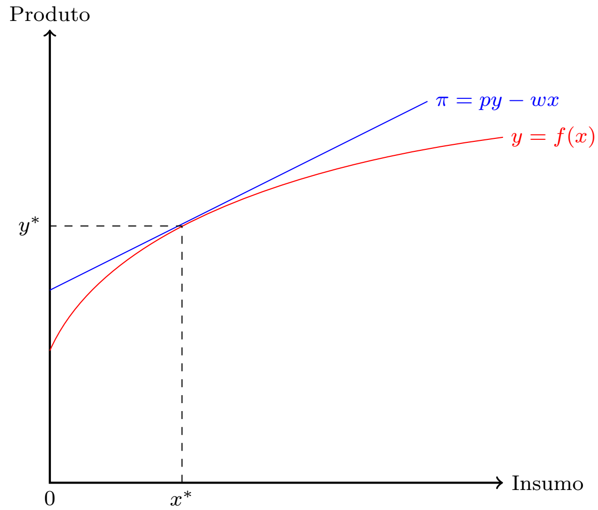
$$p\frac{\partial f(x^*)}{\partial x_i} = w_i, \quad i = 1, \dots, n, \tag{3.46}$$

Usando notação vetorial, podemos reescrever esta condição como

$$pDf(x^*) = w. \tag{3.47}$$

As condições de primeira ordem afirmam que o "valor marginal do produto de cada fator deve ser igual ao seu preço", ou seja, a receita marginal é igual ao custo marginal no plano de produção que maximiza o lucro. Esta condição de primeira ordem também pode ser exibida graficamente. Considere o conjunto de possibilidades de produção descrito na Figura 3.3. Neste caso bidimensional, os lucros são dados por $\pi = py - wx$. Os conjuntos de níveis dessa função para p e w fixos são linhas retas que podem ser representadas como funções da forma: $y = \frac{\pi}{p} + \left(\frac{w}{p}\right) x$. Aqui a inclinação da linha de isolucro dá o salário mensurado em unidades de produção, e o intercepto vertical nos dá o lucro mensurado em unidades de produção. No ponto de lucro máximo, a função de produção deve situar-se abaixo de sua linha tangente em x^*, isto é, deve ser "localmente côncava".

Figura 3.3 – Maximização do lucro



Semelhante aos argumentos da teoria do consumidor, as condições de cálculo derivadas acima só fazem sentido quando as variáveis de escolha podem ser variadas em uma vizinhança aberta da escolha ótima. As condições relevantes de primeira ordem que também incluem soluções de contorno (canto) são dadas pelas condições de Kuhn-Tucker:

$$p - \frac{\partial f(x)}{\partial x_i} - w_i \leqq 0 \quad \text{com igualdade se } x_i > 0. \tag{3.48}$$

Pode não existir um plano de produção que maximize os lucros quando uma tecnologia de produção exibe retornos constantes de escala ou retornos crescentes de escala. Por exemplo, considere o caso em que a função de produção é $f(x) = x$. Então, para $p > w$, não haverá plano maximizador de lucro. Fica claro, a partir desse exemplo, que a única posição não trivial de maximização de lucros para uma empresa de retornos constantes à escala é o caso de $p = w$ e lucros nulos.

3.4.3 Maximização do Lucro: Condições de Segunda Ordem

A condição de segunda ordem para a maximização do lucro é que a matriz das segundas derivadas da função de produção deve ser semi-definida negativa no ponto ótimo; isto é, a condição de segunda ordem requer que a matriz Hessiana

$$D^2 f(x^*) = \left(\frac{\partial^2 f(x^*)}{\partial x_i x_j} \right), \tag{3.49}$$

deve satisfazer a condição que $hD^2 f(x^*)h' \leqq 0$ para todos os vetores h. Geometricamente, a exigência de que a matriz Hessiana seja semi-definida negativa significa que a função de produção deve ser localmente côncava na vizinhança de uma escolha ótima. Anteriormente, temos a seguinte proposta.

Proposição 3.4.1. *Suponha que $f(x)$ é diferenciável e côncava em $\mathbb{R}^L_+$ e $(p, w) > 0$. Se $x > 0$ satisfaz as condições de primeira ordem dadas acima, então x é (globalmente) o plano de produção maximizador de lucros a preços (p, w).*

3.4.4 Elasticidade-Preço

A elasticidade-preço ela mede a variação percentual da quantidade demandada em relação a uma variação percentual do preço. Nada mais óbvio. Vamos ver um exemplo para entender este conceito de forma adequada.

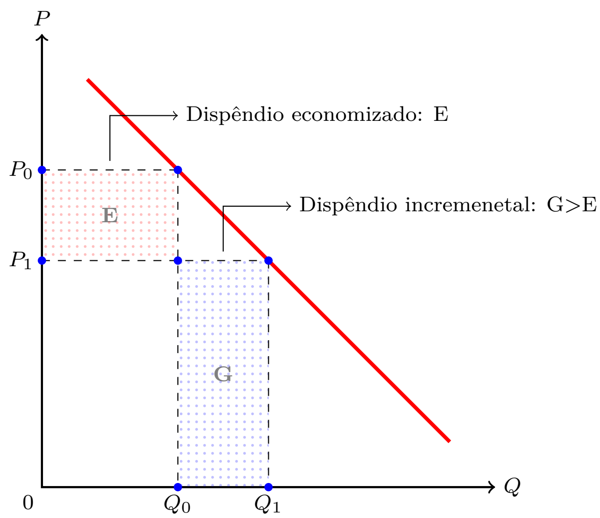
Para começar, imagine o vendedor de um certo bem que observa um volume de vendas a um dado preço. Ao testar o comportamento dos consumidores após uma pequena redução de preços, ele observa um aumento do seu volume de vendas, em decorrência da queda de preço, de tal magnitude que a receita aumenta. Diante desse cenário, ele percebe que uma pequena redução de preço proporcionar-lhe-á um aumento significativo da receita.

Antes que você pense que isso é óbvio, suponha que a \$100 o volume de vendas é de 400 unidades por mês, mas que a \$99 o volume é de 402 unidades. Então, a \$100 a receita é de \$40.000, mas a \$99, a receita é de \$39.798, ou seja, um faturamento menor. Portanto, a receita pode cair, mesmo com o aumento do volume.

O segredo da elasticidade-preço não está na comparação da variação percentual das quantidades consumidas em relação à variação percentual do preço. Esse é apenas o aspecto matemático superficial que surge a partir de algo mais fundamental. O segredo está na comparação absoluta dos dispêndios aos referidos preços. Essa comparação se apresenta idêntica tanto para os consumidores quanto para os ofertantes, pois se baseia nas observações empíricas dos dispêndios com o bem (do ponto de vista dos consumidores) ou das receitas de venda (do ponto de vista dos vendedores).

Quando o preço é P_0 e a quantidade demandada é Q_0 unidades, o gasto com o bem é P_0Q_0. Se o preço cai de P_0 para P_1 e a quantidade consumida sobe para Q_1, o gasto com o bem é P_1Q_1. Por um lado, economiza-se um dispêndio E por causa da queda do preço. Por outro, aumenta-se o gasto em G com o bem, em razão da quantidade maior demandada. Esses valores estão na Figura 3.4.

Figura 3.4 – Dispêndios decorrentes de uma mudança no preço



Note que $E = (-\Delta P)Q_0$ e $G = P_1(\Delta Q)$. Se $G > E$, então é porque a queda de preço provoca, por um lado, uma economia, mas um gasto maior por outro. Por que ocorre isso? Porque os consumidores, em termos agregados, gostam tanto do bem que a pequena queda de preço foi o gatilho para que se dispusessem a abrir mão de um valor maior dos seus outros recursos privados (na forma de consumo de outros bens) em troca do consumo maior do bem. Ora, $G > E$ significa que:

$$P_1(\Delta Q) > (-\Delta P)Q_0. \tag{3.50}$$

Se a queda de preço é pequena, então podemos substituir P_1 por P_0 e dividir tudo por P_0Q_0. Portanto:

$$\frac{\Delta Q}{Q_0} > \frac{-\Delta P}{P_0}. \tag{3.51}$$

Defina:

$$\Delta\%Q = \frac{\Delta Q}{Q_0} \tag{3.52}$$

$$\Delta\%P = \frac{\Delta P}{P_0} \tag{3.53}$$

$$\varepsilon = \frac{\Delta\%Q}{\Delta\%P}. \tag{3.54}$$

Observe que $\dfrac{-\Delta P}{P_0} = -\Delta\%P$. Então, como ΔP é negativo, na divisão inverte-se a desigualdade, de onde temos:

$$\varepsilon = \frac{\Delta\%Q}{\Delta\%P} < -1, \tag{3.55}$$

ou seja, a demanda pelo bem é preço-elástica: a quantidade demandada é muito sensível ao preço; a elasticidade é negativa, pois a quantidade demandada aumenta com a queda de preço, e em módulo ela é maior que 1. Decorre daí, analogamente, que, se $G < E$, então a demanda é preço-inelástica. Se, por fim, $G = E$, então a elasticidade-preço é unitária.

Chegamos, portanto, à seguinte interpretação alternativa da elasticidade-preço. Suponha que a um dado preço o consumidor gaste um montante de sua renda em um bem. Então:

1. A demanda por esse bem é preço-elástica se uma pequena queda do preço faz o consumidor aumentar a quantidade consumida em tal magnitude que o montante monetário dispendido no bem seja maior que antes. Em outras palavras, a demanda por esse bem é preço-elástica se a economia de gastos causada por uma pequena queda do preço é integralmente gasta na compra de mais unidades desse mesmo bem e se, além disso, o consumidor ainda sacrifica consumo de outros bens para gastar ainda mais no bem em questão.

2. A demanda pelo bem é preço-inelástica se uma pequena queda do preço faz o consumidor aumentar pouco a quantidade consumida, tão pouco a ponto de o montante monetário dispendido no bem ser menor que antes. Dito de outra forma, a demanda pelo bem é preço-inelástica se a economia causada por uma uma pequena queda do preço não é integralmente usada para comprar mais unidades desse bem: o consumidor usa parte dessa economia na compra de outros bens.

3. A demanda tem elasticidade-preço unitária se uma pequena queda do preço faz o consumidor aumentar a quantidade consumida sem alterar o montante monetário dispendido no bem, ou seja, toda economia causada por uma queda pequena do preço do bem é toda ela revertida na compra de unidades adicionais do mesmo bem.

Sabemos que a elasticidade-preço é uma propriedade local da curva de demanda. Porém, os próprios ofertantes podem captar a magnitude da elasticidade-preço da demanda que enfrentam mediante a observação empírica do volume de vendas a preços distintos, desde que próximos. Portanto, não é preciso que os ofertantes conheçam a curva de demanda para saberem se os seus consumidores têm pelo produto que demandam uma maior ou menor sensibilidade ao preço. Basta observar o faturamento a preços próximos.

Ao preço de \$100, o volume de vendas era de 400 unidades por mês. Ao preço de \$99 o volume era de 402 unidades. Assim, a \$100 a receita era de \$40.000, mas a \$99, a receita era de \$39.798, ou seja, um faturamento menor. Portanto, a receita cai, mesmo com o aumento do volume. Isso ocorre porque a demanda é preço-inelástica. Se à queda do preço de \$100 para \$99 correspondesse um aumento da quantidade demandada de 400 para 410 unidades, por exemplo, então a receita de venda subiria de \$40.000 para \$40.590. Isso ocorrerá se a demanda for preço-elástica. Se a quantidade vendida aumentasse de 400 para 404, então a receita seria a mesma. Neste caso, a elasticidade-preço é unitária.

Finalmente, o que significa a demanda ser preço-elástica? A um dado preço, o consumidor gasta determinado montante monetário no bem e o restante nos outros bens. Quando o preço cai, ele passa a gastar imediatamente um montante menor nesse bem. Isso representa uma economia de gastos imediata, realizada no preciso momento em que o consumidor compra o bem ao preço menor. Como ele deseja o bem em questão mais intensamente, então, ao adquirir as unidades adicionais desse bem em razão da queda do preço, o consumidor está disposto a abrir mão de outros bens, em variadas quantidades. Essa redução no consumo de outros bens se traduz em um valor monetário que ele decide economizar dos outros bens para gastar no bem cujo preço caiu. De fato, com a queda de preço do bem, os preços relativos entre esse bem e os demais se tornaram mais favoráveis ao bem em questão, alterando, assim, as taxas às quais o consumidor está disposto a trocar um conjunto de outros bens em certas quantidades pelo bem cujo preço caiu. Isso não se dá instantaneamente, pois o consumidor possui estoques de bens que precisam ser renovados apenas de tempos em tempos. Em outras palavras, a economia descrita por E no gráfico é imediata, mas o gasto adicional descrito por G toma ainda um certo tempo. O fato é que a demanda pelo bem será preço-elástica se o valor que ele economiza dos outros bens para gastar a mais no bem em questão (ao longo do tempo de ajuste) é maior do que a economia de gastos que ele teve com a queda de preço (no momento inicial da queda).

Se o consumidor desejasse o bem menos intensamente do que deseja os outros, então teríamos o efeito contrário: uma demanda preço-inelástica. Apesar da queda do preço do bem, ele prefere tão mais intensamente os outros bens (e não este) que, apesar da imediata redução de dispêndio em função da queda de preço, ele não está tão disposto assim a abrir mão dos outros bens para adquirir mais unidades do bem em questão. O valor gasto no bem na verdade diminui, pois ele prefere utilizar a economia imediata de riqueza nos outros bens, aumentando o consumo do bem em questão apenas um pouco, mas retirando dele parte do que gastava para gastar nos outros bens.

A elasticidade-preço, apesar de ser um conceito associado à demanda, não exige que se tenha conhecimento da curva de demanda propriamente dita e nem que não seja útil ao ofertante, que é quem enfrenta a demanda. Para qualquer agente, bastam as observações dos dispêndios a preços próximos (do ponto de vista do consumidor) ou, equivalentemente, das receitas de venda (do ponto de vista do vendedor).

3.4.5 Maximizando o Lucro: Elasticidade-Preço

Vamos usar o conceito de elasticidade-preço para mostrar como o lucro pode ser maximizado pela regra "receita marginal=custo marginal" sem que o empresário tenha qualquer função de lucro na cabeça. O conceito de elasticidade-preço é matemático. Porém, mediante um raciocínio econômico muito mais elementar e natural, devido a Armen Alchian, vamos ver que o empresário, ao agir, utiliza a elasticidade-preço da demanda sem necessidade de saber que está se valendo da elasticidade-preço e que, mesmo se soubesse, poderia estimá-la sem qualquer conhecimento da função de demanda Marshalliana do consumidor e muito menos de sua função de utilidade. Vamos em frente.

Suponha que ao preço de \$30 por entrada você vai ao cinema 40 vezes por ano, gastando, portanto, \$1.200 em cinema por ano. Se o preço da entrada cai pra \$25, então você teria uma economia de $200 se mantivesse a mesma taxa de consumo por ano, ou seja, se continuasse a ir ao cinema 40 vezes por ano.

O que você faz com essa economia? Se você gastar esses $200 em cinema, ou seja, comprar mais 8 entradas agora ao preço de $25, então a elasticidade-preço da sua demanda por cinema pode ser considerada como sendo unitária. A aproximação será tanto melhor quanto menor for a variação de preço. Neste caso, você continua gastando $1.200 em cinema por ano.

Se, porém, você resolver gastar apenas uma parte desses $200 em idas adicionais ao cinema, digamos, 6 idas a mais por ano, o que, ao preço de $25, lhe custarão $150, e destinar a parte restante, $50, pra gastar com outros bens, então a sua demanda por cinema é preço-inelástica. Neste caso, após a queda de preço você passa a gastar $1.150 em cinema por ano, menos que os $1.200 que você gastava antes.

Se, ao contrário, você usar os $200 economizados pra comprar mais 8 entradas ao preço de $25 cada e, além disso, sacrificar o consumo de outros bens, digamos, $50, pra comprar mais 2 entradas além daquelas 8 adicionais, então sua demanda por cinema é preço-elástica. Neste caso, você passa a gastar $1.250 em cinema por ano, mais que os $1.200 que você gastava quando o preço era $30.

O dono do cinema pode ter as mesmas estimativas com os mesmo dados observáveis: número de espectadores por ano a cada preço da entrada. A diferença é que a economia de $200 do consumidor é a perda de receita do vendedor em razão da queda do preço da entrada de $30 para $25. Similarmente, o dispêndio adicional do consumidor após a queda é a receita adicional do vendedor. Dessa forma, o vendedor tem a noção da receita incremental para uma dada variação de preço.

Imagine que o dono do cinema está na situação em que ao baixar o preço de $30 pra $25, a quantidade de entradas sobre de 40 pra 50 (aquelas 8 adicionais com a economia de $200 mais as outras 2, com $50), ou seja, demanda preço-elástica. Ao realizar essa variação de preço, ele tem então a noção da receita incremental, que é de $50 ($= 1250 - 1200$). Qual é o custo de oferecer essas 10 entradas a mais? Se o custo for, suponhamos, $35, então vale a pena. Ele continua fazendo essas experiências até o ponto em que a receita incremental quase não compensa o custo incremental.

O nosso dono do cinema maximizou seu lucro igualando receita marginal ao custo marginal.

Isso foi pra mostrar que podemos avaliar a elasticidade-preço da demanda por um certo bem (isto é, dizer se é preço-elástica, preço-inelástica ou de elasticidade unitária, ao menos aproximadamente) sem necessidade de termos conhecimento da curva de demanda e muito menos da função de utilidade e que essa avaliação pode ser feita tanto pelo consumidor quanto pelo vendedor a partir de dados observáveis: as quantidades transacionadas a preços distintos (e próximos). E nem é preciso que nos restrinjamos à definição de elasticidade como razão entre variações percentuais. Basta ao empresário fazer o seguinte raciocínio. “Se a um determinado preço eu vendo tantas unidades e se o preço cai, quanto de receita eu perco se vendesse essa mesma quantidade ao preço mais baixo? Ora, se o preço cai, as pessoas, por outro lado, vão comprar uma quantidade maior. Quanto de receita eu ganho, ao preço mais baixo, em decorrência dessa quantidade adicional? Comparo então a receita perdida por um lado com a receita ganha pelo outro. Se o ganho líquido é positivo, então temos uma receita incremental positiva. Mas não é só isso! Posso ainda dizer que a demanda pelo meu produto é preço-elástica. Vale a pena baixar o preço? Depende, eu tenho que saber quanto vai me custar oferecer essas unidades adicionais. Todo empresário sabe quanto isso vai custar, mesmo que seja em termos esperados. Se esse custo incremental é menor do que a receita incremental, então vale a pena. Sigo fazendo isso regularmente. Vai chegar um momento em que o preço que eu adoto é tal que, a esse preço, a receita incremental é quase igual ao custo incremental".

3.5 Função Lucro

Dado qualquer conjunto de produção Y, vimos como calcular a função de lucro, $\pi(p)$, que nos dá o lucro máximo atingível a preços p. A função lucro possui várias propriedades importantes que seguem diretamente de sua definição. Essas propriedades são muito úteis para analisar o comportamento de maximização de lucro.

3.5.1 Propriedades da Função Lucro

As propriedades dadas abaixo seguem apenas a partir da hipótese de maximização do lucro. Nenhuma suposição sobre convexidade, monotonicidade ou outros tipos de regularidade são necessárias.

Proposição 3.5.1 (Propriedades da Função Lucro)**.** *São propriedades da função lucro:*

- *Não-decrescente nos preços dos produtos e não-crescente nos preços dos insumos. Se* $p'_i \geqq p_i$ *para todos os produtos e* $p'_j \leqq p_j$ *para todos os insumos, então* $\pi(p') \geqq \pi(p)$.
- *Homogênea de grau 1 em p:* $\pi(tp) = t\pi(p)$ *para todo* $t \geqq 0$.
- *Convexa em p: seja* $p'' = tp + (1-t)p'$ *para* $0 \leqq t \leqq 1$. *Então,* $\pi(p'') \leqq t\pi(p) + (1-t)\pi(p')$.
- *Contínua em p: a função* $\pi(p)$ *é contínua, pelo menos quando* $\pi(p)$ *é bem definido e* $p_i > 0$ *para* $i = 1, \ldots, n$.

3.5.2 Derivando a Função de Oferta a partir da Função Lucro

Se nos é dada a função de oferta $y(p)$, é fácil calcular a função de lucro. Nós apenas substituímos na definição de lucro para encontrar $\pi(p) = py(p)$. Suponha que, em vez disso, tenhamos a função lucro e nos é pedido que encontremos as funções de oferta. Como isso pode ser feito? Acontece que há uma maneira muito simples de resolver esse problema: basta diferenciar a função lucro. A prova de que isso funciona é o conteúdo da próxima proposição.

Proposição 3.5.2 (Lema de Hotelling)**.** *Seja* $y_i(p)$ *a função de oferta do bem* i. *Então,*

$$y_i(p) = \frac{\partial \pi(p)}{\partial p_i}, \quad i = 1, \ldots, n, \tag{3.56}$$

assumindo que a derivada existe e que $p_i > 0$.

Demonstração. Suponha que y^* seja um vetor de produto que maximize o lucro a preços p^*. Então defina a função

$$g(p) = \pi(p) - py^*. \tag{3.57}$$

Claramente, o plano de produção que maximiza o lucro a preços p será sempre pelo menos tão lucrativo quanto o plano de produção y^*. No entanto, o plano y^* será um plano de maximização de lucro a preços p^*, de modo que a função g atinja um valor mínimo de 0 em p^*. As suposições sobre os preços implicam que isso é um mínimo interno.

As condições de primeira ordem para um mínimo, então, implicam que

$$\frac{\partial g(p^*)}{\partial p_i} = \frac{\partial \pi(p^*)}{\partial p_i} - y_i^* = 0, \quad i = 1, \ldots, n. \tag{3.58}$$

Como isso é verdade para todas as escolhas de p^*, a demonstração está satisfeita. ■

3.6 Minimização de Custo

Uma importante implicação da empresa em escolher um plano de produção que maximiza o lucro é que não há maneira de produzir as mesmas quantidades de produto com um custo total de insumos menor. Assim, a minimização de custos é uma condição necessária para a maximização do lucro. Essa observação nos leva a um estudo independente da minimização de custos da empresa. O problema é de interesse por várias razões. Primeiro, isso nos leva a vários resultados e construções tecnicamente muito úteis. Em segundo lugar, enquanto a empresa for uma tomadora de preço em seu mercado de insumos, os resultados decorrentes da minimização de custos continuarão válidos, quer o mercado de produção seja ou não competitivo e se a empresa considera ou não o preço do produto como determinado. Terceiro, quando o conjunto de produção exibe retornos de escala não decrescentes, a função custo e os vetores de otimização do problema de minimização de custos, que mantêm os níveis de produto fixos, são mais bem comportados do que a função lucro.

Para ser concreto, concentramos nossa análise no caso de um produto único. Assumimos que as empresas são perfeitamente competitivas em seus mercados de insumos e, portanto, enfrentam preços fixos. Seja $w = (w_1, w_2, \ldots, w_n)$ um vetor dos preços de mercado predominantes nos quais a empresa pode comprar insumos $x = (x_1, x_2, \ldots, x_n)$.

3.6.1 Minimização de Custos: Condições de Primeira Ordem

Vamos considerar o problema de encontrar uma maneira de minimizar custos para produzir um determinado nível de produção:

$$\min_x wx$$

$$\text{sujeito a} \quad f(x) \geqq y. \tag{3.59}$$

Analisamos esse problema de minimização restrita usando a função Lagrangeana:

$$L(\lambda, x) = wx - \lambda[f(x) - y]. \tag{3.60}$$

em que a função de produção f é diferenciável e λ é o multiplicador de Lagrange. As condições de primeira ordem que caracterizam uma solução interior x^* são

$$w_i - \lambda \frac{\partial f(x^*)}{\partial x_i} = 0, \quad i = 1, \ldots, n \tag{3.61}$$

$$f(x^*) = y, \tag{3.62}$$

ou em notação vetorial,

$$w = \lambda Df(x^*). \tag{3.63}$$

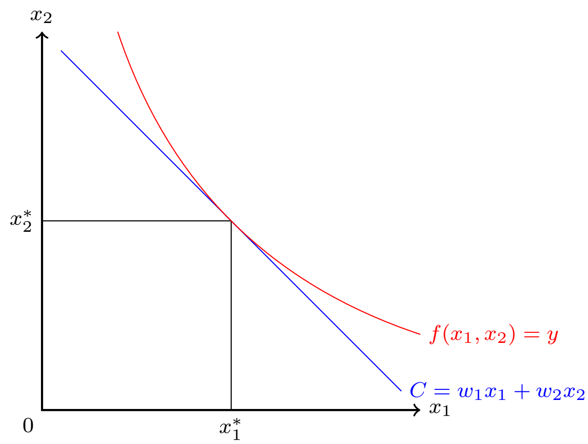
Podemos interpretar estas condições de primeira ordem, dividindo a j-ésima condição pela i-ésima condição para obter

$$\frac{w_i}{w_j} = \frac{\frac{\partial f(x^*)}{\partial x_i}}{\frac{\partial f(x^*)}{\partial x_j}}, \quad i, j = 1, \ldots, n, \tag{3.64}$$

o que significa que a taxa marginal de substituição técnica do insumo i pelo insumo j é igual à taxa econômica de substituição do insumo i pelo insumo j quando da cesta de minimização de custos.

Essa condição de primeira ordem também pode ser representada graficamente. Na Figura 3.5, as linhas curvas representam as isoquantas e as linhas retas representam as curvas de custo constantes. Quando y é fixo, o problema da empresa é encontrar um ponto de minimização de custo em uma dada isoquanta. É claro que tal ponto será caracterizado pela condição de tangência de que a inclinação da curva de custo constante deve ser igual à inclinação da isoquanta.

Figura 3.5 – Minimização do custo



Sabe-se que uma função contínua alcança um valor mínimo e máximo em um conjunto fechado e limitado. A função objetivo wx é certamente uma função contínua e o conjunto $V(y)$ é um conjunto fechado por hipótese. Tudo o que precisamos estabelecer é que podemos restringir nossa atenção a um subconjunto limitado de $V(y)$. Mas isso é fácil. Basta escolher um valor arbitrário de x, digamos x'. Claramente, a cesta com o insumo de custo mínimo deve ter um custo menor que wx'. Portanto, podemos restringir nossa atenção ao subconjunto $\{x \in V(y) \colon wx \leq wx'\}$, que certamente será um subconjunto limitado, desde que $w > 0$. Assim, a cesta de insumos de minimização de custos sempre existe.

3.6.2 Minimização de Custos: Condições de Segunda Ordem

Novamente, como o problema de otimização restrita do consumidor, as condições de primeira ordem acima são apenas condições necessárias para um ótimo local. No entanto, essas condições necessárias de primeira ordem são de fato suficientes para um ótimo global quando uma função de produção é quase côncava, o que é apresentado na seguinte proposição.

Proposição 3.6.1. *Suponha que $f(x) \colon \mathbb{R}^n_+ \to \mathbb{R}$ é diferenciável e quase-côncava no $\mathbb{R}^n_+$ e $w > 0$. Se $(x, \lambda) > 0$ satisfaz as condições de primeira ordem, então x resolve o problema de minimização de custos aos preços w.*

Para cada escolha de w e y, haverá alguma escolha de x^* que minimiza o custo de produção de unidades y de produto. Nós chamaremos a função que nos dá esta escolha ótima de função de demanda por insumo condicional e a escreveremos $x(w, y)$. Observe que as demandas de fatores condicionais dependem do nível de produto produzido, bem como dos preços dos fatores. A função custo é o custo mínimo nos preços dos fatores w e no nível de produto y, isto é, $c(w, y) = wx(w, y)$.

Exemplo 3.6.1 (Função Custo para Tecnologia Cobb-Douglas)**.** *Considere o problema de minimização de custo*

$$c(w, y) = \min_{x_1, x_2} w_1x_1 + w_2x_2$$

$$\text{tal que } Ax_1^{\alpha}x_2^{\beta} = y \tag{3.65}$$

Resolvendo a restrição para x_2, este problema é equivalente a

$$\min_{x_1} w_1x_1 + w_2A^{-\frac{1}{\beta}}y^{\frac{1}{\beta}}x_1^{-\frac{\alpha}{\beta}} \tag{3.66}$$

A condição de primeira ordem é

$$w_1 - \frac{\alpha}{\beta}w_2A^{-\frac{1}{\beta}}y^{\frac{1}{\beta}}x_1^{-\frac{(\alpha+\beta)}{\beta}} = 0 \tag{3.67}$$

que nos dá a função demanda condicional pelo insumo x_1:

$$x_1(w_1, w_2, y) = A^{-\frac{1}{(\alpha+\beta)}}\left(\frac{\alpha w_2}{\beta w_1}\right)^{\frac{\beta}{(\alpha+\beta)}}y^{\frac{1}{(\alpha+\beta)}} \tag{3.68}$$

A função demanda condicional pelo insumo x_2 é dada por:

$$x_2(w_1, w_2, y) = A^{-\frac{1}{(\alpha+\beta)}}\left(\frac{\beta w_1}{\alpha w_2}\right)^{\frac{\alpha}{(\alpha+\beta)}}y^{\frac{1}{(\alpha+\beta)}} \tag{3.69}$$

Portanto, a função custo é:

$$\begin{aligned}
c(w_1, w_2, y) &= w_1x_1(w_1, w_2, y) + w_2x_2(w_1, w_2, y) \\
&= w_1A^{-\frac{1}{(\alpha+\beta)}}\left(\frac{\alpha w_2}{\beta w_1}\right)^{\frac{\beta}{(\alpha+\beta)}}y^{\frac{1}{(\alpha+\beta)}} + w_2A^{-\frac{1}{(\alpha+\beta)}}\left(\frac{\beta w_1}{\alpha w_2}\right)^{\frac{\alpha}{(\alpha+\beta)}}y^{\frac{1}{(\alpha+\beta)}} \\
&= \left(\frac{yw_1^{\alpha}w_2^{\beta}}{A}\right)^{\frac{1}{(\alpha+\beta)}}\left[\left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{(\alpha+\beta)}} + \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{(\alpha+\beta)}}\right]
\end{aligned} \tag{3.70}$$

Como estática comparativa, temos:

- *Progresso tecnológico e desemprego*

$$\frac{\partial x_1^*}{\partial A} = -\frac{1}{(\alpha+\beta)A}x_1^* < 0 \tag{3.71}$$

$$\frac{\partial x_2^*}{\partial A} = -\frac{1}{(\alpha+\beta)A}x_2^* < 0 \tag{3.72}$$

- *Demanda por insumos é crescente com o nível de produto*

$$\frac{\partial x_1^*}{\partial y} = \frac{1}{(\alpha+\beta)y} x_1^* > 0 \tag{3.73}$$
$$\frac{\partial x_2^*}{\partial y} = \frac{1}{(\alpha+\beta)y} x_2^* > 0 \tag{3.74}$$

- *Demanda por insumos é decrescente com o preço dos insumos*

$$\frac{\partial x_1^*}{\partial w_1} = -\frac{\beta}{(\alpha+\beta)}\left(\frac{1}{w_1}\right) x_1^* < 0 \tag{3.75}$$
$$\frac{\partial x_2^*}{\partial w_2} = -\frac{\alpha}{(\alpha+\beta)}\left(\frac{1}{w_2}\right) x_2^* < 0 \tag{3.76}$$

- *Demanda por insumos é crescente com o preço dos outros insumos*

$$\frac{\partial x_1^*}{\partial w_2} = \frac{\alpha}{(\alpha+\beta)}\left(\frac{1}{w_2}\right) x_1^* > 0 \tag{3.77}$$
$$\frac{\partial x_2^*}{\partial w_1} = \frac{\beta}{(\alpha+\beta)}\left(\frac{1}{w_1}\right) x_2^* > 0 \tag{3.78}$$

Exemplo 3.6.2 (Função Custo para Tecnologia CES)**.** *Considere o problema de minimização de custo*

$$c(w,y) = \min_{x_1,x_2} w_1x_1 + w_2x_2$$
$$\text{tal que } x_1^\rho + x_2^\rho = y^\rho \tag{3.79}$$

As condições de primeira ordem:

$$w_1 - \lambda\rho x_1^{\rho-1} = 0 \tag{3.80}$$
$$w_2 - \lambda\rho x_2^{\rho-1} = 0 \tag{3.81}$$
$$x_1^\rho + x_2^\rho = y^\rho \tag{3.82}$$

Resolvendo as duas primeiras condições de primeira ordem, temos:

$$x_1^\rho = w_1^{\frac{\rho}{(\rho-1)}} (\lambda\rho)^{-\frac{\rho}{(\rho-1)}} \tag{3.83}$$

$$x_2^{\rho} = w_2^{\frac{\rho}{(\rho-1)}} (\lambda\rho)^{-\frac{\rho}{(\rho-1)}} \tag{3.84}$$

Substituindo estes resultados na função de produção, obtemos:

$$(\lambda\rho)^{-\frac{\rho}{(\rho-1)}} \left(w_1^{\frac{\rho}{(\rho-1)}} + w_2^{\frac{\rho}{(\rho-1)}} \right) = y^{\rho} \tag{3.85}$$

Resolvendo para $(\lambda\rho)^{-\frac{\rho}{\rho-1}}$ *e substituindo em* (3.83) *e* (3.84)*, obtemos as demandas condicionais pelos insumos:*

$$x_1(w_1, w_2, y) = w_1^{\frac{1}{(\rho-1)}} \left[w_1^{\frac{\rho}{(\rho-1)}} + w_2^{\frac{\rho}{(\rho-1)}} \right]^{-\frac{1}{\rho}} y \tag{3.86}$$

$$x_2(w_1, w_2, y) = w_2^{\frac{1}{(\rho-1)}} \left[w_1^{\frac{\rho}{(\rho-1)}} + w_2^{\frac{\rho}{(\rho-1)}} \right]^{-\frac{1}{\rho}} y \tag{3.87}$$

Substituindo essas funções na definição da função custo, chegamos a:

$$\begin{aligned} c(w_1, w_2, y) &= w_1 x_1(w_1, w_2, y) + w_2 x_2(w_1, w_2, y) \\ &= y \left[w_1^{\frac{\rho}{(\rho-1)}} + w_2^{\frac{\rho}{(\rho-1)}} \right] \left[w_1^{\frac{\rho}{(\rho-1)}} + w_2^{\frac{\rho}{(\rho-1)}} \right]^{-\frac{1}{\rho}} \\ &= y \left[w_1^{\frac{\rho}{(\rho-1)}} + w_2^{\frac{\rho}{(\rho-1)}} \right]^{\frac{(\rho-1)}{\rho}} \end{aligned} \tag{3.88}$$

Exemplo 3.6.3 (Função Custo para Tecnologia Leontief)**.** *Suponha que* $f(x_1, x_2) = \min\{\alpha x_1, \beta x_2\}$*. Como sabemos que a empresa não desperdiçará nenhum insumo com um preço positivo, a empresa deve operar em um ponto em que* $y = \alpha x_1 = \beta x_2$*. Portanto, se a empresa deseja produzir* y *unidades de produção, ela deve usar* $\frac{y}{\alpha}$ *unidades do insumo* 1 *e* $\frac{y}{\beta}$ *unidades do insumo* 2 *, independentemente dos preços dos insumos. Assim, a função custo é dada por*

$$c(w_1, w_2, y) = \left(w_1 \frac{y}{\alpha} + w_2 \frac{y}{\beta} \right)$$

$$= y\left(\frac{w_1}{\alpha} + \frac{w_2}{\beta}\right) \tag{3.89}$$

Exemplo 3.6.4 (Função Custo para Tecnologia Linear)**.** *Suponha que $f(x_1, x_2) = \alpha x_1 = \beta x_2$, de modo que os fatores 1 e 2 sejam substitutos perfeitos. Como será a função custo? Como os dois produtos são substitutos perfeitos, a empresa usará o que for mais barato. Assim, a função custo terá a forma $c(w_1, w_2, y) = \min\left\{\frac{w_1}{\alpha}, \frac{w_2}{\beta}\right\} y$. Nesse caso, a resposta ao problema de minimização de custos geralmente envolve uma solução de canto: um dos dois fatores será usado em uma quantidade zero. É fácil ver a resposta para esse problema específico comparando a inclinação relativa da linha de isocusto e da isoquanta. Se $\frac{\alpha}{\beta} < \frac{w_1}{w_2}$, a firma usa apenas x_2 e a função custo é dada por $c(w_1, w_2, y) = w_2 x_2 = w_2 \frac{y}{\beta}$. Se $\frac{\alpha}{\beta} > \frac{w_1}{w_2}$, a firma usa apenas x_1 e a função custo é dada por $c(w_1, w_2, y) = w_1 x_1 = w_1 \frac{y}{\alpha}$.*

3.7 Função Custo

A função custo mede o custo mínimo de produção de um dado nível de produção para alguns preços de fatores fixos. Como tal, resume informações sobre as opções tecnológicas disponíveis para as empresas. Acontece que o comportamento da função custo pode nos dizer muito sobre a natureza da tecnologia da empresa. A seguir, primeiro investigaremos o comportamento da função custo $c(w, y)$ em relação aos seus argumentos de preço e quantidade. Em seguida, definimos algumas funções relacionadas, a saber, as funções de custo médio e marginal.

3.7.1 Propriedades da Função Custo

Você pode ter notado algumas semelhanças aqui com a teoria do consumidor. Essas semelhanças são, na verdade, exatas quando se compara a função custo com a função dispêndio. De fato, considere suas definições.

$$\text{Função dispêndio: } e(p, u) \equiv \min_{x \in \mathbb{R}^n_+} px \text{ tal que } u(x) \geq u \tag{3.90}$$

$$\text{Função custo: } c(w, y) \equiv \min_{x \in \mathbb{R}^n_+} wx \text{ tal que } f(x) \geq y \tag{3.91}$$

Matematicamente, os dois problemas de otimização são idênticos. Consequentemente, para cada teorema que provamos sobre as funções dispêndio, existe um teorema equivalente para as funções de custo. Vamos declarar esses resultados aqui, mas não precisamos prová-los. Suas provas são idênticas àquelas dadas para a função dispêndio.

Proposição 3.7.1 (Propriedades da Função Custo)**.** *Suponha que a função de produção f seja contínua e estritamente crescente. Então a função custo tem as seguintes propriedades:*

- *$c(w, y)$ é não-decrescente em w.*
- *$c(w, y)$ é homogênea de grau 1 em w.*

- *$c(w, y)$ é côncava em w.*
- *$c(w, y)$ é contínua em w para $w > 0$.*
- *Para todo $w > 0$, $c(w, y)$ é estritamente crescente em y.*
- *Lema de Shephard: se $x(w, y)$ é a cesta que minimiza os custos necessários para produzir o nível y de produtos, aos preços w, então $x_i = \dfrac{\partial c(w, y)}{\partial w_i}$ para $i = 1, \ldots, n$, assumindo que a derivada existe e que $x_i > 0$.*

3.7.2 Custo Médio e Custo Marginal

Vamos considerar a estrutura da função custo. Observe que a função custo sempre pode ser expressa simplesmente como o valor da demanda condicional por insumos:

$$c(w, y) \equiv wx(x, y). \tag{3.92}$$

No curto prazo, alguns dos insumos de produção são fixados em níveis predeterminados. Seja x_f o vetor de insumos fixos, x_v, o vetor de insumos variáveis, e divida w em $w = (w_v, w_f)$, os vetores de preços dos insumos variáveis e fixos. As funções de demanda dos insumos de curto prazo geralmente dependem de x_f, então nós as escrevemos $x_v(w, y, x_f)$. Então a função custo de curto prazo pode ser escrita como

$$c(w, y, x_f) = w_v x_v(w, y, x_f) + w_f x_f. \tag{3.93}$$

O termo $w_v x_v(w, y, x_f)$ é chamado de custo variável de curto prazo (CVCP), e o termo $w_f x_f$ é o custo fixo (CF). Podemos definir vários conceitos de custos derivados dessas unidades básicas:

$$\text{Custo total de curto prazo} = CTCP = w_v x_v(w, y, x_f) + w_f x_f, \tag{3.94}$$

$$\text{Custo médio de curto prazo} = CMCP = \frac{c(w, y, x_f)}{y}, \tag{3.95}$$

$$\text{Custo variável médio de curto prazo} = CVMCP = \frac{w_v x_v(w, y, x_f)}{y}, \tag{3.96}$$

$$\text{Custo fixo médio de curto prazo} = CFMCP = \frac{w_f x_f}{y}, \tag{3.97}$$

$$\text{Custo marginal de curto prazo} = CMgCP = \frac{\partial c(w, y, x_f)}{\partial y}. \tag{3.98}$$

Quando todos os fatores são variáveis, a empresa otimizará na escolha de x_f. Portanto, a função de custo de longo prazo depende apenas dos preços dos fatores e do nível de produção, conforme indicado anteriormente. Podemos expressar essa função de longo prazo em termos da função de custo de curto prazo da seguinte maneira. Seja $x_f(w, y)$ a escolha ótima dos fatores fixos, e seja $x_v(w,y) = x_v(w,y,x_f(w,y))$ a escolha ótima de longo prazo dos fatores variáveis. Então a função custo de longo prazo pode ser escrita como

$$c(w,y) = w_v x_v(w,y) + w_f x_f(w,y) = c(w,y,x_f(w,y)). \tag{3.99}$$

De forma similar, podemos definir as funções custo médio e custo marginal de longo prazo, respectivamente, como:

$$\text{Custo médio de longo prazo} = CMLP = \frac{c(w,y)}{y}, \tag{3.100}$$

$$\text{Custo marginal de longo prazo} = CMgLP = \frac{\partial c(w,y)}{\partial y}. \tag{3.101}$$

Observe que o “custo médio de longo prazo” é igual a “custo variável médio de longo prazo”, já que todos os custos são variáveis no longo prazo; pelo mesmo motivo, os “custos fixos de longo prazo” são zero.

Exemplo 3.7.1 (Função Custo de Curto Prazo do Tipo Cobb-Douglas)**.** *Suponha que o segundo fator em uma tecnologia Cobb-Douglas seja restrito para operar em um nível k. Então o problema de minimização de custos é*

$$\begin{aligned} &\min_{x_1} w_1 x_1 + w_2 k \\ &\textit{tal que } y = x_1^{\alpha} k^{1-\alpha} \end{aligned} \tag{3.102}$$

Resolvendo a restrição para x_1 como uma função de y e k, obtemos:

$$x_1 = \left(yk^{\alpha-1}\right)^{\frac{1}{\alpha}} \tag{3.103}$$

Portanto, a função custo é dada por:

$$c(w_1,w_2,y,k) = w_1 \left(yk^{\alpha-1}\right)^{\frac{1}{\alpha}} + w_2 k \tag{3.104}$$

Da expressão do custo podemos obter:

$$CMCP = w_1 \left(\frac{y}{k}^{\frac{(1-\alpha)}{\alpha}}\right)^{\frac{1}{\alpha}} + w_2 \frac{k}{y} \tag{3.105}$$

$$CVMCP = w_1 \left(\frac{y}{k} \frac{(1-\alpha)}{\alpha} \right)^{\frac{1}{\alpha}} \tag{3.106}$$

$$CFMCP = w_2 \frac{k}{y} \tag{3.107}$$

$$CMgCP = \frac{w_1}{\alpha} \left(\frac{y}{k} \frac{(1-\alpha)}{\alpha} \right)^{\frac{1}{\alpha}} \tag{3.108}$$

Se a função de produção exibe retornos constantes de escala, fica intuitivamente claro que a função custo deve exibir custos lineares no nível do produto: se você quiser produzir duas vezes mais, custará o dobro do valor. Essa intuição é verificada na seguinte proposição.

Proposição 3.7.2. *Se a função de produção exibe retornos constantes de escala, a função custo pode ser escrita como* $c(w, y) = yc(w, 1)$.

3.7.3 Custo Incremental

Considerando o caso da firma produzir um único produto, temos que $c(y)$ é o custo mínimo para produzir y. O custo médio de produzir y, e que muitas vezes é chamado de custo unitário de produzir y, é definido por:

$$CM(y) = \frac{c(y)}{y}. \tag{3.109}$$

Outro importante conceito estudado no curso de graduação é o Custo Marginal (MC), definido por:

$$CMg(y) = \frac{\partial c(y)}{\partial y} = \lim_{\Delta y \to 0} \frac{c(y + \Delta y)}{\Delta y}. \tag{3.110}$$

A noção mais elementar para ambas as definições é o chamado Custo Incremental (CI), o qual proporciona um link natural entre o custo marginal e o custo médio.

Quanto mais custa produzir um adicional de y'?

Como $c(y + y')$ é o custo de produzir $y + y'$ e $c(y)$ é o custo de produzir y, o custo adicional para produzir uma produção adicional de y' é $c(y + y') - c(y)$. Assim, a resposta à pergunta anterior depende dos valores de y e y'. Além disso, y' não precisa ser muito pequeno como no caso da mensuração do custo marginal. Podemos definir o Custo Incremental (CI) como:

$$CI(y'|y) = c(y + y') - c(y). \tag{3.111}$$

Veja que esta definição não requer que $c(\cdot)$ seja diferenciável e contínua. O Custo Incremental Médio (CIM), por sua vez, é definido por:

$$CIM(y+y') = \frac{CI(y'|y)}{y'} = \frac{c(y+y')-c(y)}{y'}. \tag{3.112}$$

É interessante notar que o CM aumenta se e somente se o CIM aumenta mais que a média. Ou seja, formalmente temos que, para y, $y' > 0$ (teorema fundamental do CIM):

$$CIM(y'|y) > CM(y) \iff CM(y+y') > CM(y) \tag{3.113}$$
$$CIM(y'|y) = CM(y) \iff CM(y+y') = CM(y) \tag{3.114}$$
$$CIM(y'|y) < CM(y) \iff CM(y+y') < CM(y) \tag{3.115}$$

Demonstração. Veja que

$$\begin{aligned} CIM(y'|y) > CM(y) &\iff \frac{c(y+y')-c(y)}{y'} > \frac{c(y)}{y} \\ &\iff yc(y+y') - yc(y) > y'c(y) \\ &\iff yc(y+y') > (y+y')c(y) \\ &\iff \frac{c(y+y')}{y+y'} > \frac{c(y)}{y} \\ &\iff CM(y+y') > CM(y). \end{aligned} \tag{3.116}$$

A prova é a mesma para os dois outros casos. ∎

A conexão entre CIM e CMg é mais direta. Se consideramos y' como sendo um valor bem pequeno, o CIM se aproxima do CMg. Ou seja:

$$\lim_{y' \to 0} CIM(y+y') = CMg(y). \tag{3.117}$$

Também temos que:

$$CI(y'|y) = \int_{y}^{y+y'} CM(q)dq, \tag{3.118}$$

pois o lado direito é a integral definida que resulta em $c(y+y')-c(y)$. Visto desta maneira, temos:

$$CIM(y'|y) = \frac{\int_{y}^{y+y'} CM(q)dq}{y'}. \tag{3.119}$$

Apesar da relação estreita entre CIM e CMg, o teorema fundamental do CIM não é válido quando $y' \to 0$. Em outras palavras, o teorema só é válido para um valor finito de y', e não requer um y' infinitesimal. Ou seja, considere a primeira versão do teorema para o CMg: $CIM(y'|y) > CM(y) \iff CM(y+y') > CM(y)$. Se permitimos que $y' \to 0$, o lado esquerdo da desigualdade se torna $CMg(y) > CM(y)$, e o lado direito diz que o CM está aumentando em y. Assim, é claro que $CMg(y) > CM(y)$ implica que CM é crescente em y. O problema é que o inverso não é tão direto assim, ou seja, não podemos concluir que $CMg(y) > CM(y)$ a partir do CM crescente em y. Para ser preciso, mesmo se CM é crescente em y, $CMg(y) = CM(y)$ pode ocorrer. De modo similar, CM pode ser decrescente em y, e ambas as situações $CMg(y) < CM(y)$ ou $CMg(y) = CM(y)$ podem ocorrer. Colocando de outra maneira, $CMg(y) = CM(y)$ pode implicar em qualquer um dos 3 resultados do teorema fundamental definidos: CM é crescente em y, CM é (localmente) constante em y, e CM é decrescente em y.

Resumindo, temos o teorema fundamental $CIM - CM$:

$$
\begin{aligned}
CMg(y) > CM(y) &\Rightarrow CM \text{ é crescente em } y \\
CMg(y) < CM(y) &\Rightarrow CM \text{ é decrescente em } y \\
CMg(y) = CM(y) &\Rightarrow CM \text{ pode ser crescente, constante ou decrescente em } y \\
CM \text{ é crescente em } y &\Rightarrow CMg(y) \geq CM(y) \\
CM \text{ é decrescente em } y &\Rightarrow CMg(y) \leq CM(y) \\
CM \text{ é (localmente) constante em } y &\Rightarrow CMg(y) = CM(y)
\end{aligned}
$$

3.7.4 Geometria dos Custos

Vamos primeiro examinar as curvas de custo de curto prazo. Neste caso, escreveremos a função de custo simplesmente como $c(y)$, que tem dois componentes: custos fixos e custos variáveis. Podemos, portanto, escrever o custo médio de curto prazo como

$$\text{CMCP} = \frac{c(w, y, x_f)}{y} = \frac{w_f x_f}{y} + \frac{w_v x_v(w, y, x_f)}{y} = \text{CFMCP} + \text{CVMCP}. \tag{3.120}$$

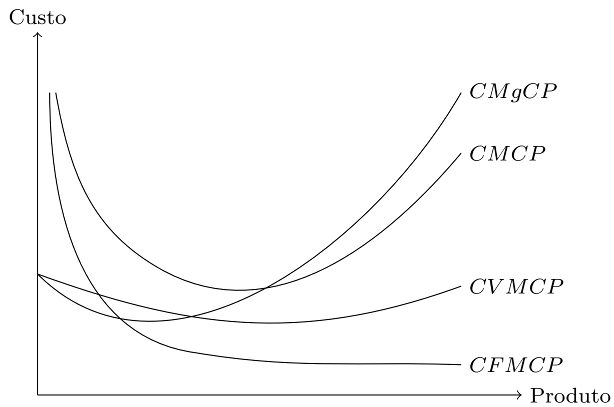
À medida que aumentamos a produção, os custos variáveis médios podem inicialmente diminuir se houver alguma região inicial de economias de escala. No entanto, parece razoável supor que os fatores variáveis requeridos acabarão por aumentar dados os retornos marginais decrescentes, conforme ilustrado na Figura 3.6. Os custos fixos médios devem, obviamente, diminuir com a produção, conforme indicado na Figura 3.6. Somando a curva de custo variável médio e a curva de custo fixo médio, obtemos a curva de custo médio em forma de U na Figura 3.6. A redução inicial nos custos médios deve-se à diminuição dos custos fixos médios. O eventual aumento nos custos médios deve-se ao aumento dos custos variáveis médios. O nível de produção em que o custo médio de produção é minimizado é conhecido como escala mínima eficiente.

Vamos agora considerar a curva de custo marginal. Qual é a sua relação com a curva de custo médio? Observe que

$$\frac{d}{dy}\left(\frac{c(y)}{y}\right) = \frac{yc'(y) - c(y)}{y^2} = \frac{1}{y}\left[c'(y) - \frac{c(y)}{y}\right], \tag{3.121}$$

em que $\frac{d}{dy}\left(\frac{c(y)}{y}\right) \leqq 0$ $(\geqq 0)$ se $c'(y) - \frac{c(y)}{y} \leqq 0$ $(\geqq 0)$.

Figura 3.6 – Curvas de custo no curto prazo



Assim, a curva de custo variável médio é decrescente quando a curva de custo marginal está abaixo da curva de custo variável médio, e é crescente quando a curva de custo marginal está acima da curva de custo variável médio. Conclui-se que o custo médio atinge seu mínimo em y^* quando a curva de custo marginal cruza a curva de custo variável médio, ou seja,

$$c'(y^*) = \frac{c(y^*)}{y^*}. \tag{3.122}$$

Toda a análise que acabamos de discutir vale tanto no longo quanto no curto prazo. No entanto, se a produção exibir retornos constantes de escala no longo prazo, de modo que a função de custo seja linear no nível de produção, o custo médio, o custo variável médio e o custo marginal são todos iguais.

3.7.5 Curvas de Custo de Curto e Longo Prazo

Vamos agora considerar a relação entre as curvas de custo de longo prazo e as curvas de custo de curto prazo. É claro que a curva de custo de longo prazo nunca deve ficar acima de qualquer curva de custo de curto prazo, uma vez que o problema de minimização de custos de curto prazo é apenas uma versão restrita do problema de minimização de custos de longo prazo.

Vamos escrever a função de custo de longo prazo como $c(y) = c(y, z(y))$. Aqui omitimos os preços dos fatores, uma vez que eles são assumidos como fixos e deixamos que $z(y)$ seja a demanda por um único fator fixo que minimiza custos. Seja y um determinado nível de produto, e seja

$z^* = z(y^*)$ a demanda associada de longo prazo para o fator fixo. O custo de curto prazo, $c(y, z^*)$, deve ser pelo menos tão grande quanto o custo de longo prazo, $c(y, z(y))$, para todos os níveis de produção, e o custo de curto prazo deve ser igual ao custo de longo prazo para produzir y^*, de tal forma que $c(y^*, z^*) = c(y^*, z(y^*))$. Assim, as curvas de curto e longo prazo devem ser tangentes em y^*.

Esta é apenas uma reformulação geométrica do teorema do envelope. A inclinação da curva de custo de longo prazo em y^* é dada por

$$\frac{dc(y^*, z(y^*))}{dy} = \frac{\partial c(y^*, z^*)}{\partial y} + \frac{\partial c(y^*, z^*)}{\partial z}\frac{\partial z(y^*)}{\partial y}. \tag{3.123}$$

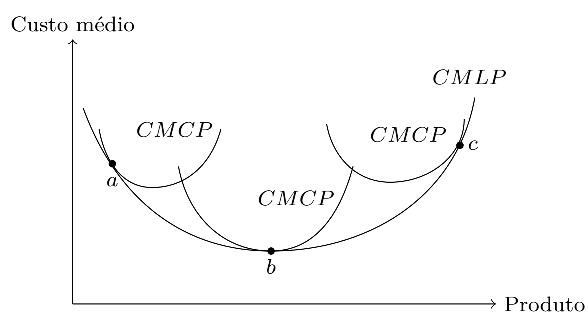
Mas como z^* é a melhor escolha dos fatores fixos no nível de produto y^*, devemos ter

$$\frac{\partial c(y^*, z^*)}{\partial z} = 0. \tag{3.124}$$

Assim, os custos marginais de longo prazo em y^* são iguais a (y^*, z^*).

Finalmente, notamos que, se as curvas de custo de curto e longo prazo são tangentes, as curvas de custo médio de curto e longo prazo também devem ser tangentes. Uma configuração típica é ilustrada na Figura 3.7.

Figura 3.7 – Curvas de custo médio de curto e longo prazo



Observe que as curvas de custo médio de curto e longo prazo devem ser tangentes, o que implica que os custos marginais de curto e longo prazo devem ser iguais.

3.7.6 Escala e Escopo

Sejam as seguintes situações:

- Economia constante de escala

$$C(aq) = aC(q), \quad a > 1 \tag{3.125}$$

Neste caso, o custo médio é constante.

- Economia de escala

$$C(aq) < aC(q), \quad a > 1 \tag{3.126}$$

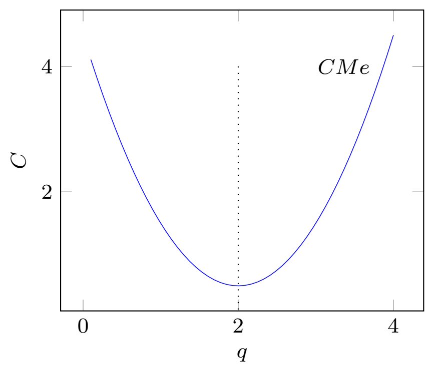
Neste caso, o custo médio é decrescente.

- Desconomia de escala

$$C(aq) > aC(q), \quad a > 1 \tag{3.127}$$

Neste caso, o custo médio é crescente.

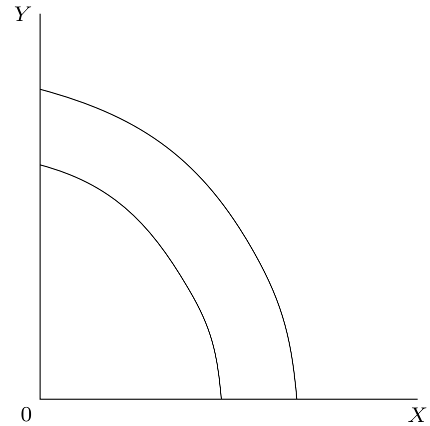
Figura 3.8 – Economia de escala



Economias de escopo: ao produzir mais de um tipo de produto que está intimamente ligado, o custo é menor do que quando produzido separadamente.

Curva de Transformação do Produto: mostra várias combinações de produtos que podem ser produzidos com um determinado conjunto de insumos.

Figura 3.9 – Curva de transformação do produto



No caso de economias de escopo, a curva de transformação do produto é negativamente inclinada e côncava. O grau de economias de escopo é definido da seguinte forma:

$$EE = \frac{C(q_1) + C(q_2) - C(q_1, q_2)}{C(q_1, q_2)}. \tag{3.128}$$

- Se $EE > 0$, temos economia de escopo.
- Se $EE < 0$, temos deseconomia de escopo.

Referências Bibliográficas

[Debreu1959] Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Yale University Press.

[Fuss and McFadden2014] Fuss, M. and McFadden, D. (2014). *Production Economics: A Dual Approach to Theory and Applications: Applications of the Theory of Production*, volume 2. Elsevier.

[Hicks1946] Hicks, J. (1946). *Value and Capital*. Clarendon Press.

[Jehle2001] Jehle, G. A. (2001). *Advanced Microeconomic Theory*. Pearson Education India.

[Mas-Colell et al.1995] Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic Theory*, volume 1. Oxford University Press, New York.

[Samuelson1947] Samuelson, P. (1947). *Foundations of Economic Analysis*. Harvard University Pressk.

[Shepherd2015] Shepherd, R. W. (2015). *Theory of Cost and Production Functions*. Princeton University Press.

Capítulo 4

Teoria do Mercado

Contents

[...] 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

Nos capítulos anteriores, estudamos o comportamento de consumidores individuais e empresas. Aqui, exploramos as consequências desse comportamento quando consumidores e empresas se reúnem em mercados. Consideraremos a determinação de preço e quantidade de equilíbrio em um mercado único ou grupo de mercados relacionados pelas ações dos agentes individuais para diferentes estruturas de mercados. Essa análise de equilíbrio é chamada de análise de equilíbrio parcial porque se concentra em um mercado único, assumindo implicitamente que as mudanças nos mercados considerados não alteram nem os preços de outros bens e não perturbam o equilíbrio que detém nesses mercados. Vamos tratar todos os mercados simultaneamente na teoria do equilíbrio geral.

Vamos nos concentrar em modelar o comportamento do mercado da empresa. Como as empresas determinam o preço pelo qual venderão sua produção ou os preços pelos quais estão dispostas a comprar insumos? Veremos que, em certas situações, o "comportamento de tomada de preços" pode ser uma aproximação razoável do comportamento ideal, mas em outras situações, teremos que explorar modelos do processo de fixação de preços. Vamos primeiro considerar o caso ideal (benchmark) de concorrência perfeita. Então nos voltamos para o estudo de configurações em que alguns agentes têm poder de mercado. Essas configurações incluem estruturas de mercado de monopólio puro, concorrência monopolística, oligopólio e monopsônio.

4.2 Preços

4.2.1 O Papel dos Preços

O principal insight da Riqueza das Nações de Adam Smith é simples: se uma troca entre duas partes é voluntária, não ocorrerá a menos que ambos acreditem que se beneficiarão dela. Como isso é verdade para qualquer número de partes e para o caso de produção? O sistema de preços é o mecanismo que executa esta tarefa muito bem sem direção central.

Os preços desempenham três funções na organização de atividades econômicas em uma economia de mercado livre:

1. Eles transmitem informações sobre produção e consumo. O sistema de preços transmite apenas as informações importantes e apenas para as pessoas que precisam saber. Assim, transmite informação de forma eficiente.

2. Eles fornecem incentivos corretos. Uma das belezas de um sistema de preço livre é que os preços que trazem as informações também fornecem um incentivo para reagir à informação

não apenas sobre a demanda por produto, mas também sobre a maneira mais eficiente de produzir um produto. Eles fornecem incentivos para adotar os métodos de produção que são menos onerosos e, portanto, usam os recursos disponíveis para as finalidades mais valorizadas.

3. Eles determinam a distribuição de renda. Eles determinam quem recebe quanto do produto. Em geral, não se pode usar os preços para transmitir informações e fornecer um incentivo para agir sem utilizar preços para afetar a distribuição de renda. Se o que uma pessoa recebe não depende do preço que ele recebe pelos serviços desses recursos, que incentivo ele tem para buscar informações sobre preços ou para agir com base nessas informações?

4.2.2 Formação dos Preços

Um princípio básico da teoria dos preços é que todo preço final de um bem X para o consumidor é uma soma de *shadow-prices* que surgem toda vez que aparece alguma restrição ao longo do processo de produção da primeira unidade de X a ser consumida pelo consumidor supra-marginal, ou seja, aquele que mais valoriza a próxima unidade dentre os que não estão no mercado. As restrições podem ocorrer mesmo que a produção da unidade adicional do bem X requeira toda uma cadeia de produtos intermediários $A, B, C, \ldots$, sobre os quais o produtor da próxima unidade do bem X não tenha qualquer controle. As restrições que geram *shadow-prices* podem ser de quaisquer tipos: tecnológicas, informacionais, mercadológicas ou mesmo institucionais.

Antes, algumas definições. Vamos imaginar que cada consumidor consome uma e apenas uma unidade discreta. Consumidor marginal é o consumidor da última unidade produzida, a quantidade que equilibra o mercado. Cada consumidor possui uma disposição a pagar (também chamada de preço de demanda) pela unidade do bem e ela representa o valor que o consumidor está disposto a sacrificar do seu consumo de outros bens para adquirir, em troca, a unidade do bem em questão. Os consumidores são ordenados de forma decrescente de acordo com suas disposições marginais a pagar. É nisso que se resume a curva de demanda agregada. O custo marginal de produção de uma unidade representa o valor que a sociedade atribui aos recursos deslocados da economia para a produção dessa unidade.

Vamos ver um exemplo. Suponha que o custo marginal de produção do bem X seja constante igual a \$100, mas que o produtor possui uma capacidade instalada máxima de 10 unidades. Existem 11 demandantes e cada um demanda uma única unidade. Os 10 primeiros consumidores têm disposições marginais a pagar de \$200; o 11º, \$150. Como a capacidade máxima é de 10 unidades, então o 11º demandante fica de fora do mercado e somente os 10 primeiros (com maiores disposições a pagar) são atendidos. O preço de equilíbrio é qualquer preço P entre \$150 e \$200. Mesmo que o 11º consumidor não seja atendido, é a sua disposição a pagar que determina o preço mínimo a partir do qual o 10º consumidor e o produtor podem barganhar. Por simplicidade, suponha que o preço de equilíbrio seja \$150.

Evidentemente o custo marginal da 10º unidade é \$100. A diferença de \$50 entre o preço final e o custo marginal nada tem de anti-competitivo. Esse valor de \$50 corresponde ao preço-sombra da capacidade instalada máxima, que se configura numa restrição à produção da unidade adicional que atenderia à demanda do 11º consumidor, que é o primeiro da fila para entrar no mercado se a próxima unidade for ofertada ao preço de \$150 ou menos.

De fato, o 11º consumidor está disposto a pagar \$150 pela 11º unidade do ofertante, mas o ofertante não tem como atendê-lo, já que está operando em sua capacidade máxima de 10 unidades. Se ele quiser atendê-lo, deverá expandir sua capacidade instalada máxima o suficiente para ser capaz de produzir exatamente uma unidade a mais. Para fazer isso, ele deverá gastar um valor V. Depois deverá produzir a unidade adicional cujo custo marginal é \$100 para vendê-la por \$150. Logo, só vale a pena expandir a capacidade instalada se o preço-sombra V da capacidade instalada incremental não for maior que \$50. Se fosse menor, o ofertante poderia acrescentar uma segunda ou mesmo terceira unidade de capacidade até o ponto em que ficasse indiferente entre expandir ou não, pois a cada passo teria lucro (incremental positivo, mas decrescente). Portanto, suponha, sem perda de generalidade, que isso já ocorre nessa primeira unidade adicional de capacidade instalada.

Note que quem determina o preço-sombra da capacidade instalada é o consumidor supra-marginal: mais especificamente, a diferença entre sua disposição a pagar, \$150, e o custo marginal, \$100, da unidade que ele adquiriria, se pudesse. Essa diferença de \$50 se configuraria numa renda do produtor, caso ele não enfrentasse a limitação da capacidade instalada. Como essa limitação existe, é esse excedente que deve ser gasto na expansão da capacidade para a produção da unidade adicional. Dessa forma, o preço-sombra da capacidade instalada não é intrínseco à planta da firma: ele é a valoração do consumidor supra-marginal em excesso ao custo marginal. Em outras palavras, é o preço de demanda que o 11º consumidor está disposto a pagar que determina o custo incremental que será incorrido pelo ofertante. Não é o custo que determina o preço, mas o preço é que determina o custo.

Já com base nesse raciocínio, vemos que o preço final de \$150 pode ser decomposto em \$100 + \$50, que é o custo marginal operacional da última unidade ofertada mais o preço-sombra da capacidade instalada. O custo marginal operacional da última unidade ofertada é o valor que a sociedade atribui aos recursos deslocados da economia para produção dessa unidade e é, por isso mesmo, o preço-sombra social dos insumos. Chamemos esse preço-sombra de P_1. Temos, assim, o primeiro resultado: o preço final é a soma de dois preços-sombra. Neste caso, o custo marginal operacional mais o preço-sombra da capacidade instalada. Este último preço-sombra, que chamaremos de P_2, é o quanto a sociedade está disposta a pagar pelo aumento da capacidade com vistas à oferta da 11ª unidade. Assim, o preço do produto é $\$150 = P_1 + P_2$.

Suponha que o produto X é adquirido por um monopolista que determina seu preço pela regra de mark-up. Cada unidade de X lhe custa \$150, de modo que seu custo marginal é constante e igual a \$150. O bem X é o único insumo variável. O mark-up é de \$40. Logo, o preço de monopólio é \$190. Esse mark-up também provém de uma limitação imposta ao monopolista e se configura, portanto, em um *shadow-price* de natureza muito especial. Já explico o que significa essa restrição. Antes, porém, observe que é claro que esse excesso de \$40 sobre o custo marginal evidentemente decorre da interseção com a curva de receita marginal. A RMg da última unidade é de tal monta que $RMg = CMg$.

Recorde que o monopolista, assim como qualquer outro produtor, atua segundo a regra $RMg = CMg$ (receita marginal = custo marginal). A diferença é que a receita marginal adquire uma forma especial. Se $R = PQ$ denota a receita e C o custo, então a regra $RMg = CMg$ pode ser expressa como:

$$(\Delta P)Q + P(\Delta Q) = \Delta C. \tag{4.1}$$

Seja P^* o preço de monopólio e seja P' a disposição a pagar do consumidor supra-marginal pela próxima unidade além da quantidade de monopólio. Seja Q^* a quantidade de monopólio (aquela à qual a receita marginal é igual ao custo marginal ou, se maior, aquela tal que, para a próxima unidade, se torna menor que o custo incremental correspondente). Produzir essa unidade para o consumidor supra-marginal, o primeiro na fila para entrar no mercado (ou seja, o consumidor com a maior disposição marginal a pagar dentre os consumidores não-atendidos pelo monopolista), implica uma variação $\Delta Q = 1$ da quantidade produzida. Então a expressão acima se torna:

$$P = \Delta C + (-\Delta P)Q^*, \tag{4.2}$$

em que $(\Delta P)Q^* = (P' - P^*)Q^*$, ou seja, $(\Delta P)Q^* = P'Q^* - P^*Q^*$. Em palavras bem claras, $(\Delta P)Q^* = P'Q^* - P^*Q^*$ é a perda de receita que o monopolista teria se vendesse a quantidade de monopólio Q^* ao preço dado pela disposição marginal a pagar, P', do consumidor supra-marginal, isto é, se o preço cobrado fosse a disposição a pagar do consumidor supra-marginal, P', e não P^*. Como $P' < P^*$, então essa perda de receita vem com sinal negativo. Portanto, o termo $(-\Delta P)Q^*$ é positivo e representa um potencial custo adicional para o monopolista, caso ele aplique mark-up conforme o preço do consumidor supra-marginal e não o do marginal. Note ainda que a expressão $(\Delta P)Q^* = P'Q^* - P^*Q^*$ pode ser escrita como $(\Delta P)Q^* = P'(Q^* + 1) - P^*(Q^* + 1) + (P^* - P')$. O primeiro termo, $P'(Q^* + 1) - P^*(Q^* + 1)$, é a perda de receita que o monopolista enfrenta se incluir o consumidor supra-marginal na regra de mark-up. Neste caso, a quantidade de monopólio é $Q^* + 1$, já que a última unidade produzida seria aquela destinada a satisfazer o consumidor supra-marginal. Esse termo evidentemente tem sinal negativo, pois é uma perda para o monopolista. De fato, como já mencionei, $P' < P^*$. Já o segundo termo, $P^* - P'$, é positivo e representa a economia de custo de que o consumidor supra-marginal se aproveita ao comprar a unidade supra-marginal ao preço P' e não ao preço correto P^*. Portanto, $(\Delta P)Q^*$ é o efeito líquido, em termos sociais, da inclusão do consumidor supra-marginal na regra de mark-up. É, dessa forma, o preço-sombra do consumidor supra-marginal para os ganhos sociais de troca. Esse preço-sombra nasce em razão do uso da precificação linear, que, no caso do monopólio, se materializa na regra de mark-up. Logo, a regra de mark-up, quando vista na margem, ou seja, $P^* = \Delta C + (-\Delta P)Q^*$, nos diz que o preço de monopólio é igual ao custo marginal (incremental) ΔC da última unidade produzida (a de equilíbrio) mais o preço-sombra do consumidor supra-marginal decorrente da restrição imposta pela precificação linear. Foi esse valor $(-\Delta P)Q^*$ que se materializou no mark-up de \$40 em nosso exemplo.

A restrição que o monopolista enfrenta é que ele se vê forçado a adotar uma regra de preço linear, ou seja, ele deve cobrar o mesmo preço de \$190 por cada unidade produzida e de cada consumidor. O monopolista poderia extrair todo o excedente do mercado se fizesse discriminação perfeita de preço ou se adotasse uma tarifa bipartite.

Há duas explicações para não adotar estas ou outras regras de preço alternativas à precificação linear. A primeira e mais óbvia é uma restrição institucional. O governo simplesmente proíbe que o monopolista faça quaisquer discriminações de preço.

Outra explicação é de natureza informacional. Mesmo que não houvesse restrições institucionais à discriminação de preços, poderia ser simplesmente muito custoso extrair de cada demandante sua disposição a pagar. Essa informação não é difícil de ser obtida, mas é custosa. O monopolista poderia leiloar cada unidade produzida por um leilão ascendente de maior preço (ou mesmo de segundo maior preço, que é o leilão de Vickrey), ofertando o objeto a quem desse o maior lance. Esse mecanismo gera incentivos para os demandantes revelarem suas verdadeiras disposições a pagar quando estas são informação privada. O problema é que ofertar cada unidade por meio de um leilão pode simplesmente ser inviável.

Nesse sentido, dadas as restrições listadas acima, que, na falta de um termo que englobe tanto a potencial restrição institucional quanto a informacional, chamarei de mercadológicas, o próprio monopolista prefere adotar a regra de preço linear e, ao tentar extrair todo o potencial excedente de mercado, o máximo que ele consegue extrair é uma parte desse excedente e o faz pela regra de mark-up. O mark-up de \$40, como vimos, está intimamente relacionado com o consumidor supra-marginal. Este consumidor é a personificação da restrição mercadológica enfrentada pelo monopolista. A estrutura de mercado lhe impõe essa restrição. É como se o mercado dissesse ao monopolista: "Use uma regra de preço linear e use-a até o ponto em que a quantidade produzida bata naquele consumidor supra-marginal a partir do qual você fará lucros incrementais negativos". Observe que, de um ponto de vista essencial, esse consumidor supra-marginal funciona exatamente como aquela primeira unidade a ser produzida além da capacidade máxima instalada e cujo preço-sombra deve ser acrescido ao custo marginal. Dessa forma, o mark-up se constitui numa espécie de preço-sombra que deve ser acrescido ao custo marginal do monopolista em razão da restrição mercadológica de linear pricing. Quem determina esse preço-sombra é precisamente o consumidor (ou unidade) supra-marginal. Chame esse preço-sombra, \$40, de P_3. Esse valor é um preço-sombra porque reflete a valoração que o mercado atribui à restrição mercadológica de linear pricing. Essa restrição é exógena e, se o consumidor supra-marginal mudar seu preço de demanda ou sair da fila de entrada no mercado, o mark-up muda e o excedente total também.

O monopolista não tem que conhecer a curva de demanda agregada por seu produto. No modelo matemático esses elementos são dados e nos passam a ideia incorreta de que o monopolista, ao procurar maximizar o seu lucro, maximiza a função lucro e que, para isso, deve conhecer a curva de demanda agregada e a sua função custo, de modo a deduzir a regra de mark-up. Não é isso que a Microeconomia diz. O modelo matemático simplesmente chega aos mesmos resultados, só que pelo caminho formal. A partir de princípios muito mais elementares, o raciocínio conduz aos resultados que conhecemos, sem necessidade de se pressupor conhecimento perfeito de preços e funções de qualquer tipo. Por exemplo, o monopolista, instado a praticar linear pricing, vai testando mark-ups diferentes sobre os custos incrementais incorridos, mudando, assim, os preços cobrados. Aumentos pequenos de preços alteram sua receita marginal. Ele compara a receita proveniente do preço mais alto e da quantidade comercializada mais baixa no período (digamos, no mês) com a receita proveniente do preço anterior, mais baixo, e quantidade comercializada maior. Deduzidos os respectivos custos incrementais, ele apenas compara qual deu mais lucro. Cada mudança é um passo e a cada passo ele compara os lucros incrementais alternativos, escolhendo, obviamente, aquela ação que dê o lucro incremental maior. Ao continuar nesse processo, eventualmente as oportunidades de ganho serão exauridas e, se nada mudar nesse caminho, encontrará o preço de monopólio. Esse procedimento passo-a-passo é precisamente o que devemos entender por marginalismo. O fato é que esse procedimento é otimizatório. Pra ser mais objetivo e enfático nesse ponto, toda vez que tomamos uma decisão sobre alternativas distintas e escolhemos a que tem maior valor (ou menor custo de oportunidade), estamos maximizando. Quando se diz que o indivíduo é racional e maximizador, é nesse sentido que deve ser entendido, não no sentido matemático do termo. Ainda assim, os dois sentidos são equivalentes.

Não percebemos isso porque as variações são infinitesimais e as funções envolvidas são contínuas, de modo que o consumidor marginal e o supra-marginal se confundem no mesmo preço de demanda. Dado que $P_3 = \$40$ e $P_1 + P_2 = \$150$, já temos o seguinte resultado: o preço final do monopolista, \$190, é igual à soma dos *shadow-prices* $P_1 + P_2 + P_3$, em que P_1 é o custo marginal do produtor do bem X (ou seja, o preço-sombra dos recursos deslocados da economia para a produção da unidade do bem X, isto é, o preço-sombra da restrição da escassez dos recursos para a sociedade), P_2 é o preço-sombra da capacidade instalada máxima do produtor do bem X (ou seja, o preço-sombra da restrição tecnológica da capacidade instalada) e P_3 é o mark-up do monopolista (isto é, o preço-sombra instituído pela restrição mercadológica de linear pricing):

$$P_1 + P_2 + P_3 = \$190. \tag{4.3}$$

Podemos alongar ainda mais a cadeia de produção até chegar ao produto final a ser consumido, aquele que Menger chamava de bem de última ordem, um bem que finalmente será consumido e não utilizado como insumo para a produção de algum outro bem. É nesse sentido que o consumidor é quem movimenta a produção. Os preços de demanda desses bens de última ordem se decompõem na soma de *shadow-prices* ao longo de toda a cadeia de produção. Toda vez que o produtor enfrenta uma restrição que limita a satisfação de consumidores adicionais, deve-se acrescer o correspondente preço-sombra aos custos marginais. A ideia de preço-sombra aqui é muito mais pervasiva: representa o valor que a sociedade está disposta a pagar pela superação da restrição.

Uma observação final pode ajudar a esclarecer esses princípios. Em Economia, a alocação ótima dos recursos tem a sua contraparte formal na otimização. Essa otimização é o que chamamos de problema primal. Todo problema primal tem um problema gêmeo e complementar: o problema dual. A solução do problema dual nos dá os multiplicadores de Lagrange associados a cada restrição do problema primal. Esses multiplicadores de Lagrange são exatamente os preços-sombras que procuramos. Como no problema primal as restrições são dadas por inequações distintas, uma a uma, então no problema primal os respectivos preços-sombras entram aditivamente na função-objetivo do problema dual. Cada preço-sombra vem multiplicado pela respectiva variável exógena da restrição primal, a qual exerce o papel de recurso limitado cujo uso deve ser racionalizado. Lá na ponta da cadeia, a satisfação do consumidor supra-marginal requer a variação de um montante de recurso necessário que vai se estendendo cadeia abaixo. Cada montante necessário variado corresponde à unidade necessária para realizar a produção da unidade supra-marginal. É por isso que os preços-sombra são somados. O que o dono de um clube precisa para atender um certo pacote adicional de crianças a mais na piscina infantil? Se a densidade de crianças na piscina deve obedecer a um padrão, então a piscina deve ser aumentada de modo a preservar a densidade. Ao expandir a piscina pelo volume adicional necessário, o dono do clube contrata uma empresa e paga um tanto. Esse tanto deve bater com a receita incremental que o dono do clube terá com o pacote adicional de usuários. O montante expandido é, assim, a "unidade" de expansão e o custo incorrido na expansão reflete o preço-sombra da capacidade instalada da piscina. Observe uma vez mais que esse preço-sombra pode, equivalentemente, ser medido pela disposição que os potenciais usuários têm a pagar pela piscina. Embora haja essa equivalência, quem a determina são os consumidores: o dono da piscina só vai gastar se vislumbrar potenciais ganhos com isso.

4.3 Demanda e Oferta

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, ou seja, desde aquele que está disposto a pagar mais até aquele disposto a pagar menos, como na tabela abaixo.

Tabela 4.1 – 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

Assim, o número 3 relativo a Maria significa que Maria possui a terceira maior disposição a pagar pelo frango ($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.

Tabela 4.2 – 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

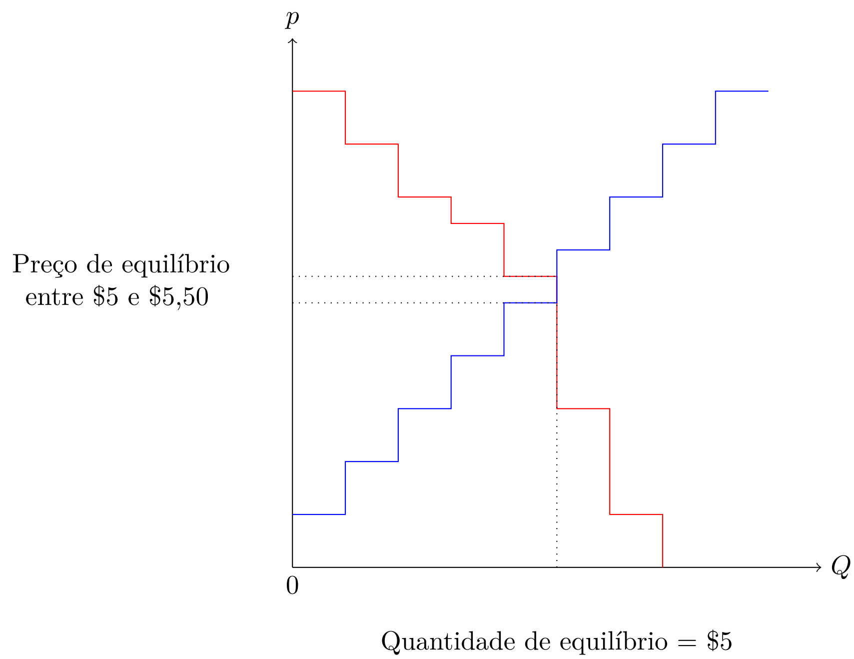
Juntando tudo numa tabela só, temos o seguinte quadro de preços de demanda e de oferta para as diversas quantidades de frango transacionadas:

Tabela 4.3 – 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 mais bem 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 4.1.

Figura 4.1 – 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 $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 $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 $5,50 pelo frango e a ofertante Paula aceitaria $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 $6, mas a sexta demandante, Dora, só estaria disposta a pagar $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 $5 a unidade. Ora, outro equilíbrio seria o comércio de 5 frangos ao preço de $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 $5 para ofertar a quinta unidade e Antônio está disposto a pagar no máximo $5,50 por ela. Portanto, qualquer preço entre $5 e $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.

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.

Exemplo 4.3.1. *Suponha que* $\widehat{x}(p) = a - bp$ *e* $c(y) = y^2 + 1$*. Dado que* $CMg = 2y$*, temos:*

$$y = \frac{p}{2}, \tag{4.4}$$

e, portanto,

$$\widehat{y}(p) = \frac{Jp}{2}. \tag{4.5}$$

Fazendo $a - bp = \dfrac{Jp}{2}$*, chegamos a:*

$$p^* = \frac{a}{b + \frac{J}{2}}. \tag{4.6}$$

Agora, para o caso geral de $D(p)$ *e* $S(p)$*, o que acontece com o preço de equilíbrio se o número de firmas aumenta? Vemos que:*

$$D(p(J)) = Jy(p(J)), \tag{4.7}$$

o que implica em

$$\begin{aligned} D'(p(J))p'(J) &= y(p) + Jy'(p(J))p'(J) \\ p'(J) &= \frac{y(p)}{D'(p) - Jy'(p)} < 0, \end{aligned} \tag{4.8}$$

o que significa que o preço de equilíbrio decresce quando o número de firmas aumenta.

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.

4.4 Teorema de Allen-Alchian

O preço final é uma soma de preços-sombras de etapas entre a fonte e o destino. A cada etapa, o preço vai sendo acrescido do respectivo preço-sombra (ou multiplicador de Lagrange). O exemplo mais simples é o custo de transporte. Isso é típico de lojas que compram no atacado para vender no varejo. O preço do bem no varejo é mais alto do que no atacado porque ele embute o preço-sombra do transporte ou da intermediação entre fonte e destino. Outro exemplo é o preço da energia elétrica em horários de pico de consumo. Enquanto a demanda é baixa, a tarifa da energia reflete o custo marginal operacional da produção de energia. Quando a demanda sobe no início da noite, a empresa produtora atinge a capacidade instalada e, por isso, acrescenta ao custo marginal operacional da energia o preço-sombra da capacidade instalada. É por isso que existe horário de verão. Não é para economizar energia: é para distribuir o consumo no início da noite de modo a evitar que se atinja a capacidade instalada e, com isso, evitar cobrar a tarifa mais alta de verão. Quando houver investimento em infra-estrutura e aumento da capacidade instalada, não precisaremos mais de horário de verão.

A chave para entender é o Teorema de Allen-Alchian. O Teorema de Allen-Alchian é também conhecido como a Terceira Lei da Demanda.

Considere um bem que pode ser de boa qualidade ou de qualidade ruim. O Teorema de Allen-Alchian explica como a distribuição do bem de qualidades substitutas é afetada pelos custos fixos. Como mencionei acima, para o comerciante que compra no atacado para vender no varejo a uma distância fixa, o preço-sombra do transporte é um custo fixo. Não entenda esse custo fixo como o custo fixo que você aprende no livro-texto, ou seja, como o custo fixo em que se incorre independentemente da quantidade produzida. Não é isso! O adjetivo fixo significa apenas que o preço final de cada lote do bem inclui o custo incremental de transporte. Subentende-se que a distância entre a fonte e o destino é fixa. Embora na aparência se confunda com um custo fixo, é, na essência, distinto. É, portanto, como uma parcela discreta do custo incremental. Para o produtor, o custo marginal é o custo marginal de produzir a unidade. Para produzir, ele não paga transporte. O transporte configura-se noutra atividade econômica, ou seja, outra etapa. O custo de oportunidade da realização dessa etapa é o valor do transporte. Como esta etapa se resume à ação binária de transportar ou não, o preço-sombra é o custo incremental do transporte, com aparência de fixo. Pode ser o mesmo independentemente da quantidade comprada no atacado, mas é variável conforme o número de viagens da origem para o destino.

Suponha que todos os consumidores de vinho têm as mesmas preferências e renda. Uma garrafa de vinho bom custa \$100. A de um vinho ruim, \$40. Se você consome o vinho no local em que foi produzido, você vai pagar $100 pelo vinho bom e $40 pelo vinho ruim. A quantidade que você consome de cada vinho deverá ser tal que sua taxa marginal de substituição entre vinho bom e ruim se iguale aos preços relativos, 100/40=2,5. Pra você, uma garrafa de vinho bom equivale a 2,5 de ruim.

Imagine agora que as duas garrafas são transportadas para um local distante. O custo de transporte é $20. O mesmo custo aplica-se aos dois tipos de vinho. Entendeu o sentido do "fixo"? Nesse local distante, os preços dos vinhos serão $120 e $60. Agora, a quantidade que você consome de cada vinho será aquela à qual a taxa marginal de substituição é igual a 120/60=2. Cada garrafa de vinho bom agora equivale a duas de ruim.

O vinho bom ficou relativamente mais barato para o consumidor distante. Como os consumidores são idênticos em termos de preferências e renda, a distribuição de consumo de vinhos bons e ruins no local distante é diferente da distribuição no local de produção. Os consumidores distantes vão consumir mais vinhos de boa qualidade. Os consumidores locais tenderão a consumir os vinhos piores. Quanto mais distante o local de consumo, maior a proporção do bem de boa qualidade relativamente ao bem ruim, pois o bem de boa qualidade torna-se relativamente mais barato do que nos locais mais próximos à produção.

4.5 Concorrência Perfeita

4.5.1 Uma Introdução

O que é concorrência perfeita? São poucas as respostas, mas significativas:

1. todo equilíbrio de competição perfeita é walrasiano, mas nem todo equilíbrio walrasiano é competitivo;

2. não são os preços que precedem à maximização (no sentido de os agentes tomarem os preços como dados nas maximizações de seus lucros e utilidades), mas sim o comportamento maximizador que precede os preços (no sentido de que os preços de equilíbrio surgem mediante os processos de arbitragem ou de exploração de oportunidades de ganhos em razão das diferenças entre as taxas marginais de troca dos agentes, quando percebidas);

3. comportamento otimizador significa apenas a comparação racional entre custos e benefícios de cada decisão, na margem, passo a passo;

4. competição perfeita é um conceito de natureza não só otimizatória, mas principalmente distributiva.

O que define competição perfeita na moderna Teoria de Equilíbrio Geral é a condição de *full appropriation*: todo agente internaliza na forma de excedente privado a sua contribuição marginal para o excedente social da economia. Nas palavras de Makowski & Ostroy (2001)[1]:

[1] Makowski, L. & Ostroy, J. (2001): "Perfect competition and the creativity of the market", Journal of Economic Literature, 39, pp. 479-535.

> We will define a perfectly competitive equilibrium as one in which every individual fully appropriates his social contribution, when viewed as the marginal individual added to the economy. The experiment is analogous to standard marginal productivity theory where, to calculate any worker's marginal product, we view him as the marginal worker added to the firm. A fundamental lesson comes from the question "which worker is the marginal one?" The teacher smiles and says, "They all are!".

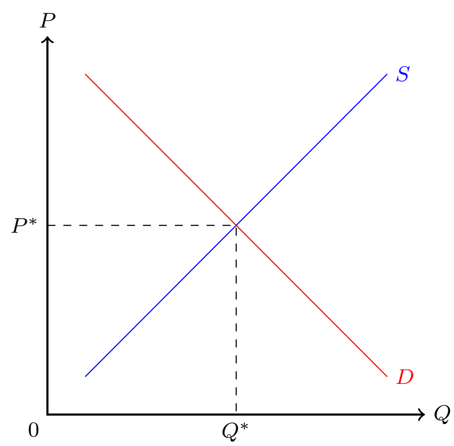
Esta condição equivale ainda a duas outras: (a) os agentes enfrentam ofertas e demandas perfeitamente elásticas (condição de Joan Robinson) e (b) existência de opções externas perfeitamente substituíveis.

A confusão acerca do que é competição perfeita reside na equivalência, incorreta, entre competição perfeita e price-taking. Nas palavras de Makowski e Ostroy (2001):

> To us, however, the problem of applicability stems not so much from its implicit informational restrictions as from its identification of perfect competition with price-taking. This identification hides the remarkable properties of perfect competition itself. The link between perfect competition and price-taking had its origins in the early developments of neoclassical theory. Portraying the individual as a price-taker was extremely useful for displaying the new equi-marginal principle underlying individual choice. But it had the unfortunate consequence of suppressing the entrepreneurial side of competition. In the end, price-taking and perfect competition became synonymous, leading to the conclusion that perfect competition is an inhospitable environment for the exercise of market creativity.

Para entendermos, afinal, do que se trata competição perfeita, seja a Figura 4.2. Essas curvas denotam a oferta agregada (em azul) e a demanda agregada (em vermelho). O equilíbrio de mercado ocorre à quantidade Q^* e ao preço P^*. A confusão começa na forma como chegamos a essas curvas. O consumidor toma os preços como dados, conhece-os todos, e maximiza sua utilidade, encontrando, assim, sua demanda Marshalliana pelo bem cujo mercado simbolizamos no gráfico acima. Os produtores tomam os preços fatoriais e de produto como dados e maximizam seus lucros (ou minimizam seus custos), encontrando, assim, suas curvas de oferta. Dessas curvas individuais obtemos as curvas agregadas. É importante ressaltar que os modelos matemáticos dessas decisões individuais são descrições alternativas do modus cogitandi do Homo Oeconomicus que chegam às mesmas conclusões por outras vias. Não são a forma exclusiva como o microeconomista raciocina. Enquanto o Homo Oeconomicus chega à sua decisão mediante um comportamento otimizatório simples, que consiste na mera comparação entre custos e benefícios incrementais, passo a passo, sob a condição de valoração marginal decrescente e taxas marginais de troca decrescentes (que são princípios naturais da psicologia humana e não teses que devam ser testadas empiricamente), sem necessariamente conhecimento perfeito de todos os preços e definitivamente sem qualquer função matemática na cabeça, o modelo matemático da decisão microeconômica, por outro lado, chega às condições de primeira ordem via otimização matemática, razão pela qual se faz necessário incluir na formulação inicial todas as variáveis exógenas (inclusive os preços) e as funções maximizadas. Na formulação matemática, os preços são dados antes e só então é que se procede à maximização. Na formulação microeconômica, o comportamento otimizador é dado antes e só quando se esgotam as possibilidades de arbitragem é que se chega a um vetor de preços invariante por desvios unilaterais, razão pela qual se diz que os agentes não mais conseguem alterar os preços.

Figura 4.2 – Equilíbrio



Em qualquer dos dois caminhos, os resultados lógicos são os mesmos. O gráfico acima esconde um mundo. Em primeiro lugar, as curvas são agregadas e por isso perdemos a perspectiva do que se passa no nível descentralizado. É lá que encontraremos a competição perfeita. Em segundo lugar, as quantidades são infinitamente divisíveis, tornando as funções, que normalmente seriam discretas, contínuas. O equilíbrio de mercado se resume a um ponto de interseção: a interseção entre as curvas agregadas. Em equilíbrio as oportunidades de arbitragem são esgotadas. Explorar oportunidades de arbitragem significa fazer lucro a partir das diferenças entre taxas marginais de substituição.

O fato é que a moderna modelagem do caminho para o equilíbrio não é mais tâtonment (como proposto por Walras), embora até hoje seja capítulo de livros-textos. O modo aceito hoje é o da exploração das oportunidades de arbitragem, que nada mais são do que ganhar com as diferenças entre as TMS's dos agentes. O seguinte excerto de Makowski e Ostroy (1998) resume a ideia: "using a model with a continuum of agents, we show that competitive equilibrium can be regarded as resulting from the elimination of arbitrage opportunities, rather than from the elimination of Walrasian excess demands"[2].

O que devemos fazer é tomar uma lupa e direcioná-la para esse pontinho, que, na verdade, possui toda uma geometria. O que acontecesse nesse ponto?

Seja o seguinte exemplo. Em uma economia há 4 ofertantes e 4 demandantes. Cada agente transaciona uma única e apenas uma única unidade. Os ofertantes são os agentes $i = 1, 2, 3, 4$ e os demandantes são os agentes $i = 5, 6, 7, 8$. Os custos incrementais $c(i)$ do ofertante $i = 1, 2, 3, 4$ e as valorações incrementais $v(i)$ do demandante $i = 5, 6, 7, 8$ são dados por:

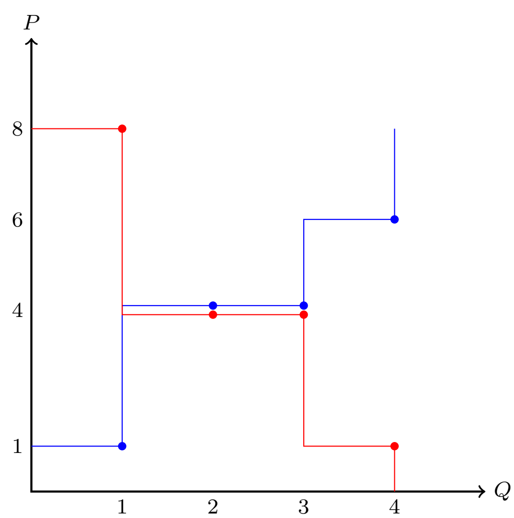
[2] Makowski, L. & J. Ostroy (1998), "Arbitrage and the flattening effect of large numbers", Journal of Economic Theory 78, 1-31.

Agente	$c(i)$	Agente	$v(i)$
1	1	5	8
2	4	6	4
3	4	7	4
4	6	8	1

Assim, o agente $i = 1$ é um ofertante que oferta uma unidade do bem a um custo incremental de $c(1) = \$1$. Esse é o valor que a sociedade atribui aos recursos deslocados da economia para a produção dessa unidade pelo ofertante $i = 1$. Analogamente, o agente $i = 5$ é um demandante cuja disposição incremental a pagar pela primeira e única unidade do bem é $v(5) = \$8$. Esse é o valor do consumo privado de outros bens que ele está disposto a sacrificar para adquirir, em troca, essa unidade do bem.

No gráfico abaixo temos a oferta (inversa, isto é, preço contra quantidade) agregada, que é a soma horizontal das ofertas (inversas) individuais. A curva de demanda agregada é analogamente obtida pela soma horizontal das demandas inversas e é como no gráfico a seguir. O equilíbrio se obtém pela interseção das curvas de oferta e demanda agregadas, como abaixo.

Figura 4.3 – Oferta e demanda



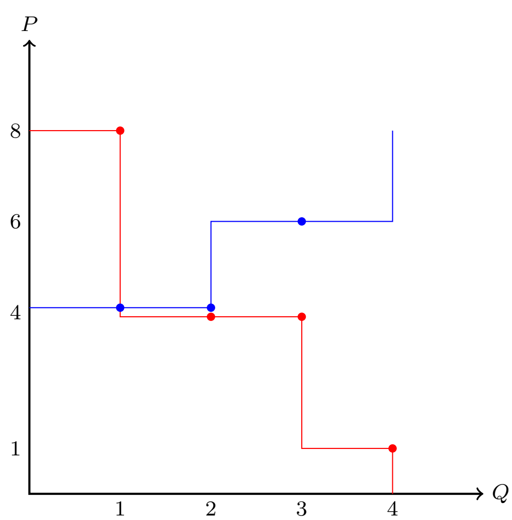
Note que a interseção ocorre no trecho horizontal que corresponde às unidades 2 e 3. O preço de equilíbrio é $P^* = 4$ e a quantidade de equilíbrio é $Q^* = 3$ unidades. Convém calcular o excedente social e os excedentes privados dos agentes. Isso é fácil de calcular. O ofertante $i = 1$ recebe \$4 pela unidade que lhe custa \$1 produzir, de modo que seu excedente privado é \$3. Os ofertantes $i = 2, 3$ têm excedente privado nulo. O ofertante $i = 4$ não participa do mercado, pois o preço de \$4 é inferior ao seu custo de \$6. O demandante $i = 5$ está disposto a pagar \$8, mas paga \$4, de modo que seu excedente privado é \$4. Os demandantes $i = 6, 7$ têm excedente nulo (estão indiferentes entre participar ou não) e o demandante $i = 8$ não participa. A lista dos excedentes privados $\pi(i)$, para $i = 1, \ldots, 8$, é, portanto:

Excedentes Privados	
$\pi(1) = 3$	$\pi(5) = 4$
$\pi(2) = 0$	$\pi(6) = 0$
$\pi(3) = 0$	$\pi(7) = 0$
$\pi(4) = 0$	$\pi(8) = 0$

O excedente social TS (total surplus) é a soma dos excedentes privados, de modo que $TS =$ \$7. Observe que o excedente social é a área entre as curvas de oferta e de demanda agregadas.

Para verificarmos se há competição perfeita, devemos calcular o preço-sombra de cada agente, ou seja, sua contribuição marginal para os ganhos sociais de troca (ou excedente social). Para sabermos o preço-sombra do ofertante $i = 1$, devemos retirá-lo da economia e recalcular o excedente social sem a sua presença. Quando o ofertante $i = 1$ está ausente da economia, a curva de oferta (inversa) agregada se desloca uma unidade para a esquerda, pois o ofertante $i = 1$ contribui com uma única unidade. A nova situação é descrita no gráfico a seguir.

Figura 4.4 – Oferta e demanda sem o agente $i = 1$



O preço de equilíbrio continua sendo $P^* =$ \$4, mas a quantidade de equilíbrio cai para $Q' = 2$ unidades. O excedente social sem a presença do agente $i = 1$ cai para $TS(-1) =$ \$4. De fato, esta é a área entre a curva de demanda agregada e a nova curva de oferta agregada (sem a participação do ofertante $i = 1$).

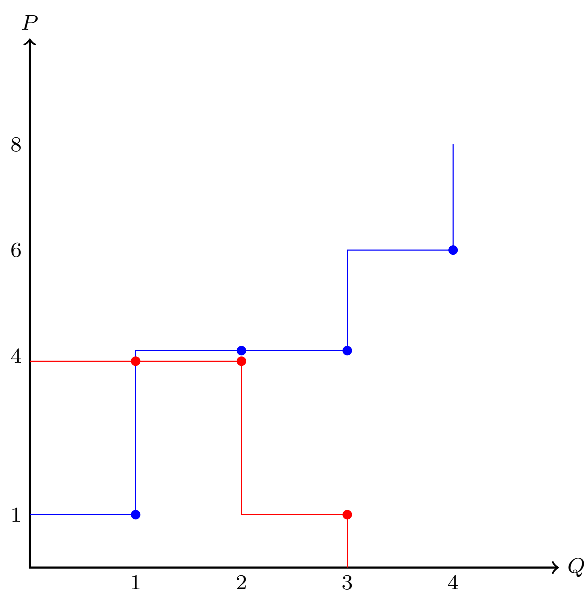
Outra maneira de ver isso é pela soma dos excedentes privados. Todos os agentes continuam com os mesmos excedentes de antes, menos o agente $i = 1$ que saiu da economia, levando consigo seu excedente de \$3.

Essa é uma observação importante, pois mostra que os agentes da economia são indiferentes à participação do agente $i = 1$. Portanto, a contribuição marginal do agente $i = 1$ para os ganhos sociais de troca (total gains from trade ou excedente social) é $TS - TS(-1) = \$7 - \$4 = \$3$. Essa diferença entre os valores ótimos de bem-estar com e sem a presença do agente $i = 1$ é o seu preço-sombra social, equivale ao seu multiplicador de Lagrange para o problema de otimização alocativa. Refraseando, a sociedade é indiferente à participação do ofertante $i = 1$, pois tudo que a sociedade ganha com a participação desse agente é inteiramente internalizada por ele na forma de excedente privado. Sem o agente $i = 1$, a sociedade tem um excedente social de \$4. Com a participação desse agente, a sociedade aumenta seu excedente social para \$7, mas esse incremento é apropriado pelo agente $i = 1$ na forma de excedente privado. Ele recebe de forma justa o quanto ele contribui. Denotamos o preço-sombra do agente i (ou sua contribuição marginal social) por $CMS(i) = TS - TS(-i)$, em que TS denota o excedente total e $TS(-i)$ denota o excedente total sem a presença do agente i. O que acabamos de ver foi que $CMS(1) = \pi(1)$.

Essa propriedade, $CMS(i) = \pi(i)$, quando válida para todos os agentes, é denominada *full appropriation* (apropriação plena). Sua natureza é essencialmente pigoviana, pois foi Pigou, em seu livro The Economics of Welfare, que examinou à exaustão os efeitos alocativos, em termos de eficiência, das divergências entre os benefícios e custos marginais sociais e privados.

Vamos calcular o preço-sombra do agente $i = 5$, que é o demandante que mais valoriza o bem. Voltando à situação inicial, com a presença de todos os 8 agentes, retiramos o agente $i = 5$. O efeito de sua ausência é um deslocamento da curva de demanda agregada uma unidade para a esquerda, conforme o gráfico a seguir:

Figura 4.5 – Oferta e demanda sem o agente $i = 5$



O preço de equilíbrio continua sendo $P^* = \$4$ e a quantidade de equilíbrio cai para $Q' = 2$ unidades. O excedente social sem a presença do agente $i = 5$ cai para $TS(-5) = \$3$, que é a área entre a curva de demanda agregada sem o agente $i = 5$ e a curva de oferta agregada. Portanto, a contribuição marginal do agente $i = 5$ para os ganhos sociais de troca é $TS - TS(-5) = \$7 - \$3 = \$4$. Mais uma vez, $CMS(5) = \pi(5)$.

Se repetirmos esse raciocínio para todos os agentes $i = 1, \ldots, 8$, vamos encontrar o seguinte resultado: $CMS(i) = \pi(i)$, para todo agente $i = 1, \ldots, 8$. Em outras palavras, *full appropriation*:

todos os agentes internalizam, na forma de excedente privado, as suas contribuições marginais para os ganhos sociais de troca desse mercado. Isso é competição perfeita!

Se considerarmos a função excedente total TS como uma função cujas variáveis são os agentes (e não bens e serviços), então cada agente será privadamente remunerado pelo seu impacto incremental sobre o excedente social. É como se, em termos discretos, a função excedente social fosse homogênea de grau 1 nos agentes. Isso nos recorda o teorema de Euler para funções homogêneas de grau 1, também conhecido por teorema da exaustão do produto: o valor do produto é distribuído integralmente entre os fatores de produção de acordo com os produtos marginais de cada um. A diferença é que os fatores, neste caso, são os agentes. Traduzindo para o nosso contexto: o excedente social é distribuído integralmente entre os agentes de acordo com os produtos marginais de cada um. É essa condição de *full appropriation* que caracteriza a competição: existe competição perfeita se, e somente se, existe *full appropriation*.

O mercado de nosso exemplo é, portanto, perfeitamente competitivo.

Uma caracterização equivalente à condição de *full appropriation* deve-se a Joan Robinson: competição perfeita ocorre quando os agentes enfrentam ofertas e demandas perfeitamente elásticas. O sentido dessa caracterização é este: mudanças ocupacionais unilaterais de agentes não afetam os preços de equilíbrio. Por mudanças ocupacionais entende-se, entre outras coisas, entrar e sair do mercado, ou seja, ajustes na margem extensiva. A literatura moderna de Equilíbrio Geral incorpora também mudanças de preferências ou tecnologias, de possibilidades de troca etc. Nosso exemplo é bem simples, mas capta perfeitamente esses avanços teóricos.

Outra caracterização equivalente de competição perfeita é esta: os agentes marginais na vizinhança do equilíbrio possuem opções externas perfeitamente substituíveis. Agentes marginais são aqueles cujas valorações marginais determinam o preço de equilíbrio. No nosso exemplo, os agentes marginais são os ofertantes $i = 2, 3$ e os demandantes $i = 6, 7$, os quais comercializam a terceira e a quarta unidades de equilíbrio.

Nas exposições de livro-texto, as funções são contínuas e o nível individual se dilui no agregado, tirando-nos a chance de enxergar a competição perfeita em equilíbrio, pois tudo se colapsa em um ponto. Na competição perfeita, o ponto é, na verdade, um trecho horizontal coincidente das curvas de oferta e demanda agregadas. Esse trecho horizontal será mais provável de existir quanto maior o número de agentes, pois assim também maior será a chance de que alguns agentes tenham valorações marginais iguais ou muito parecidas (sejam elas preços de demanda, isto é, disposições marginais a pagar, ou preços de oferta, isto é, custos marginais), de modo que, na soma horizontal, várias unidades sejam transacionadas ao mesmo preço. É nesse fato que reside a condição de *full appropriation*. Ela não é tão imediata assim, pois depende do que ocorre não só na infra-margem, mas com todos os agentes anteriores. Ela realça o aspecto essencialmente distributivo da competição perfeita. Já a condição equivalente de ofertas e demandas perfeitamente elásticas, elucidada por Joan Robinson nos anos 30, é muito patente no trecho horizontal que determina o preço de equilíbrio. Os agentes não conseguem alterar o preço de forma unilateral.

Um ponto que se faz óbvio com o exemplo é que a existência de grandes números de agentes não é condição inerente à competição perfeita, ao contrário do que vemos por aí e muitos repetem sem a devida ciência de como a Teoria evoluiu.

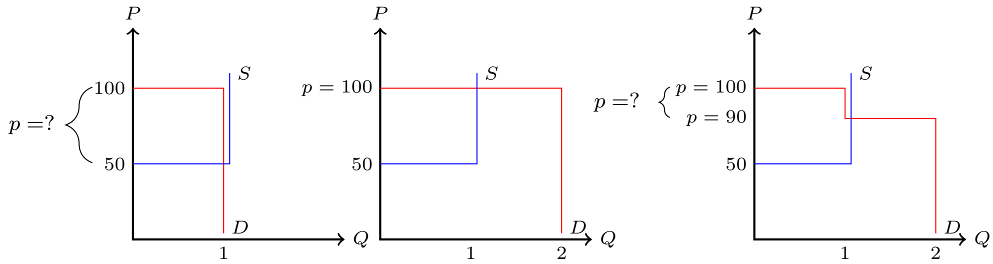
Note ainda que, sob competição perfeita, o excedente social é igual à soma dos preços-sombras dos agentes. Sabemos que, por definição, o excedente social é igual à soma dos excedentes privados. Existe um teorema em Equilíbrio Geral, devido a Joseph Ostroy e Louis Makowski ("Apropriation and efficiency: a revision of the first theorem of welfare economics", American Economic Review, 85: 808-827 (1995)), que diz que o excedente privado de qualquer agente é sempre majorado pelo seu preço-sombra. Sob competição perfeita, como há *full appropriation*, então o excedente fosse um bolo cujas fatias fossem determinadas pelas contribuições marginais de cada agente para o tamanho do bolo, seria possível repartir o bolo integralmente segundo esse critério.

A condição de *full appropriation* chamava-se antes no-surplus condition e surgiu no artigo seminal de Ostroy: "The no-surplus condition as a characterization of perfectly competitive equilibrium", Journal of Economic Theory, 22: 183-207 (1980). A grande contribuição de Ostroy para a Teoria do Equilíbrio Geral foi, tendo mostrado que o marginalismo possui dois pilares, o da mercadoria e o do indivíduo, ter erigido o pilar marginalista do indivíduo, que fora antes apenas abarcado apenas localmente por Vickrey, Hurwicz, Pigou, Coase, John Bates Clarke, Wicksteed e outros. (Ostroy: "A reformulation of the marginal productivity theory of distribution", Econometrica, 52: 599-630 (1984)).

4.5.2 Oferta da Firma no Curto Prazo

Na linguagem comum, "competição" está associada à rivalidade ativa entre alternativos vendedores potenciais e entre alternativos compradores potenciais. Mas na teoria do equilíbrio geral Walrasiana, "competição" está associada ao comportamento passivo de precificação (passive price-taking behavior). As duas visões são consistentes? Se a competição ativa entre potenciais vendedores e entre potenciais compradores for suficientemente intensa, quando a poeira for eliminada (ou seja, em equilíbrio), nenhum vendedor ou comprador poderá influenciar os termos de troca resultantes da competição ativa. Assim, em equilíbrio, as pessoas devem agir como tomadoras de preços. A competição ativa envolve uma forma particular de rivalidade. Mesmo sob monopólio bilateral, a rivalidade na forma de barganha sobre um preço pode ser muito intensa. A competição adiciona um novo elemento: opções externas. Para ilustrar o ponto, considere um mercado com apenas um vendedor e um comprador, o comprador está disposto a pagar até \$100 por uma unidade de um objeto indivisível, enquanto o vendedor está disposto a aceitar \$50 ou mais pelo objeto (seu custo de oportunidade = \$50). Supondo que eles façam um acordo, qualquer preço entre \$50 e \$100 é possível, o preço real dependerá de suas habilidades relativas de barganha, como ilustrado na Figura 4.6.

Figura 4.6 – A competição ativa envolve uma forma particular de rivalidade: rivalidade entre compradores alternativos e/ou vendedores alternativos – Sem competição, Competição perfeita, Competição imperfeita



Agora, considere o que acontece se outro comprador com disposição de pagar \$100 entrar no mercado. Mesmo que ele não seja um negociador muito duro, a situação do vendedor melhorou consideravelmente, pois agora ele tem a opção externa de negociar com o segundo comprador e não com o primeiro. Na verdade, ele pode jogar um comprador contra o outro. A competição entre os dois compradores forçará o preço até \$100, enquanto o vendedor se senta e aproveita o espetáculo. Inicialmente, a segunda situação pode até parecer envolver muito menos rivalidade do que a primeira: pode haver muito menos persuasão, bajulação ou vociferantes ameaças no ar na segunda situação, porque o primeiro comprador percebe que suas ameaças não mais ser eficazes; o vendedor agora tem a opção externa de lidar com o segundo comprador. Para desenvolver o exemplo mais um passo, suponha agora que a disposição do segundo comprador em pagar é de apenas \$90, em vez de \$100. Assim, o segundo comprador não é mais um substituto perfeito para o primeiro, no que diz respeito ao vendedor. Ele ainda pode jogar com um comprador contra o outro, mas agora os lances entre os compradores serão de \$90, em seguida, o segundo comprador vai cair fora da negociação. Assim, nesta variação permanece \$10 do excesso de monopólio bilateral para o vendedor e o comprador da primeira para negociar. Em relação à situação (a), a concorrência reduziu a faixa de indeterminação pelo preço final, mas não a eliminou. Exprimindo um pouco diferente, nesta variação, porque os dois compradores concorrentes têm diferentes disposições para pagar pelo objeto do vendedor, a competição entre eles não é mais “perfeita”: a concorrência perfeita requer competição entre compradores substitutos perfeitos.

Como a empresa competitiva deve ver o preço de mercado como determinado (no equilíbrio não há mais possibilidade de ganhar algo com a arbitragem), seu problema de maximização do lucro é o seguinte. A empresa determina o nível de produto y para resolver o seguinte problema:

$$\max_{y} py - c(y), \tag{4.9}$$

em que y é o produto produzido pela empresa, p é o preço do produto e $c(y)$ é a função de custo da produção.

A condição de primeira ordem para uma solução interior nos dá

$$p = c'(y) \equiv CM(y). \tag{4.10}$$

A condição de primeira ordem torna-se uma condição suficiente se a condição de segunda ordem é satisfeita

$$c''(y) > 0. \tag{4.11}$$

Tomadas em conjunto, essas duas condições determinam a função de oferta de uma empresa competitiva: a qualquer preço p, a firma fornecerá uma quantidade de produto $y(p)$ tal que $p = c'(y(p))$. Como $p = c'(y(p))$, temos

$$\begin{aligned} \frac{\partial}{\partial p} p &= \frac{\partial}{\partial p} c'(y(p)) \\ 1 &= c''(y(p))y'(p), \end{aligned} \tag{4.12}$$

e, portanto, $y'(p) > 0$, o que significa que a lei da oferta é válida.

Lembre-se que $p = c'(y^*)$ é a condição de primeira ordem que caracteriza o ótimo apenas se $y^* > 0$, isto é, somente se y^* é um ótimo interior. Pode acontecer que, a um preço baixo, uma empresa possa muito bem decidir sair do negócio. Para o caso de curto prazo,

$$c(y) = c_v(y) + c_f. \tag{4.13}$$

O lucro da firma é $\pi(y) = py(p) - c_v(y(p)) - c_f$ se $y > 0$ e $\pi(0) = -c_f$ se $y = 0$. Assim, a empresa deve produzir se $\pi(y) > \pi(0)$, isto é,

$$py(p) - c_v(y(p)) - c_f \geqq -c_f, \tag{4.14}$$

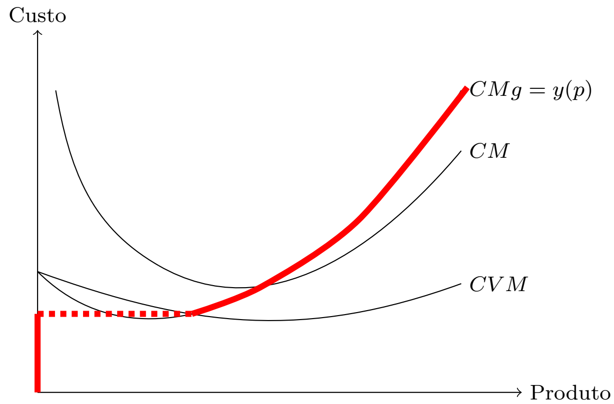
o que implica que

$$p \geqq \frac{c_v(y(p))}{y(p)} \equiv \text{CVM}, \tag{4.15}$$

ou seja, a condição necessária para a empresa produzir uma quantidade positiva de produto é que o preço seja maior ou igual ao custo variável médio.

Assim, a curva de oferta para a firma competitiva é em geral dada por: $p = c'(y)$ se $p \geqq \frac{c_v(y(p))}{y(p)}$ e $y = 0$ se $p \leqq \frac{c_v(y(p))}{y(p)}$. Ou seja, a curva de oferta coincide com a porção inclinada para cima da curva de custo marginal, desde que o preço cubra o custo variável médio, e a curva de oferta é zero se o preço for menor do que o custo variável médio.

Figura 4.7 – Curva de oferta da firma competitiva



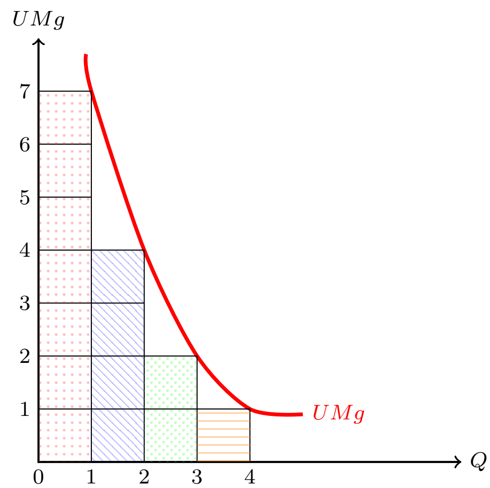
Se é no longo prazo, não há custo fixo e, portanto, o custo variável médio coincide com o custo médio; se é no curto-prazo, há diferença.

Vamos ver a curva de oferta de outra forma. Ela mede, na verdade, o valor que a sociedade atribui aos recursos econômicos que são deslocados para a produção do bem em questão. Para tornar inteligível o significado da curva de oferta, vamos conectar com o princípio da utilidade marginal decrescente. Sim, a curva de oferta é crescente e convexa precisamente por causa do princípio da utilidade marginal decrescente.

Para produzir um bem, o produtor se vale de insumos. Esses insumos são recursos existentes na economia. Não importa se você os chama de trabalho, capital ou terra ou qualquer outra coisa. São recursos escassos que a sociedade valoriza precisamente porque são escassos. Quando o produtor produz uma unidade do bem, ele tem que deslocar recursos da economia para a produção dessa unidade. Imagine que os recursos estão disponíveis em pacotinhos de igual magnitude. Cada pacotinho é um quantum de recursos contendo bens e serviços que, aos olhos do produtor, são vistos como insumos ou fatores de produção. Para simplificar, suponha que existam 4 quanta (plural de quantum) de recursos na sociedade.

Isso é o que a sociedade tem de recursos. Como eles são escassos, a sociedade os valoriza. É aqui que entra o princípio da utilidade marginal decrescente. Se a sociedade não tivesse nada e adquirisse o 1º quantum de recursos, isso lhe daria uma imensa satisfação, uma utilidade incremental pela qual ela estivesse disposta a pagar, no máximo, digamos, \$7. Ao adquirir o 2º quantum de recursos, a sociedade também fica feliz, mas numa intensidade menor. Suponha que ela estivesse disposta a pagar por esse 2º quantum \$4. Tendo adquirido o 2º quantum, ela adquire o 3º e por ele está disposta a pagar \$2. O 4º quantum vale para a sociedade \$1. Note que expressei a satisfação pelos quanta em termos monetários. É que estou supondo efeito-renda nulo, só para simplificar, de modo que a utilidade marginal possa ser expressa como valor monetário. Os valores acima ilustram o princípio da utilidade marginal decrescente e estão dispostos na Figura 4.8:

Figura 4.8 – Utilidade marginal da sociedade



Esses recursos estão, neste momento, sendo usados pelas pessoas na economia em suas trocas (produções e consumos). Você, que faz parte da economia, decide, de repente, produzir um bem. Para produzir a 1ª unidade, você tem que deslocar um quantum de recursos da economia para o seu processo de produção.

Dos 4 quanta de recursos de que a sociedade dispunha, ela tem que abrir mão de 1 quantum para que você produza a 1ª unidade de seu bem. Quanto valem esses recursos deslocados da sociedade? Valem \$1. Você tem de ler o gráfico da utilidade marginal da sociedade de trás para a frente. Ela não está numa situação em que aos poucos adquire unidades adicionais, mas numa em que aos poucos perde as que já possui. Ela tem bastante e não se importa tanto em perder um pouco. Assim, o custo econômico social da 1ª unidade produzida é $C(1) = \$1$. Se não há externalidades, esse custo incremental social é exatamente igual ao seu custo incremental privado. Como isso é possível? É possível desde que os preços sejam livremente determinados pelo mercado segundo a lei de oferta e demanda, pois assim os preços refletirão as valorações marginais das unidades comercializadas. Como o custo de produzir zero é $C(0) = \$0$ (já que nada é deslocado da economia nesse caso), então o custo incremental da 1ª unidade é $CMg(1) = \$1$.

Qual é o preço que você deve cobrar por essa 1ª unidade? Deve ser o preço que cubra o valor do qual a sociedade abriu mão: $P = \$1$. Nem adianta cobrar mais, pois a receita incremental não seria igual ao custo incremental e, portanto, não seria uma boa ideia: você teria tomado a decisão errada.

Você agora quer produzir a 2ª unidade. Como você já deslocou um quantum de recursos da economia, privando-a de um pacotinho, tem de deslocar um segundo quantum.

Para a sociedade, esse 2º quantum vale \$2. Esse valor é o custo marginal social da 2ª unidade: $CMg(2) = \$2$. De fato, com a 2ª unidade deslocada, a sociedade já abriu mão de $\$3 = \$1 + \$2$. Com apenas a 1ª, ela tinha aberto mão de \$1. A diferença é \$2.

Qual é o preço que você deve cobrar por essa 2ª unidade? Deve ser o preço que cubra o valor incremental do qual a sociedade abriu mão: $P = \$2$.

Você agora quer produzir a 3ª unidade. Como você já deslocou 2 quanta de recursos da economia, privando-a de dois pacotinhos, tem de deslocar um 3º quantum.

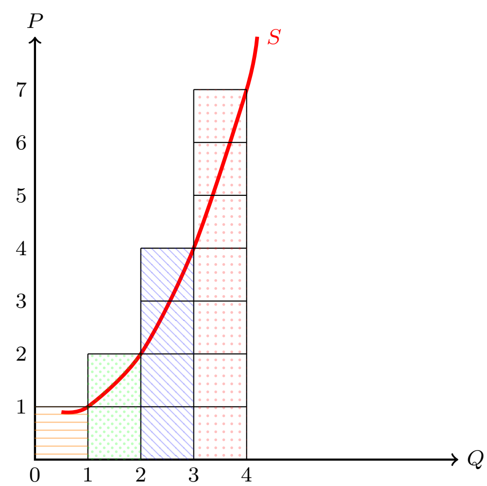
Para a sociedade, esse 3º quantum vale \$4. Esse valor é o custo marginal social da 3ª unidade, $CMg(3) = \$4$, e, pelas mesmas razões anteriores, o preço que você deve cobrar por essa 3ª unidade é $P = \$4$.

Finalmente, você quer produzir a 4ª e última unidade. Como você já deslocou 3 quanta de recursos da economia, privando-a de três pacotinhos, tem de deslocar agora o 4º quantum.

Para a sociedade, esse 4º quantum vale \$7. Esse valor é o custo marginal social da 4ª unidade: $CMg(4) = \$7$ e, pelas mesmas razões anteriores, o preço que você deve cobrar por essa 4ª unidade é $P = \$7$. A sociedade ficou sem nada, voltou à estaca zero. Se ela pudesse reaver um pacotinho, seria o 1º pacotinho em relação a zero. Como vimos, esse 1º pacotinho vale \$7 para a sociedade.

Temos, finalmente, a sua curva de oferta (ou de custo marginal privado, que coincide com o custo marginal social), como pode ser visto na Figura 4.9.

Figura 4.9 – Curva de oferta



Observe como a curva de oferta é o reflexo especular da curva de utilidade marginal da sociedade sobre os recursos escassos disponíveis. A conexão é entre a curva de oferta e a valoração marginal que a sociedade inteira atribui aos recursos escassos disponíveis. A ideia de custo marginal como explicadora da curva de oferta é uma ideia intermediária. Ela é a parte visível, mas por trás dela está a sociedade. No fundo, ler a curva de oferta dessa maneira equivale também a compreender o que é a função custo econômico, cuja diferença em relação ao custo contábil torna-se assim patente. A valoração marginal que a sociedade atribui a cada pacotinho condensa o valor que a sociedade atribui a tudo aquilo que pode resultar do uso ótimo de cada pacotinho. Se você desloca um pacotinho e o vende por um preço igual a essa valoração marginal, você está simplesmente deixando a sociedade indiferente entre (a) usar esse pacotinho para os usos que ela tinha em mente e (b) deixar esse pacotinho na sua mão para que você produza um bem que ela valoriza tanto quanto o pacotinho deslocado.

A demanda tem uma natureza essencialmente privada. Já a produção, além de seu caráter privado, possui também uma natureza social, que se manifesta na relação entre a curva de oferta privada e a valoração marginal da sociedade sobre os recursos deslocados para a produção do bem. Como dizia William Stanley Jevons e, depois dele, John Bates Clark, a produção é um fenômeno social e, por isso, distinto da demanda.

Suponha que tenhamos J firmas no mercado. A função de oferta da indústria é simplesmente a soma das funções de oferta de todas as firmas, de modo que é dada por:

$$\widehat{y}(p) = \sum_{j=1}^{J} y_j(p), \tag{4.16}$$

em que $y_j(p)$ é a função de oferta da firma j para $j = 1, \ldots, J$. Como cada empresa escolhe um nível de produção em que o preço é igual ao custo marginal, cada empresa que produz uma quantidade positiva de produto deve ter o mesmo custo marginal. A função de oferta da indústria mede a relação entre a produção da indústria e o custo comum de produzir esse produto. A função de demanda agregada (indústria) mede a produção total demandada a qualquer preço que é dado por

$$\widehat{x}(p) = \sum_{i=1}^{n} x_i(p), \tag{4.17}$$

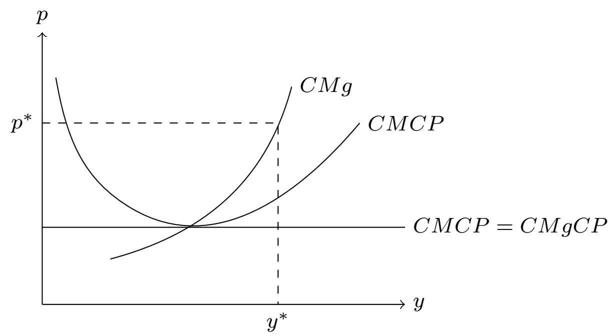
em que $x_i(p)$ é a função de demanda do consumidor i para $i = 1, \ldots, n$.

4.5.3 Concorrência no Longo Prazo

O comportamento de longo prazo de uma indústria competitiva é determinado por dois tipos de efeitos. O primeiro efeito é o fenômeno de entrada e saída livre, de modo que os lucros obtidos por todas as empresas sejam zero. Se alguma empresa está tendo lucros negativos, esperamos que eventualmente tenha que mudar seu comportamento. Por outro lado, se uma empresa está obtendo lucros positivos, esperamos que isso acabe incentivando a entrada no setor. Se temos uma indústria caracterizada pela entrada e saída sem custo, fica claro que, a longo prazo, todas as empresas devem obter o mesmo nível de lucros. Como resultado, toda empresa ganha lucro zero no equilíbrio competitivo de longo prazo.

A segunda influência no comportamento de longo prazo de uma indústria competitiva é a do ajuste tecnológico. A longo prazo, as empresas tentarão ajustar seus fatores fixos de modo a produzir o nível de equilíbrio da produção da maneira mais barata. Suponha, por exemplo, que tenhamos uma empresa competitiva com uma tecnologia constante de retornos à escala de longo prazo que esteja operando na posição ilustrada na Figura 4.10. Então, a longo prazo, é claro que a empresa paga para mudar seus fatores fixos de modo a operar em um ponto de custo médio mínimo. Mas, se toda empresa tentar fazer isso, o preço de equilíbrio certamente mudará.

Figura 4.10 – Ajustamento de longo prazo com custos constantes



4.6 Monopólio Puro

A partir de agora vamos supor que os indivíduos se comportem estrategicamente; por exemplo, os compradores prometem pagar o mínimo que podem. Da mesma forma, os vendedores negociam muito para conseguir o máximo que podem. Para ilustrar, em vez de agir como um tomador de preços e optar pela "precificação de custo marginal", um vendedor que enfrente uma curva de demanda negativamente inclinada pode reduzir sua produção para obter um preço unitário mais alto. Ou ele pode tentar discriminar os preços entre os clientes, para aumentar ainda mais seu lucro. Por outro lado, os compradores tentarão resistir aos esforços do vendedor para discriminar preços fingindo valorizar seu bem menos do que realmente fazem, para aumentar o excedente de seus consumidores. Assim, as barganhas dos compradores geralmente acarretam "problemas de revelação" para o vendedor.

A maioria das empresas consegue escolher a que preço vendem seus bens? Vamos agora analisar explicitamente essa situação. Vamos começar pensando em um mercado no qual há uma empresa monopolista, e ela escolhe o preço pelo qual vende seus produtos. O que deve impedir uma empresa desse tipo de ganhar uma quantia infinita de dinheiro (e assim tornar nosso problema chato)? Bem, é claro que teremos que abandonar a suposição de que o monopolista pode vender o quanto quiser pelo preço que escolher. Em vez disso, o monopolista será limitado pela curva de demanda do produto que está vendendo. Isso diz a eles quanto podem vender a qualquer preço: o problema do monopolista pode, portanto, ser visto como o problema de selecionar a localização que maximiza o lucro na curva de demanda. Antes de analisarmos tal situação, vale a pena tirar um minuto para pensar em como tal situação pode surgir. Como poderíamos ter uma indústria em que há apenas uma empresa, fazendo lucros positivos? Normalmente, existem duas razões dadas para a existência de monopólios:

- Barreiras à entrada: em alguns casos, pode ser impossível que outras empresas entrem no mercado. Considere, por exemplo, uma empresa que detenha a patente de um determinado medicamento. Outras empresas podem querer começar a vender a mesma droga, mas elas simplesmente não podem por lei.

- Aumentar os retornos de escala: se a estrutura de custos em um determinado setor exibir retornos crescentes de escala (custos marginais decrescentes), pode ser impossível para o setor suportar mais de uma empresa: se houvesse duas empresas, cada uma estaria sempre tentando crescer; portanto, tornar-se mais eficiente e forçar o outro a sair do negócio. Exemplos típicos de tais indústrias são aqueles em que há custos fixos muito elevados - por exemplo, o fornecimento de eletricidade.

4.6.1 Precificação Homogênea

O objetivo da empresa é maximizar o lucro. No entanto, o preço que o monopolista cobra afeta a quantidade que vende. A relação entre a quantidade vendida e o preço cobrado é regida pela curva de demanda (agregada), $q(p)$. Note, para focar na relação entre q e p, suprimimos os argumentos de riqueza na função de demanda agregada.

Podemos assim afirmar o problema do monopolista da seguinte forma:

$$\max_q pq(p) - c(q(p)). \tag{4.18}$$

Note, no entanto, que existe uma correspondência de um para um entre o preço cobrado e a quantidade que o monopolista vende. Assim, podemos reescrever o problema em termos de quantidade vendida, em vez do preço cobrado. Seja $p(q)$ a função de demanda inversa. Ou seja, $p(q(p)) = p$. O problema de maximização do lucro da empresa pode então ser escrito como

$$\max_q p(q)q - c(q). \tag{4.19}$$

Acontece que normalmente é mais fácil olhar para o problema em termos de definir a quantidade e deixar que o preço seja determinado pelo mercado. Por esse motivo, usaremos a abordagem de definição de quantidade.

Para que a solução seja única, precisamos que a função objetivo seja estritamente côncava, ou seja, $\frac{d^2}{dq^2} < 0$. Temos que

$$\frac{d}{dq} = 0 \quad \Longleftrightarrow \quad p(q) + qp'(q) - c'(q). \tag{4.20}$$

A segunda derivada do lucro em relação a q é dada por

$$\frac{d^2}{dq^2} = p''(q)q + 2p'(q) - c''(q). \tag{4.21}$$

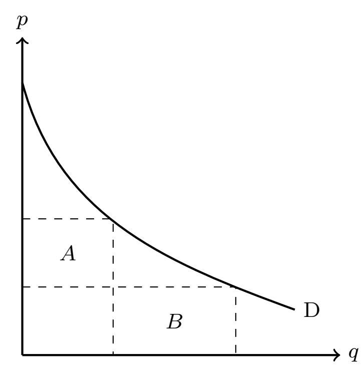
Se o custo é estritamente convexo, temos $c''(q) > 0$. Como a demanda se inclina para baixo, $p'(q) < 0$. Portanto, o segundo e terceiro termos são negativos. Por causa disso, não precisamos que a demanda inversa seja côncava. No entanto, não pode ser “muito convexo”. De modo geral, vamos supor que a função objetivo é côncava sem fazer suposições adicionais sobre $p(\cdot)$. Na verdade, para garantir que a quantidade maximizadora seja finita, precisamos assumir que, eventualmente, os custos se tornam grandes o suficiente em relação à demanda. Isso sempre será satisfeito se, por exemplo, as curvas de demanda e de custo marginal se cruzam.

A função objetivo é maximizada observando a primeira derivada. Na quantidade ideal, q^*, encontramos:

$$p'(q^*)q^* + p(q^*) = c'(q^*). \tag{4.22}$$

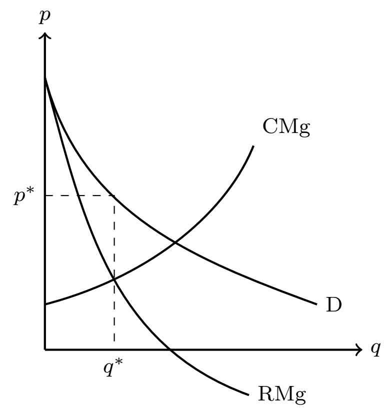
No lado esquerdo da expressão está a receita marginal do aumento da produção. Esse aumento tem duas partes – a receita adicional devido à venda de mais uma unidade, $p(q^*)$ (área B na Figura 4.11), e a diminuição da receita devido ao fato de a empresa receber um preço mais baixo em todas as unidades que vende (área A na Figura 4.11). Assim, a quantidade ótima do monopolista é onde a receita marginal é igual ao custo marginal, e o preço é definido pela curva de demanda $p(q^*)$[3]. Ver Figura 4.12 para uma descrição gráfica do ótimo.

Figura 4.11 – Receita marginal do monopolista



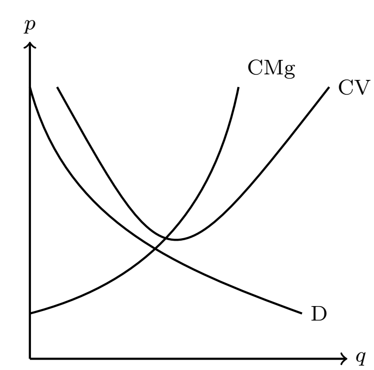
[3] Isso também é verdade para a empresa competitiva. No entanto, como as empresas competitivas são tomadoras de preço, sua receita marginal é igual ao preço.

Figura 4.12 – Preço e quantidade ótima do monopolista



Se o lucro do monopolista é maximizado em $q = 0$, devemos ter $p(0) \leq c'(0)$. Isso corresponde ao caso em que o custo de produzir até mesmo a primeira unidade é maior do que o que os consumidores estão dispostos a pagar. Geralmente, assumimos que $p(0) > c'(0)$ para focarmos no caso interessante em que o monopolista deseja produzir uma quantidade positiva de produto. No entanto, mesmo se assumirmos que $p(0) > c'(0)$, o monopolista pode não querer escolher um nível de produto positivo. Pode ser que "fechar" ainda seja preferível a produzir um produto positivo. Ou seja, o monopolista pode ter custos fixos tão grandes que preferiria sair do setor. Tal situação é ilustrada na Figura 4.13. Assim, interpretamos a condição $p(0) > c'(0)$ (juntamente com as condições de segunda ordem apropriadas) dizendo que, se o monopolista não sair da indústria, produzirá uma quantidade positiva de produto.

Figura 4.13 – Caso em que o monopolista não apresenta lucro



Se houver um máximo em um nível positivo de produto, deve ser tal que a primeira derivada seja igual a zero ou:

$$p'(q^*)q^* + p(q^*) = c'(q^*). \tag{4.23}$$

Note que podemos reescrever o lado esquerdo da equação acima como:

$$p'(q^*)q^* + p(q^*) = p(q^*)\left(\frac{dp(q^*)}{dq^*}\frac{q^*}{p(q^*)} + 1\right) = p(q^*)\left(1 - \frac{1}{|\varepsilon_p^*|}\right), \tag{4.24}$$

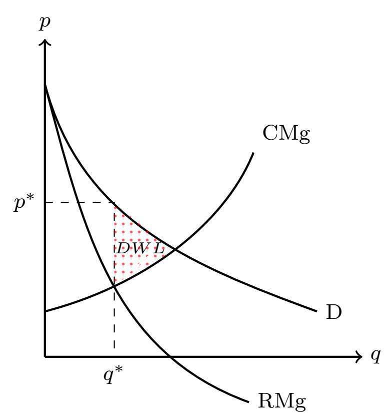
em que ε_p^* é a elasticidade-preço da demanda avaliada em (q^*, p^*). Agora podemos reescrever a condição de primeira ordem do monopolista como:

$$\frac{p(q^*) - c'(q^*)}{p(q^*)} = p'(q^*)\frac{-q^*}{p(q^*)} = \frac{1}{|\varepsilon_p^*|}. \tag{4.25}$$

A quantia $\frac{p(q^*) - c'(q^*)}{p(q^*)}$ é o mark-up do preço sobre o custo marginal e é expressa como uma fração do preço. Essa quantidade, chamada de índice de Lerner, é frequentemente usada para medir o grau de poder de mercado em uma indústria.

Note que na quantidade onde $RMg = CMg$, temos $p > CMg$ (já que $p'(q)q$ é negativo). Assim, o monopolista cobra mais do que o custo marginal[4]. O ideal social seria o monopolista vender desde que os consumidores estejam dispostos a pagar mais pela última unidade produzida do que o custo para produzi-la. Ou seja, produzir até o ponto em que $p = CMg$. Mas, o monopolista reduz a produção porque se preocupa com o lucro, não com a otimização social. E está disposto a reduzir q para aumentar o lucro. Isso resulta no que é conhecido como peso morto do monopólio, e é igual à área entre a curva de demanda e a curva de custo marginal, e à direita da quantidade ótima, como na Figura 4.14. Esta área representa o excedente social que pode ser gerado, mas não está no resultado do monopólio.

Figura 4.14 – Peso morto do monopolista



Assim, o preço cobrado pelo monopolista será:

$$p(q^*) = c'(q^*)\left(\frac{|\varepsilon_p^*|}{|\varepsilon_p^*| - 1}\right). \tag{4.26}$$

[4] Além disso, a partir dessas duas equações, podemos mostrar que o monopolista sempre escolherá uma quantidade tal que o preço seja elástico, ou seja, $|\varepsilon_p^*| > 1$.

4.6.2 Precificação Não Homogênea

O fato de que o monopolista vende menos do que a quantidade idealmente social da produção surge da exigência de que o monopolista venda todos os bens pelo mesmo preço. Assim, se quiser ganhar lucros maiores em qualquer item em particular, deve aumentar o preço de todos os itens, o que diminui a quantidade vendida. Se o monopolista pudesse aumentar o preço de alguns itens, mas não de outros, poderia obter maiores lucros e ainda vender a quantidade eficiente. Consideramos agora dois exemplos de mecanismos de preços mais complicados ao longo destas linhas, preços não lineares e tarifas de duas partes.

4.6.2.1 Precificação Não Linear

Considere o caso em que o monopolista tem um sistema de preços em que cada unidade é vendida por um preço diferente.

Por enquanto, assumimos que todos os consumidores são idênticos e que os consumidores não podem revender itens depois de comprá-los. Nesse caso, o monopolista resolve seu problema de maximização de lucro projetando um esquema que maximiza o lucro obtido com qualquer consumidor e, em seguida, aplica esse esquema a todos os consumidores.

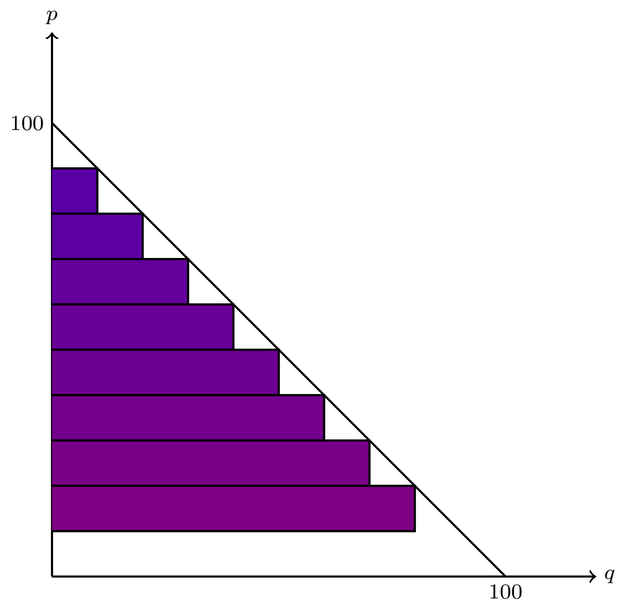
No contexto do modelo quase-linear, sabemos que a altura da função de demanda de um consumidor em uma determinada quantidade representa sua utilidade marginal para aquela unidade de produção. Em outras palavras, a altura da curva de demanda representa o máximo que um consumidor pagará por essa unidade de produção.

Com isto em mente, se o monopolista vai cobrar um preço diferente por cada unidade de produção, como deve definir esse preço? Obviamente, ele quer definir o preço de cada unidade igual à disposição máxima do consumidor em pagar por aquela unidade, ou seja, igual à altura da curva de demanda naquele q. Se o monopolista emprega um esquema de preços decrescente, em que $p(q) = D^{-1}(q)$, até o ponto em que a demanda e o custo marginal se cruzam, o monopolista pode realmente extrair todo o excedente social. Observe que se estamos preocupados que o produto só possa ser vendido em unidades inteiras, então o preço da unidade q deve ser dado por:

$$p_q = \int_{q^{-1}}^{q} D^{-1}(q)dq. \tag{4.27}$$

Para ilustrar os preços não lineares, considere um consumidor que tenha curva de demanda $P = 100 - Q$ e suponha que o custo marginal do monopolista seja igual a 10. A um preço de 90, o consumidor exige 10 unidades de produção. Enquanto o consumidor não compraria mais nenhuma unidade a um preço de 90, compraria mais se o preço fosse mais baixo. Assim, suponha que o monopolista venda as 10 primeiras unidades de produção a um preço de 90, e as segundas 10 unidades a um preço de 80. Similarmente, suponha que o monopolista venda unidades de 21 a 30 a um preço de 70, 31-40 a um preço de 60, 41-50 a um preço de 50, etc., até as unidades 71-80, que são vendidas a um preço de 20. Isso produz a Figura 4.15.

Figura 4.15 – Preço não-linear

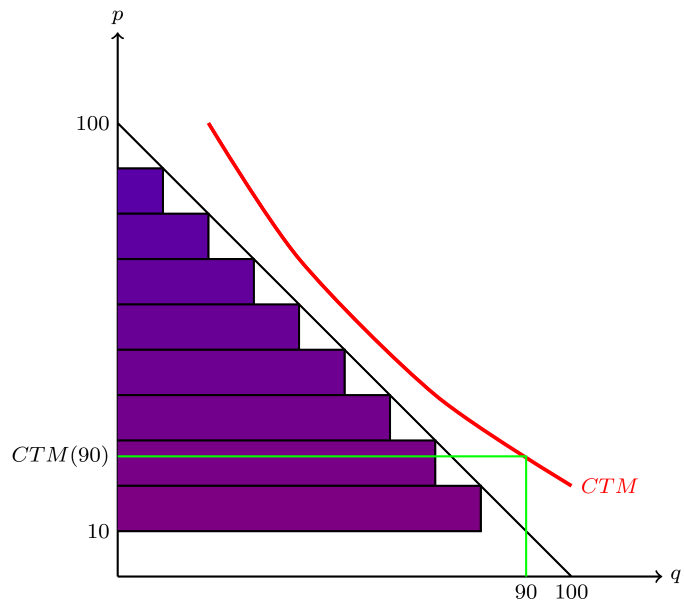


O excedente de produção do monopolista é igual à região sombreada. Assim, diminuindo o preço à medida que o número de unidades compradas aumenta, o monopolista pode se apropriar de grande parte do excedente do consumidor. Na verdade, à medida que o número de "barras" no esquema de preços aumenta, o excedente do produtor se aproxima de todo o triângulo delimitado pela curva de demanda e pelo custo marginal. Assim, se o monopolista foi capaz de cobrar do consumidor $p(q) = \int_{q^{-1}}^{q} D^{-1}(q)dq$ para o bloco de produto consistindo da unidade q, ele se apropriaria de todo o excedente do consumidor.

Na prática, o preço de bloco em declínio é encontrado com mais frequência nos preços de serviços públicos, em que um grande comprador pode ser cobrado por um preço alto pelas primeiras X unidades de produção, um preço mais baixo pelas próximas X unidades e assim por diante.

Lembre-se de que dissemos que um monopolista que cobra um único preço só pode obter um lucro positivo se a curva de demanda estiver acima da curva de custo total médio (CTM) da empresa em algum momento. Se a empresa pode se envolver em preços não lineares, isso não é mais o caso. Como a receita da empresa é toda a região sombreada no diagrama acima, pode ser possível para a empresa obter um lucro positivo, mesmo se a curva de demanda estiver totalmente abaixo da curva do CTM. Por exemplo, se o CTM no exemplo anterior é como na Figura 4.16, então o monopolista pode ter lucro sempre que a área da região sombreada for maior que $CT(90) = CTM(90) \times 90$.

Figura 4.16 – Lucro mesmo quando a demanda está abaixo do CTM



4.6.2.2 Tarifa em Duas Partes

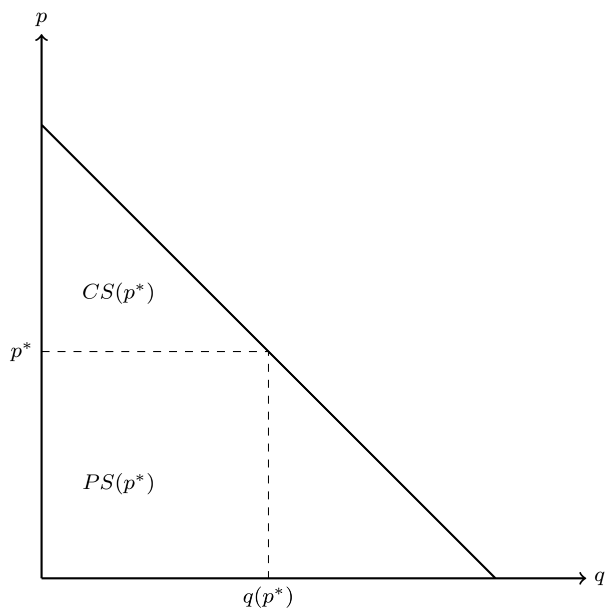
Dizemos que uma empresa pratica tarifa em duas partes caso ela cobre um preço, chamado tarifa de acesso, independente da quantidade consumida pelo acesso ao produto mais um preço constante por unidade consumida. Exemplos: alguns serviços de telefonia; provedor de banda larga (tarifa de acesso positiva e preço nulo); bares e restaurantes com couvert.

A questão que queremos abordar é como, nesse contexto, a taxa fixa (que chamaremos de F) e a taxa de uso (que chamaremos de p) devem ser definidas. Como no exemplo de precificação não-linear, continuamos a supor que todos os consumidores são idênticos e que o monopolista produz um produto a um custo marginal constante. Para simplificar, assuma que $CMg = 0$, mas observe que um custo marginal positivo é facilmente incorporado ao modelo. Além disso, continuamos a pensar no consumidor como tendo utilidade quase-linear. Nesse caso, sabemos que a curva de demanda inversa do consumidor, $p(q)$, dá ao consumidor a utilidade marginal de consumir a unidade q, e que o benefício do consumidor é dado pelo excedente do consumidor nessa quantidade, $CS(q)$.

O lucro da empresa com base em duas partes (F, p) é dado por $F + p^*q(p)$. Nós dividimos nossa análise em duas etapas. Primeiro, para qualquer p, como deve ser definido F? Em segundo lugar, qual p deve ser escolhido? Como o lucro do monopolista é crescente em F, o monopolista quer definir o maior valor possível para o F, ao mesmo tempo em que induz o consumidor a participar. Ou seja, o benefício líquido do consumidor deve ser pelo menos zero.

Ao preço p, o consumidor escolhe consumir $q(p)$ unidades de produto e ganha o excedente $CS(q(p))$. Portanto, o excedente líquido é $CS(q(p)) - F$, e queremos que isso seja não-negativo: $CS(q(p)) - F \geq 0$. Como discutido acima, a firma quer definir F tão grande quanto possível. Portanto, para qualquer p, a configuração $F(p) = CS(q(p))$ maximiza o lucro. Isso faz sentido, já que a empresa quer definir a taxa de acesso para extrair todo o excedente do consumidor. Em seguida, como p deve ser escolhido? Dado $F(p)$ como definido acima, o lucro do monopolista da tarifa de duas partes é dado por: $CS(q(p)) + pq(p)$. Observe que essa é a soma do excedente do consumidor e do produtor (veja a Figura 4.17). Como é facilmente visto no diagrama, esta expressão é maximizada definindo a taxa de acesso igual a zero (custo marginal) e a taxa fixa igual a $CS(q(0))$. Assim, a tarifa ótima em duas partes envolve $p = 0$ e $F = CS(0)$.

Figura 4.17 – Tarifa em duas partes



Podemos entender a tarifa ótima em duas partes de outra maneira. Quando o monopolista escolhe F de maneira ideal, ele reivindica todo o excedente criado no mercado. Assim, o monopolista tem um incentivo para maximizar o excedente criado nesse mercado. Pelo primeiro teorema do bem-estar, sabemos que o excedente total é maximizado pelo resultado perfeitamente competitivo. Isto é, quando p é tal que $p^* = CMg$. Assim, para maximizar o excedente total, a firma deve definir $p = CMg$. Se isso acontecer, o "excedente do produtor" é zero. No entanto, a empresa define $F = CS(q(p^*))$ e extrai todo o excedente social na forma da taxa fixa.

4.6.3 Discriminação de Preços

Seção baseada em Varian, H. (1989). "Price Discrimination". In: Handbook of Industrial Organization, volume I, edited by R. Schmalensee and R.D. Willig. Elsevier Science Publishers B. V.

A discriminação de preços é uma das formas mais comuns de práticas de marketing. Pode-se ocasionalmente duvidar que as empresas realmente se envolvam em alguns dos sofisticados raciocínios estratégicos que os economistas gostam de examinar, mas não há dúvida de que as empresas estão bem cientes dos benefícios da discriminação de preços. Considere, por exemplo, a seguinte passagem retirada de um folheto publicado pelo Boston Consulting Group:

> Um passo fundamental é evitar o preço médio. Precificação para grupos de clientes específicos deve refletir o verdadeiro valor competitivo do que está sendo fornecido. Quando isso é alcançado, nenhum dinheiro é deixado na mesa desnecessariamente, por um lado, enquanto não há oportunidades abertas para os concorrentes. O preço é uma ação precisa e confiável que aproveita ao máximo a combinação da sensibilidade ao preço dos clientes e dos fornecedores alternativos que eles têm ou poderiam ter [Miles (1986)].

Embora um economista possa ter usado um pouco mais de terminologia técnica, as idéias centrais de discriminação de preços são bastante evidentes nessa passagem. Cada livro de microeconomia de graduação contém uma lista de exemplos de discriminação de preços; as ilustrações mais populares parecem ser as de descontos para estudantes, descontos para idosos e similares. Dada a prevalência da discriminação de preços como fenômeno econômico, é surpreendentemente difícil chegar a uma definição inteiramente satisfatória.

A definição convencional é que a discriminação de preços está presente quando a mesma mercadoria é vendida a preços diferentes para diferentes consumidores. No entanto, essa definição falha em dois aspectos: preços diferentes cobrados de diferentes consumidores poderiam simplesmente refletir os custos de transporte ou custos semelhantes de venda do bem; e a discriminação de preços pode estar presente mesmo quando todos os consumidores recebem o mesmo preço - considere o caso de um preço de entrega uniforme. Vamos ver a definição de Stigler (1987): a discriminação de preços está presente quando dois ou mais bens similares são vendidos a preços que estão em proporções diferentes dos custos marginais. Como ilustração, Stigler usa o exemplo de um livro que vende em capa dura por R$ 15 e em brochura por R$ 5. Aqui, argumenta ele, há uma presunção de discriminação, uma vez que os custos não são suficientes para explicar a diferença de preço. Naturalmente, essa definição ainda deixa em aberto o significado preciso de "similar", mas a definição será útil para nossos propósitos.

Três condições são necessárias para que a discriminação de preços seja uma solução viável para o problema de preço da empresa. Primeiro, a empresa deve ter algum poder de mercado. Em segundo lugar, a empresa deve ter a capacidade de classificar os clientes. E terceiro, a empresa deve ser capaz de impedir a revenda. Discutiremos brevemente cada um desses pontos e os desenvolveremos com muito mais detalhes no decorrer da seção.

Nós nos voltamos primeiro para a questão do poder de mercado. A discriminação de preços surge naturalmente na teoria do monopólio e do oligopólio. Sempre que um bem é vendido a um preço superior ao seu custo marginal, há um incentivo para se envolver na discriminação de preços. Pois dizer que o preço está acima do custo marginal é dizer que há alguém disposto a pagar mais do que o custo de produção por uma unidade extra do bem. A redução do preço para todos os consumidores pode não ser rentável, mas a redução do preço apenas para o consumidor marginal provavelmente será lucrativa. Para diminuir o preço apenas para o consumidor marginal, ou mais geralmente para alguma classe específica de consumidores, a empresa deve ter uma maneira de separar os consumidores. O caso mais fácil é quando a empresa pode ordenar explicitamente os consumidores em relação a alguma categoria exógena, como a idade. Uma análise mais complexa é necessária quando a empresa deve discriminar preços com base em alguma categoria endógena, como o momento da compra. Nesse caso, o monopolista enfrenta o problema de estruturar seus preços para que os consumidores se "auto-selecionem" em categorias apropriadas. Finalmente, se a empresa pretende vender a preços diferentes para diferentes consumidores, a empresa deve ter uma maneira de impedir que os consumidores que compram a um preço com desconto e revendam para outros consumidores. Carlton e Perloff (1990) discutem vários mecanismos que podem ser usados para evitar a revenda:

1. Alguns bens, como serviços, energia elétrica, etc., são difíceis de revender por causa da natureza do bem.

2. Tarifas, impostos e custos de transporte podem impor barreiras à revenda. Por exemplo, é comum os editores venderem livros a preços diferentes em países diferentes e confiarem nos custos de transporte ou tarifas para restringir a revenda.

3. Uma empresa pode legalmente restringir a revenda. Por exemplo, os fabricantes de computadores geralmente oferecem descontos educacionais, juntamente com uma disposição contratual que restringe a revenda.

4. Uma empresa pode modificar seu produto. Por exemplo, algumas empresas vendem edições de estudantes de software com capacidades mais limitadas do que as versões padrão.

O analista econômico está, naturalmente, interessado em ter um modelo detalhado e preciso do comportamento da empresa. Mas, além disso, o economista quer ser capaz de julgar esse comportamento. Até que ponto a discriminação de preços de vários tipos promove o bem-estar econômico? Que tipos de discriminação devem ser encorajados e quais tipos são desencorajados? A discriminação de preços é ilegal apenas na medida em que "reduz substancialmente a concorrência". Como devemos interpretar essa frase? Estas são algumas das questões que examinaremos nesta seção.

4.6.3.1 Tipos de Discriminação de Preços

A classificação tradicional das formas de discriminação de preços deve-se a Pigou (1920).

Discriminação de preço de primeiro grau (ou perfeita) envolve o vendedor cobrando um preço diferente para cada unidade do bem, de tal forma que o preço cobrado por cada unidade seja igual à disposição máxima a pagar por aquela unidade.

Discriminação de preço de segundo grau, ou precificação não linear, ocorre quando os preços diferem dependendo do número de unidades do bem comprado, mas não dos consumidores. Ou seja, cada consumidor enfrenta o mesmo esquema de preços, mas o esquema envolve preços diferentes para quantidades diferentes do bem adquirido. Descontos de quantidade ou prêmios são os exemplos óbvios.

Discriminação de preço de terceiro grau significa que diferentes compradores recebem preços diferentes, mas cada comprador paga uma quantia constante por cada unidade do bem comprado. Esta é talvez a forma mais comum de discriminação de preços; exemplos são descontos para estudantes, ou cobrar preços diferentes em dias diferentes da semana.

Seguiremos a classificação de Pigou nesta aula, discutindo as formas de discriminação de preços na ordem em que ele as sugeriu. Posteriormente, abordaremos alguns tópicos mais especializados que não parecem se encaixar convenientemente nesse esquema de classificação.

4.6.3.2 Discriminação de Preços de Primeiro Grau

Discriminação de preço de primeiro grau, ou discriminação de preço perfeita, significa que o vendedor vende cada unidade do bem ao preço máximo que alguém está disposto a pagar por aquela unidade do bem. Alternativamente, a discriminação perfeita de preços é às vezes definida como ocorrendo quando o vendedor faz uma única oferta a cada consumidor que extrai o máximo possível do mercado.

Embora a equivalência dessas duas definições tenha sido afirmada há muito tempo - Pigou a menciona em sua discussão sobre a discriminação de preço em primeiro grau - não está inteiramente claro como geralmente as duas definições coincidem. A equivalência é verdadeira apenas no caso da utilidade quase-linear, ou ela é verdadeira de maneira mais geral? Como se vê, a proposição é válida em circunstâncias bastante gerais.

Para ver isso, considere um modelo simples com dois bens, x e y, e um único consumidor. Nós escolhemos y como o bem numerário, e normalizamos seu preço para 1. O consumidor está consumindo inicialmente 0 unidades do bem x, e o monopolista deseja vender x^* unidades para a maior quantidade possível do bem y. Seja y^* a quantidade do bem y que o consumidor tem após efetuar esse pagamento; então y^* é a solução para a equação

$$u(x^*, y^*) = u(0, y), \tag{4.28}$$

e o pagamento é simplesmente $y - y^*$. Esta é claramente a maior quantidade possível do bem x que o consumidor pagaria em uma base de take-it ou leave-it para consumir x^* unidades do bem x.

Suponha, em vez disso, que o monopolista divida x^* em n partes de tamanho Δx e venda cada peça ao consumidor ao preço máximo que o consumidor estaria disposto a pagar por essa unidade. Seja (x_i, y_i) a quantidade que o consumidor tem no estágio i deste processo, de modo que $y_{i-1} - y_i$ é o valor pago pela unidade i do bem x. Como a utilidade permanece constante durante este processo, temos:

$$\begin{aligned} u(x_1, y_1) - u(0, y) &= 0, \\ u(x_2, y_2) - u(x_1, y_1) &= 0, \\ &\vdots \\ u(x^*, y_n) - u(x_{n-1}, y_{n-1}) &= 0. \end{aligned} \tag{4.29}$$

Queremos mostrar que y_n, o montante total detido do bem y depois de concluído este processo, é igual a y^*, o montante pago pela oferta take-or-leave-it descrita acima.

Mas isso é fácil. Basta somar as equações em (4.29) para encontrar

$$u(x^*, y_n) - u(0, y) = 0. \tag{4.30}$$

Examinando (4.28), vemos que $y_n = y^*$, como deveria ser mostrado.

É bem sabido que um monopolista perfeitamente discriminador produz uma quantidade Pareto eficiente de produto. Seja $u(x, y)$ a função de utilidade do consumidor, como antes, e por simplicidade suponha que o monopolista se preocupa apenas com o consumo do bem y. O monopolista é dotado de uma tecnologia que lhe permite produzir x unidades do bem x usando $c(x)$ unidades do bem y. A dotação inicial do consumidor é denotada por (x_c, y_c) e, por suposição, o monopolista tem uma dotação inicial de zero de cada bem.

O monopolista quer escolher um nível de produção (positivo) x_c e um pagamento (negativo) y do bem y que maximize sua utilidade sujeita à restrição de que o consumidor realmente compra o bem x do monopolista. Assim, o problema de maximização se torna:

$$\begin{aligned} &\max_{x,y} y - c(x) \\ \text{sujeito a} \quad &u(x_c + x, y_c - y) \geq u(x_c, y_c). \end{aligned} \tag{4.31}$$

Mas esse problema simplesmente nos pede para encontrar uma alocação viável que maximize a utilidade de uma parte, o monopolista, sujeita à restrição de que a outra parte, o consumidor, tenha algum nível de utilidade. Esta é a definição de uma alocação eficiente de Pareto. Portanto, um monopolista perfeitamente discriminador escolherá um nível de produção Pareto-eficiente.

Pelo Segundo Teorema do Bem-Estar e pelas condições apropriadas de convexidade, esse nível de produção eficiente em Pareto é um equilíbrio competitivo para algumas dotações. Para ver isso diretamente, denote a solução para o problema de maximização do monopolista por (x^*, y^*). Esta solução deve satisfazer as condições de primeira ordem:

$$1 - \lambda \frac{\partial u(x_c + x^*, y_c - y^*)}{\partial y} = 0, \tag{4.32}$$

$$-c'(x^*) + \lambda \frac{\partial u(x_c + x^*, y_c - y^*)}{\partial x} = 0. \tag{4.33}$$

Dividindo a segunda equação pela primeira e rearranjando, temos:

$$\frac{\dfrac{\partial u(x_c + x^*, y_c - y^*)}{\partial x}}{\dfrac{\partial u(x_c + x^*, y_c - y^*)}{\partial y}} = c'(x^*). \tag{4.34}$$

Se o consumidor tem uma dotação de $(x_c + x^*, y_c - y^*)$ e a firma se defronta com um preço determinado por $p^* = \dfrac{\dfrac{\partial u(x_c + x^*, y_c - y^*)}{\partial x}}{\dfrac{\partial u(x_c + x^*, y_c - y^*)}{\partial y}}$, então o problema de maximização da firma se torna:

$$\max_x p^* x - c(x). \tag{4.35}$$

Nesse caso, fica claro que a empresa escolherá otimamente produzir x^* unidades de produto, conforme necessário.

É claro que a prova de que o nível de produção de um monopolista perfeitamente discriminatório é o mesmo que o de uma firma competitiva só é válida se a realocação adequada das dotações iniciais for feita. No entanto, se estivermos dispostos a descartar os efeitos de renda, essa advertência poderá ser eliminada.

Para ver isso, vamos supor agora que a função de utilidade para o consumidor assume a forma quase-linear $u(x) + y$. Neste caso $\dfrac{\partial u}{\partial y} = 1$, então as condições de primeira ordem se reduzem a

$$\frac{\partial u(x_c + x^*)}{\partial x} = c'(x^*). \tag{4.36}$$

Isso mostra que o nível de produção eficiente de Pareto produzido pelo monopolista perfeitamente discriminativo é independente da dotação de y, que é o que exigimos. Claramente, a quantidade do bem x produzida é a mesma que a de uma empresa competitiva que enfrenta um preço dado por $p^* = \dfrac{\partial u(x_c + x^*)}{\partial x}$.

As ofertas de "pegar ou largar" não são formas muito comuns de negociação por dois motivos. Em primeiro lugar, a ameaça "leave-it" carece de credibilidade: normalmente, um vendedor não tem como se comprometer a romper as negociações se uma oferta for rejeitada. E uma vez que uma oferta inicial tenha sido rejeitada, é geralmente racional que o vendedor continue negociando.

Em segundo lugar, mesmo que o vendedor tenha uma maneira de se comprometer com o fim das negociações, ele normalmente não tem informações completas sobre as preferências dos compradores. Assim, o vendedor não pode determinar com certeza se sua oferta será realmente aceita e deve negociar os custos da rejeição com os benefícios de lucros adicionais.

Se o vendedor fosse capaz de antecipar o take-it-or-leave-it, e tivesse informações perfeitas sobre as preferências dos compradores, seria de se esperar que as transações fossem feitas de acordo com esse mecanismo. Afinal, ele oferece ao vendedor o maior lucro possível.

No entanto, vestígios da tentativa de discriminação de preços em primeiro grau ainda podem ser detectados em alguns acordos de marketing. Alguns tipos de mercadorias - desde aeronaves, por um lado, a refrigeradores e aparelhos de som, por outro - ainda são vendidos por pechinchar. Certamente isso deve ser devido a uma tentativa de discriminação de preços entre os clientes em potencial.

Na medida em que essa negociação é bem-sucedida na extração do excedente total dos consumidores, ela tende a estimular a produção de uma quantidade eficiente de produto. Uma análise completa do bem-estar das tentativas de se envolver nesse tipo de discriminação de preços não pode negligenciar os custos de transação envolvidos na própria negociação.

4.6.3.3 Discriminação de Preços de Segundo Grau

A discriminação de preço de segundo grau, ou discriminação de preço não-linear, ocorre quando os indivíduos enfrentam esquemas de preços não-lineares, ou seja, o preço pago depende da quantidade comprada. O exemplo padrão dessa forma de discriminação de preços são os descontos por quantidade.

Curiosamente, a determinação de preços ótimos não lineares não foi cuidadosamente examinada até Spence (1976). Desde então, tem havido uma série de contribuições nesta área; veja a pesquisa bibliográfica em Brown e Sibley (1986). Grande parte de seu trabalho usa técnicas originalmente desenvolvidas por Mirrlees (1971, 1976), Roberts (1979) e outros, com o propósito de analisar problemas na tributação ótima. Grande parte do trabalho descrito por Brown e Sibley é motivado pelo preço dos serviços públicos. Vamos seguir a excelente discussão em Tirole (1988), que por sua vez é baseada em Maskin e Riley (1984). No entanto, conduzimos a derivação principal usando uma estrutura de utilidade geral e recorremos ao caso especial considerado por esses autores apenas quando necessário.

É útil começar considerando uma situação em que existem apenas dois tipos de consumidores, uma fração f_1 do tipo t_1 e uma fração f_2 do tipo t_2. O monopolista quer vender x_1 para os consumidores tipo t_1 e x_2 para os consumidores tipo t_2, recebendo os pagamentos totais de r_1 e r_2 de cada tipo.

As funções de utilidade dos consumidores são da forma quase-linear $u(x_i, t_i) + y_i$ em que y_i é o consumo do bem numerário. Por conveniência, consideramos a dotação do numerário como zero. Também assumimos que $u(x, t_2) > u(x, t_1)$ e que $\dfrac{\partial u(x, t_2)}{\partial x} > \dfrac{\partial u(x, t_1)}{\partial x}$. Essas premissas implicam que não apenas o consumidor 2 está disposto a pagar mais do que o consumidor 1 por um determinado montante do bem, mas também que a disposição marginal a pagar do consumidor 2 excede a do consumidor 1. Vamos nos referir ao consumidor 2 como consumidor de alta demanda e o consumidor 1 como consumidor de baixa demanda. As suposições implicam que a função de demanda para o consumidor de alta demanda é sempre mais alta do que a função de demanda para o consumidor de baixa demanda, uma propriedade às vezes conhecida como condição de não-cruzamento.

As restrições de demanda enfrentadas pelo monopolista são as seguintes. Primeiro, cada consumidor deve querer consumir a quantidade x_i e estar disposto a pagar o preço r_i, isto é, deve satisfazer as seguintes restrições de participação (ou restrição de racionalidade individual):

$$u(x_1, t_1) - r_1 \geq 0, \tag{4.37}$$

$$u(x_2, t_2) - r_2 \geq 0. \tag{4.38}$$

Isso é simplesmente definir o domínio do problema que analisaremos. Em segundo lugar, cada consumidor deve preferir esse consumo ao consumo do outro consumidor:

$$u(x_1, t_1) - r_1 \geq u(x_2, t_1) - r_2, \tag{4.39}$$
$$u(x_2, t_2) - r_2 \geq u(x_1, t_2) - r_1. \tag{4.40}$$

Estas são as chamadas restrições de auto-seleção. Se o plano (x_1, x_2) for viável no sentido de que será voluntariamente escolhido pelos consumidores, então cada consumidor deve preferir consumir a cesta destinada a ele em comparação ao consumo da cesta da outra pessoa. Em outras palavras, essas restrições evitam a arbitragem pessoal: cada consumidor obtém um excedente tão grande quanto a escolha do pacote projetado para ele como ele/ela escolhe o pacote projetado para o outro consumidor.

Nossas suposições sobre as funções de utilidade e o fato de que o monopolista quer que os preços sejam os mais altos possíveis, implicam que duas das quatro desigualdades acima indicadas serão restrições vinculantes (binding).

Demonstração de quais restrições são satisfeitas com igualdade.

De (4.37) e (4.39), temos que:

$$r_1 \leq u(x_1, t_1) \tag{4.41}$$
$$r_1 \leq u(x_1, t_1) - u(x_2, t_1) + r_2 \tag{4.42}$$

De (4.38) e (4.40), temos que:

$$r_2 \leq u(x_2, t_2) \tag{4.43}$$
$$r_2 \leq u(x_2, t_2) - u(x_1, t_2) + r_1 \tag{4.44}$$

O monopolista deseja maximizar os lucros e, portanto, escolherá os valores mais alto possíveis de r_1 e r_2. Como consequência, apenas uma das duas primeiras desigualdades e apenas uma das duas desigualdades serão vinculativas (ou seja, elas serão satisfeitas com a igualdade). A suposição de que o consumidor 2 é o consumidor de alta demanda e o consumidor 1 é o consumidor de baixa demanda $(u(x_2, t_2) > u(x_1, t_1)\ \forall x$ e $u'(x_2, t_2) > u'(x_1, t_1)\ \forall x)$ é suficiente para determinar quais restrições são binding.

Suponha que (4.43) é satisfeita com a igualdade e, portanto, $r_2 = u(x_2, t_2)$. Então $r_2 \leq r_2 - u(x_1, t_2) + r_1 \Rightarrow r_1 \geq u(x_1, t_2)$. Dado que o consumidor 2 é o consumidor de alta demanda $u_2(x_2, t_2) > u_1(x_1, t_1)\ \forall x$. Ou seja, $r_1 \geq u(x_1, t_2) > u_1(x_1, t_1)$, o que significa que a restrição (4.41) não seria binding, o que é uma contradição. O fato de a restrição de participação do consumidor de alta demanda estar satisfeita com a igualdade não é compatível com o fato de o consumidor de baixa demanda comprar o bem. Como conclusão, (4.41) não é binding e (4.42) é satisfeita com a igualdade

$$r_2 = u(x_2, t_2) - u(x_1, t_2) + r_1 \tag{4.45}$$

Especificamente, o consumidor de baixa demanda será cobrado por sua disposição máxima a pagar, e o consumidor de alta demanda será cobrado pelo preço mais alto que apenas o induzirá a consumir x_2 em vez de x_1.

Suponha que a condição (4.42) seja satisfeita com a igualdade e, portanto, que $r_1 = u(x_1, t_1) - u(x_2, t_1) + r_2$. Ao substituir r_2 da condição (4.45), obtemos:

$$r_1 = u(x_1, t_1) - u(x_2, t_1) + u(x_2, t_2) - u(x_1, t_2) + r_1 \tag{4.46}$$

o que implica que

$$\begin{aligned} u(x_2, t_2) - u(x_1, t_2) &= u(x_2, t_1) - u(x_1, t_1) \\ \int_{x_1}^{x_2} u'(t_2) dt_2 &= \int_{x_1}^{x_2} u'(t_1) dt_1 \end{aligned} \tag{4.47}$$

Mas isso contradiz a suposição de que o consumidor 2 é o consumidor de alta demanda, $u'(t_2) > u'(t_1)\ \forall x$. Portanto, (4.42) não é binding e (4.41) é satisfeita com a igualdade.

Assim,

$$r_1 = u(x_1, t_1), \tag{4.48}$$

$$r_2 = u(x_2, t_2) - u(x_1, t_2) + u(x_1, t_1). \tag{4.49}$$

O monopolista cobra do consumidor 1 uma tarifa igual à sua disposição máxima a pagar, uma vez que o consumidor de baixa demanda não tem incentivo para se envolver em arbitragem pessoal. Dado que o consumidor de alta demanda tem incentivo para se envolver em arbitragem pessoal (e imitar o consumidor de baixa demanda), o monopolista cobra o preço máximo que o induz a escolher o pacote projetado para ele (a quantidade de dinheiro que o deixa indiferente entre a sua cesta e a que foi designada para o consumidor de baixa demanda).

O problema do monopolista era

$$\begin{aligned} &\max_{r_1, x_1, r_2, x_2} \pi = [r_1 - cx_1]f_1 + [r_2 - cx_2]f_2 \\ \text{sujeito a } & u(x_1, t_1) - r_1 \geq 0 \\ & u(x_2, t_2) - r_2 \geq 0 \\ & u(x_1, t_1) - r_1 \geq u(x_2, t_1) - r_2 \\ & u(x_2, t_2) - r_2 \geq u(x_1, t_2) - r_1 \end{aligned}$$

$$(4.50)$$

e com as novas restrições (binding) se torna:

$$\max_{x_1,x_2} \pi = [r_1 - cx_1]f_1 + [r_2 - cx_2]f_2 \tag{4.51}$$

$$= [u(x_1,t_1) - cx_1]f_1 + [u(x_2,t_2) - u(x_1,t_2) + u(x_1,t_1) - cx_2]f_2. \tag{4.52}$$

Esta expressão deve ser maximizada em relação a x_1 e x_2. Diferenciando, nós temos

$$\left[\frac{\partial u(x_1,t_1)}{\partial x_1} - c\right] f_1 + \left[\frac{\partial u(x_1,t_1)}{\partial x_1} - \frac{\partial u(x_1,t_2)}{\partial x_1}\right] f_2 = 0, \tag{4.53}$$

$$\frac{\partial u(x_2,t_2)}{\partial x_2} - c = 0. \tag{4.54}$$

em que $\left(\frac{\partial u(x_1,t_1)}{\partial x_1} - c\right)$ é o lucro marginal do consumidor 1: uma mudança na quantidade fornecida a esse consumidor implica uma mudança no lucro obtido pelo monopolista e $\left(\frac{\partial u(x_1,t_1)}{\partial x_1} - \frac{\partial u(x_1,t_2)}{\partial x_1}\right)$ é o lucro marginal do consumidor 2: uma mudança na quantidade fornecida ao consumidor 1 implica uma mudança no excedente que o monopolista deve deixar ao consumidor 2 para evitar arbitragem pessoal.

A equação (4.53) pode ser rearranjada para nos dar:

$$\frac{\partial u(x_1,t_1)}{\partial x_1} = c + \left[\frac{\partial u(x_1,t_1)}{\partial x_1} - \frac{\partial u(x_1,t_2)}{\partial x_1}\right] \frac{f_2}{f_1}, \tag{4.55}$$

o que significa que o consumidor de baixa demanda tem um valor (marginal) para o bem que excede o custo marginal. Assim, ele consome uma quantidade ineficientemente pequena do bem. A equação (4.54) diz que aos preços não lineares ideais, o consumidor de alta demanda tem uma disposição a pagar marginal que é igual ao custo marginal. Assim, ele consome a quantidade socialmente correta, ou seja, $P = CMg$ e, portanto, é a solução de concorrência perfeita.

O resultado de que o consumidor com a maior demanda enfrenta um preço marginal igual ao custo marginal é muito geral. Se o consumidor com maior demanda enfrentasse um preço marginal superior ao custo marginal, o monopolista poderia reduzir o preço marginal cobrado ao maior consumidor em uma pequena quantia, induzindo-o a comprar mais. Como o preço marginal ainda excede o custo marginal, o monopolista teria lucro com essas vendas. Além disso, tal política não afetaria os lucros do monopolista de nenhum outro consumidor, já que todos eles são otimizados com valores de consumo mais baixos. Para obter resultados mais explícitos sobre o esquema ótimo de preços, é necessário fazer suposições mais explícitas sobre os gostos. Por exemplo, é comum observar descontos de preços em certos tipos de mercadorias - consumidores de alta demanda pagam um custo por unidade menor do que os consumidores de baixa demanda. Maskin e Riley (1984) mostram que se as preferências tomam a forma específica $u(x,t) + y = tv(x) + y$, então a política de preços ótima exibirá descontos por quantidade.

Observe que podemos reescrever o lucro marginal associado ao consumidor do tipo 1:

$$\underbrace{\left[\frac{\partial u(x_1,t_1)}{\partial x_1} - c\right] f_1}_{>0} - \underbrace{\left[\frac{\partial u(x_1,t_2)}{\partial x_1} - \frac{\partial u(x_1,t_1)}{\partial x_1}\right] f_2}_{>0} \tag{4.56}$$

Se esta expressão for negativa, o monopolista teria aumento do lucro ao reduzir a quantidade vendida ao grupo do tipo 1. Dito de outra forma, o monopolista irá escolher oferecer o bem aos dois grupos de consumidores de o lucro conjunto for superior ao lucro obtido com a venda apenas para o grupo 2.

Principais Resultados

1. O monopolista fornece ao consumidor de alta demanda a quantidade eficiente e deixa um excedente positivo. De (4.48) e de (4.49), vemos que o consumidor 2, de alta demanda, ganha do monopolista um excedente dado por $u(x_2,t_2) - r_2 = u(x_1,t_2) - u(x_1,t_1)$. E como $\dfrac{\partial u(x_2,t_2)}{\partial x_2} - c = 0$, a quantidade consumida é a eficiente.
2. O monopolista oferece ao consumidor de baixa demanda uma quantidade menor que a quantidade eficiente e não o deixa com excedente.

Suponha agora que há um continuum de tipos e seja $f(t)$ a densidade de consumidores do tipo t. Por conveniência, os tipos variam de 0 a T. A função de utilidade de um consumidor do tipo t é dada por $u(x,t)+y$. Mais uma vez, assumimos que o aumento de t aumenta tanto a disposição como a remuneração total a pagar, o que, nesse contexto, significa que

$$\frac{\partial u(x(t),t)}{\partial t} > 0, \tag{4.57}$$

$$\frac{\partial^2 u(x(t),t)}{\partial t \partial x} > 0. \tag{4.58}$$

Seja $x(t)$ o consumo ótimo de um consumidor do tipo t. As restrições de auto-seleção implicam que um consumidor do tipo t prefere seu consumo a um consumidor do tipo s, o que significa

$$u(x(t),t) - r(x(t)) \geq u(x(s),t) - r(x(s)). \tag{4.59}$$

Considere a função $g(s)$ definida por:

$$g(s) = [u(x(t),t) - r(x(t))] - [u(x(s),t) - r(x(s))]. \tag{4.60}$$

Acabamos de ver que $g(s) \geq 0$ e, claro, $g(t) = 0$. Segue que $g(s)$ alcança seu valor mínimo quando $s = t$. Portanto, a derivada de g em relação a s deve ser nula em $s = t$, o que implica:

$$\frac{\partial u(x(t), t)}{\partial x} - \frac{\partial r(x(t))}{\partial x} = 0. \tag{4.61}$$

Este é o análogo da restrição de auto-seleção dada acima.

Seja $V(t)$ a utilidade maximizada de um agente do tipo t, isto é,

$$V(t) \equiv u(x(t), t) - r(x(t)). \tag{4.62}$$

Derivando $V(t)$, obtemos:

$$\begin{aligned} V'(t) &= \left(\frac{\partial u(x(t), t)}{\partial x} - \frac{\partial r(x(t))}{\partial x} \right) \frac{dx}{dt} + \frac{\partial u(x(t), t)}{\partial t} \\ &= \frac{\partial u(x(t), t)}{\partial t}. \end{aligned} \tag{4.63}$$

Este é simplesmente o teorema do envelope - a derivada total da utilidade reduz a derivada parcial após a substituição nas condições de primeira ordem para maximização. O monopolista quer escolher $x(t)$ para maximizar os lucros sujeitos às restrições de auto-seleção. Os lucros são dados por

$$\pi = \int_0^T [r(x(t)) - cx(t)] f(t) dt. \tag{4.64}$$

O truque é construir as restrições de auto-seleção na função objetivo de uma maneira útil. Usando (4.62) podemos reescrever os lucros como

$$\pi = \int_0^T [u(x(t), t) - cx(t)] f(t) dt - \int_0^T V(t) f(t) dt. \tag{4.65}$$

Integrando o último termo por partes, temos

$$\int_0^T V(t) f(t) dt = V(t)(F(t) - 1)|_0^T - \int_0^T V'(t)[F(t) - 1] dt. \tag{4.66}$$

Aqui nós usamos $F(t) - 1$ como a integral de $f(t)$. A utilidade do tipo 0 é normalizada para ser 0, e $F(T) = 1$. Portanto, o primeiro termo do lado direito dessa expressão desaparece. Logo,

$$\int_0^t V(t)f(t)dt = -\int_0^T \frac{\partial u(x(t),t)}{\partial t}[F(t)-1]dt. \tag{4.67}$$

Substituindo isso de volta na função objetivo, a equação (4.65) nos dá a forma final da função lucro:

$$\pi = \int_0^T \left\{[u(x(t),t) - cx(t)]f(t) - \frac{\partial u(x(t),t)}{\partial t}[1-F(t)]\right\} dt. \tag{4.68}$$

Ao longo do caminho ótimo, a derivada do integrando com respeito a cada $x(t)$ deve desaparecer. Isso nos dá a condição de primeira ordem:

$$\left[\frac{\partial u(x(t),t)}{\partial x} - c\right] f(t) - \frac{\partial^2 u(x(t),t)}{\partial t \partial x}[1-F(t)] = 0. \tag{4.69}$$

Resolvendo para $\dfrac{\partial u(x(t),t)}{\partial x}$, obtemos:

$$\frac{\partial u(x(t),t)}{\partial x} = c + \frac{\partial^2 u(x(t),t)}{\partial t \partial x}\left(\frac{1-F(t)}{f(t)}\right). \tag{4.70}$$

Como no caso de dois consumidores, todos os consumidores pagam um preço superior ao custo marginal, exceto para o consumidor com a maior disposição a pagar, o consumidor T.

Exemplo 4.6.1. *Suponha que um monopolista se defronte com N consumidores de alta renda e n consumidores de baixa renda com as respectivas funções de demanda:*

$$P_H = A - Q \tag{4.71}$$

$$P_L = a - Q \tag{4.72}$$

com $A > a > 0$. E o custo é igual a $C = cQ$ com $0 < c < a$. Denote por $W_H(Q)$ a disposição a pagar dos consumidores de alta renda e $W_L(Q)$ a disposição a pagar dos consumidores de baixa renda.

Podemos calcular essas disposições como:

$$W_H(Q) = \int_0^Q P_H(x)dx = \int_0^Q (A-x)dx = Ax - \frac{x^2}{2}\bigg|_0^Q = AQ - \frac{Q^2}{2} \tag{4.73}$$

$$W_L(Q) = \int_0^Q P_L(x)dx = \int_0^Q (a-x)dx = Ax - \frac{x^2}{2}\bigg|_0^Q = aQ - \frac{Q^2}{2} \tag{4.74}$$

O monopolista decide vender pacotes (Q, V) consistindo de Q unidades ao preço P. Ele tem três opções:

1. *Oferecer somente um pacote para os consumidores de alta renda, uma vez que $V > W_L(Q)$ e $V \leq W_H(Q)$. A função objetivo do monopolista será*

$$\begin{aligned} \pi_1 &= N[W_H(Q) - cQ] \\ &= N\left[AQ - \frac{Q^2}{2} - cQ\right] \end{aligned} \tag{4.75}$$

A condição de primeira ordem implica que

$$\frac{\partial \pi_1}{\partial Q} = 0 \quad \Longleftrightarrow \quad N[A - c - Q] = 0 \Rightarrow Q_1^* = A - c \tag{4.76}$$

Assim

$$\begin{aligned} V_1^* = W_H(Q_1^*) &= AQ_1^* - \frac{(Q_1^*)^2}{2} \\ &= Q_1^*\left(A - \frac{A-c}{2}\right) \\ &= \frac{A^2 - c^2}{2} \end{aligned} \tag{4.77}$$

Logo, o lucro será:

$$\begin{aligned} \pi_1^* &= N\left[Q_1^*(A-c) - \frac{(Q_1^*)^2}{2}\right] \\ &= N\left[\frac{(A-c)^2}{2}\right] \end{aligned} \tag{4.78}$$

2. *Oferecer somente um pacote para os dois grupos. Nesse caso, $V = W_L(Q)$. A função objetivo do monopolista será*

$$\begin{aligned} \pi_2 &= (N+n)[W_L(Q) - cQ] \\ &= (N+n)\left[aQ - \frac{Q^2}{2} - cQ\right] \end{aligned} \tag{4.79}$$

A condição de primeira ordem implica que

$$\frac{\partial \pi_2}{\partial Q} = 0 \iff (N+n)[a-c-Q] = 0 \Rightarrow Q_2^* = a-c \tag{4.80}$$

Assim

$$\begin{aligned} V_2^* = W_L(Q_2^*) &= aQ_2^* - \frac{(Q_2^*)^2}{2} \\ &= Q_2^*\left(a - \frac{a-c}{2}\right) \\ &= \frac{a^2-c^2}{2} \end{aligned} \tag{4.81}$$

Logo, o lucro será:

$$\begin{aligned} \pi_2^* &= (N+n)\left[Q_2^*(a-c) - \frac{(Q_2^*)^2}{2}\right] \\ &= (N+n)\left[\frac{(a-c)^2}{2}\right] \end{aligned} \tag{4.82}$$

3. *Oferecer dois pacotes:* (Q_H, V_H) *para os consumidores de alta renda e* (Q_L, V_L) *para os consumidores de baixa renda.*

   Nesse caso, temos que as seguintes condições devem ser satisfeitas:

   (a) $V_L \leq W_L(Q_L)$*: os consumidores de baixa renda estão dispostos a comprar o "seu" pacote.*

   (b) $W_L(Q_L) - V_L \geq W_L(Q_H) - V_H$*: é o incentivo de compatibilidade, ou seja, os consumidores de baixa renda não preferem o pacote dos consumidores de alta renda.*

   (c) $V_H \leq W_H(Q_H)$*: os consumidores de alta renda estão dispostos a comprar o "seu" pacote.*

   (d) $W_H(Q_H) - V_H \geq W_H(Q_L) - V_L$*: é o incentivo de compatibilidade, ou seja, os consumidores de alta renda não preferem o pacote dos consumidores de baixa renda.*

   Segue de (a) e de (d) que:

$$\begin{aligned} V_L - W_L(Q_L) &\leq 0 \\ W_L(Q_L) - V_L &\geq 0 \\ W_H(Q_L) - V_L &\geq 0 \quad [\textit{usando o fato de que } W_H(Q_L) > W_L(Q_L)] \\ W_H(Q_H) - V_H &\geq W_H(Q_L) - V_L \\ W_H(Q_H) - V_H &> 0 \\ W_H(Q_H) &> V_H \end{aligned} \tag{4.83}$$

Portanto, a partir do incentivo de compatibilidade para os consumidores de alta renda eles terão um excedente positivo.

Também segue das condições (a) e (d) que:

$$V_L = W_L(Q_L) \tag{4.84}$$

e

$$\begin{aligned} V_H &= W_H(Q_H) - W_H(Q_L) + W_L(Q_L) \\ &= AQ_H - \frac{Q_H^2}{2} - AQ_L + \frac{Q_L^2}{2} + aQ_L - \frac{Q_L^2}{2} \\ &= AQ_H - \frac{Q_H^2}{2} - (A-a)Q_L \end{aligned} \tag{4.85}$$

Com isso, a função lucro se torna:

$$\begin{aligned} \pi_3 &= N(V_H - cQ_H) + n(V_L - cQ_L) \\ &= N\left[AQ_H - \frac{Q_H^2}{2} - (A-a)Q_L - cQ_H\right] + n\left[aQ_L - \frac{Q_L^2}{2} - cQ_L\right] \end{aligned} \tag{4.86}$$

O objetivo do monopolista é

$$\max_{Q_H, Q_L} N\left[AQ_H - \frac{Q_H^2}{2} - (A-a)Q_L - cQ_H\right] + n\left[aQ_L - \frac{Q_L^2}{2} - cQ_L\right] \tag{4.87}$$

As condições de primeira ordem são:

$$\frac{\partial \pi_3}{\partial Q_H} = 0 \quad \Longleftrightarrow \quad N[A - Q_H - c] = 0 \quad \Longrightarrow Q_H^* = A - c \tag{4.88}$$

$$\frac{\partial \pi_3}{\partial Q_L} = 0 \quad \Longleftrightarrow \quad N[-(A-a)] + n[a - Q_L - c] = 0 \quad \Longrightarrow Q_L^* = a - c - \frac{N}{n}(A-a) \tag{4.89}$$

Note que $A > a$, o que implica que $Q_H^* > Q_L^*$. Note que esta solução é aceitável se $Q_L^* > 0$, o que é verdade se e somente se $n(a-c) > N(A-a)$.

Agora, de (a) tínhamos que $W_L(Q_L) = V_L$. Substituindo em (b), encontramos que

$$W_L(Q_H^*) - V_H^* = -(A-a)(Q_H^* - Q_L^*) < 0 \tag{4.90}$$

Assim,

$$\begin{aligned}
\pi_3^* &= N\left[AQ_H^* - \frac{(Q_H^*)^2}{2} - (A-a)Q_L^* - cQ_H^*\right]^+ n\left[aQ_L^* - \frac{(Q_L^*)^2}{2} - cQ_L^*\right] \\
&= N(A-c)\left(A-c-\frac{A-c}{2}\right) + \left(a-c-\frac{N}{n}(A-a)\right)\left(-N(A-a)+na-nc-\frac{n}{2}\left(a-c-\frac{N}{n}(A-a)\right)\right) \\
&= \frac{N}{2}(A-c)^2 + \frac{n}{2}\left(a-c-\frac{N}{n}(A-a)\right)^2 \\
&= \pi_2^* + \frac{(N+n)N(A-a)^2}{2n}
\end{aligned} \tag{4.91}$$

Isso implica que a opção 3 é a escolhida.

Agora, devemos comparar a opção 1 e a opção 3. A opção 3 será escolhida se $\pi_3^* > \pi_1^*$*. Isto é,*

$$\begin{aligned}
\pi_3^* - \pi_1^* &> 0 \\
\frac{n}{2}\left(a-c-\frac{N}{n}(A-a)^2\right) &> 0 \\
\frac{1}{2n}\left[n(a-c)-N(A-a)\right]^2 &> 0
\end{aligned} \tag{4.92}$$

Se $n(a-c) < N(A-a)$*, o monopolista vende somente para os consumidores de alta renda. Se* $n(a-c) > N(A-a)$*, o monopolista vende para os dois grupos e ada um paga um preço diferente.*

4.6.3.4 Discriminação de Preços de Terceiro Grau

Discriminação de preço de terceiro grau ocorre quando dos consumidores são cobrados preços diferentes, mas cada consumidor enfrenta um preço constante para todas as unidades de produção compradas. Esta é provavelmente a forma mais comum de discriminação de preços. Um exemplo seria discriminação por idade, como descontos para idosos no cinema. Se deixarmos $p(x_i)$ ser a função de demanda inversa para o grupo i e supor que existem dois grupos, então o problema de maximização do lucro do monopolista é

$$\max_{x_1,x_2} p_1(x_1)x_1 + p_2(x_2)x_2 - cx_1 - cx_2. \tag{4.93}$$

As condições de primeira ordem são:

$$p_1(x_1) + p_1'(x_1)x_1 = c, \tag{4.94}$$
$$p_2(x_2) + p_2'(x_2)x_2 = c. \tag{4.95}$$

Seja ε_i a elasticidade da demanda no mercado i. Podemos escrever essas expressões como

$$p_1(x_1)\left[1 - \frac{1}{|\varepsilon_1|}\right] = c, \tag{4.96}$$
$$p_2(x_2)\left[1 - \frac{1}{|\varepsilon_2|}\right] = c. \tag{4.97}$$

Segue-se que $p_1(x_1) > p_2(x_2)$ se e somente se $|\varepsilon_1| < |\varepsilon_2|$. Assim, o mercado com demanda mais elástica - o mercado mais sensível ao preço - é cobrado pelo preço mais baixo.

Suponha agora que o monopolista é incapaz de separar os mercados tão claramente quanto se supõe, de modo que o preço cobrado em um mercado influencie a demanda em outro mercado. Por exemplo, considere um cinema que tenha uma noite de barganha na segunda-feira; o preço mais baixo na segunda-feira, presumivelmente, influenciaria a demanda na terça-feira em algum grau.

Então o problema de maximização do lucro do monopolista seria:

$$\max_{x_1,x_2} p_1(x_1,x_2)x_1 + p_2(x_1,x_2)x_2 - cx_1 - cx_2, \tag{4.98}$$

e as condições de primeira ordem se tornariam:

$$p_1 + \frac{\partial p_1}{\partial x_1}x_1 + \frac{\partial p_2}{\partial x_1}x_2 = c, \tag{4.99}$$
$$p_2 + \frac{\partial p_2}{\partial x_2}x_2 + \frac{\partial p_1}{\partial x_2}x_1 = c. \tag{4.100}$$

Podemos rearranjar estas equações para encontrarmos:

$$p_1\left[1 - \frac{1}{|\varepsilon_1|}\right] + \frac{\partial p_2}{\partial x_1}x_2 = c, \tag{4.101}$$
$$p_2\left[1 - \frac{1}{|\varepsilon_2|}\right] + \frac{\partial p_1}{\partial x_2}x_1 = c. \tag{4.102}$$

Não é fácil dizer algo muito interessante sobre essas equações, mas vamos tentar. Uma simplificação que podemos fazer é assumir que não há efeitos renda, de modo que $\dfrac{\partial p_1}{\partial x_2} = \dfrac{\partial p_2}{\partial x_1}$, ou seja, os efeitos cruzados de preço são simétricos[5]. Subtraindo a segunda equação do primeiro e

[5] Assumir que não há efeito-renda é uma suposição mais forte do que precisamos. Willig (1976) e Varian (1978) mostraram que tudo o que é necessário é que as elasticidades de renda dos bens 1 e 2 sejam localmente constantes em algumas regiões no espaço preço-renda.

rearranjando, temos:

$$p_1\left[1-\frac{1}{|\varepsilon_1|}\right]-p_2\left[1-\frac{1}{|\varepsilon_2|}\right]=(x_1-x_2)\frac{\partial p_2}{\partial x_1}. \tag{4.103}$$

É natural supor que os dois bens são substitutos - afinal, eles são o mesmo bem sendo vendido a grupos diferentes - de modo que $\frac{\partial p_2}{\partial x_1}>0$. Sem perda de generalidade, assuma que $x_1>x_2$, que, pela equação imediatamente acima, implica que:

$$p_1\left[1-\frac{1}{|\varepsilon_1|}\right]-p_2\left[1-\frac{1}{|\varepsilon_2|}\right]>0. \tag{4.104}$$

Rearranjando, temos:

$$\frac{p_1}{p_2}>\frac{1-\frac{1}{|\varepsilon_2|}}{1-\frac{1}{|\varepsilon_1|}}. \tag{4.105}$$

Segue-se desta expressão que se $|\varepsilon_2|>|\varepsilon_1|$, devemos ter $p_1>p_2$. Ou seja, se o mercado menor tem demanda mais elástica, deve ter o preço mais baixo. Assim, a intuição dos mercados separados é transferida para o caso mais geral sob essas suposições adicionais.

4.6.4 Aplicações de Discriminação de Preços

Como vimos nas observações introdutórias, a discriminação de preços é uma tática de marketing muito comumente usada. Vamos discutir várias maneiras que as empresas podem usar para discriminar preços entre seus clientes.

4.6.4.1 Discriminação de Preços Espacial

As empresas geralmente usam taxas de entrega - ou a ausência de taxas de entrega - para discriminar preços entre os clientes. Embora os custos de entrega desses bens possam ser uma fração significativa de seu valor, as empresas podem achar lucrativo cobrar um preço único.

A fim de modelar o efeito dos custos de transporte na política de preços do monopolista, seja x a quantidade vendida em um determinado local, seja a função de demanda inversa para o bem denotada por $p(x)$ e seja o custo tx. A função de demanda líquida que a empresa enfrenta é então $p(x)-t$. Assumindo custos marginais constantes, o problema de maximização do lucro para a firma se torna:

$$\max_x [p(x)-t-c]x, \tag{4.106}$$

cuja condição de primeira ordem é:

$$p(x) + p'(x)x = c + t. \tag{4.107}$$

Como o preço líquido para o consumidor mudará quando os custos de transporte mudarem? A formulação acima deixa claro que isso é o mesmo que perguntar como o preço para os consumidores muda à medida que mudamos um imposto de consumo em um caso monopolista.

Diferenciando implicitamente (4.107) com relação a t e resolvendo para $\frac{dx}{dt}$, temos:

$$\frac{dx}{dt} = \frac{1}{2p'(x) + p''(x)x}. \tag{4.108}$$

Portanto,

$$\frac{dp}{dt} = \frac{dp}{dx}\frac{dx}{dt} = \frac{1}{2 + \frac{p''(x)x}{p'(x)}}. \tag{4.109}$$

Assim, o montante dos custos de transporte que são repassados depende da segunda derivada da função de demanda inversa, $p''(x)$. Podemos examinar alguns casos especiais. Se a curva de demanda inversa for linear, $p''(x) = 0$, metade dos custos de transporte serão pagos pelos consumidores. Nesse caso, o monopolista pratica a absorção de frete e discrimina clientes com menores custos de transporte. Isso é semelhante a os exemplos de cimento e placas de gesso descritos acima: a empresa efetivamente absorve parte ou todos os custos de transporte.

Se a curva de demanda inversa tiver uma elasticidade constante de ε, então o monopolista cobra uma margem constante sobre o custo marginal, de modo que os clientes em cada distância efetivamente paguem mais pela entrega do que os custos reais de entrega. No entanto, essa forma de discriminação de preços é especialmente sensível à arbitragem do consumidor - os consumidores com custos de transporte mais baixos podem transbordar para aqueles com custos de transporte mais altos, prejudicando essencialmente a posição do monopolista.

Os sistemas de preços de entrega às vezes foram atacados como anticoncorrenciais sob a Lei Robinson-Patman; para uma discussão de algumas das questões envolvidas e citações de vários casos relevantes, veja Neale e Goyder (1980, pp.245-248).

Também temos o PLS 49/2015. Ementa: Institui a Política Nacional do Livro e regulação de preços. Explicação da Ementa: Institui a política nacional de fixação do preço do livro, estabelecendo regras para a comercialização e difusão do livro, e definindo infrações, penalidades de multa pecuniária e ação judicial. Todo livro receberá da editora precificação única por prazo de 1 ano, a partir de seu lançamento ou importação. Constituem infrações praticar tratamento não isonômico aos comerciantes intermediários e a oferta de livros a preços inferiores ao estabelecido.

4.6.4.2 Discrimanção Intertemporal

Muitas vezes novos produtos são introduzidos a um preço alto, que depois declina. Por exemplo, os livros são normalmente introduzidos em edições caras de capa dura e só mais tarde publicados como livros de bolso menos caros. Tal política parece ser uma forma de discriminação intertemporal de preços: o monopolista tenta primeiro extrair o excedente dos consumidores de alta demanda e só depois vende para os consumidores de baixa demanda.

Este tipo de discriminação de preços intertemporais foi analisado em detalhe pela primeira vez por Stokey (1979)[6]. Ela considerou um modelo com um continum de consumidores e tempos em que os consumidores e a empresa têm a mesma taxa de desconto. No contexto desse modelo, Stokey provou ser um resultado surpreendente: a política de maximização do lucro da empresa é cobrar um preço uniforme e não se envolver em discriminação de preços.

Investigaremos o resultado de Stokey no contexto de um modelo muito mais simples. Suponha que haja dois consumidores com preços de reserva r_1 e r_2 e um fator de desconto comum de $0 < \alpha < 1$. Sem perda de generalidade, suponha que $r_1 > r_2$. O monopolista, que também tem uma taxa de desconto de α, pode fornecer gratuitamente até duas unidades do bem.

Consideramos agora se é ótimo discriminar preços neste modelo. Suponha que o monopolista defina os preços p_1 e p_2 que conseguem induzir o consumidor 1 a consumir no primeiro período e o consumidor 2 a consumir no segundo período. Em seguida, cada consumidor deve preferir o valor líquido descontado de sua compra ao do outro consumidor; isso nos dá as restrições de auto-seleção:

$$r_1 - p_1 \geq \alpha(r_1 - p_2), \tag{4.110}$$

$$\alpha(r_2 - p_2) \geq r_2 - p_1. \tag{4.111}$$

A primeira desigualdade diz que o comprador do período 1 prefere sua escolha de comprar no período 2, enquanto a segunda desigualdade estabelece a condição análoga para o consumidor 2. Se multiplicarmos essas duas desigualdades juntas e cancelarmos α de cada lado, temos:

$$\begin{aligned}(r_1 - p_1)(r_2 - p_2) &\geq (r_1 - p_2)(r_2 - p_1)\\ (p_2 - p_1)(r_2 - r_1) &\geq 0. \end{aligned} \tag{4.112}$$

Como $r_1 > r_2$ por suposição, devemos ter $p_1 > p_2$, o que significa simplesmente que o preço no primeiro período deve ser maior que o preço no segundo, um resultado dificilmente surpreendente.

O problema enfrentado pelo monopolista discriminador de preços pode agora ser escrito como

$$\max_{p_1, p_2} p_1 + \alpha p_2, \tag{4.113}$$

$$\text{sujeito a } r_1 - p_1 \geq \alpha(r_1 - p_2), \tag{4.114}$$

6 Stokey, N. (1979) "Intertemporal price discrimination", Quarterly Journal of Economics, 93:355–371.

$$\alpha(r_2 - p_2) \geq r_2 - p_1, \tag{4.115}$$
$$r_1 \geq p_1, \tag{4.116}$$
$$r_2 \geq p_2. \tag{4.117}$$

Embora existam quatro restrições neste programa linear, elas não são todas independentes. Reorganizando (4.114) temos:

$$r_1 \geq \frac{1}{1-\alpha}(p_1 - \alpha p_2). \tag{4.118}$$

Substituindo p_1 de cada lado, obtemos:

$$\begin{aligned} r_1 - p_1 &\geq \frac{1}{1-\alpha}(p_1 - \alpha p_2) - p_1 \\ &= \frac{\alpha}{1-\alpha}(p_1 - p_2) \geq 0. \end{aligned} \tag{4.119}$$

Segue que (4.114) implica (4.116), então nós descartamos (4.116) do conjunto de restrições.

Na solução para este problema de programação linear, pelo menos duas das restrições serão ativas. Eliminando os casos triviais, isso nos deixa com três possibilidades.

1. $r_1 - p_1 = \alpha(r_1 - p_2)$ e $\alpha(r_2 - p_2) = r_2 - p_1$. Manipulando essas desigualdades nos dão $(1-\alpha)r_1 = p_1 - \alpha p_2 = (1-\alpha)r_2$, uma contradição.

2. $r_2 = p_2$ e $\alpha(r_2 - p_2) = r_2 - p_1$. Estas duas equações implicam que $p_1 = p_2$, então a precificação uniforme é ótima.

3. $r_2 = p_2$ e $r_1 - p_1 = \alpha(r_1 - p_2)$. Resolvendo para p_1 e substituindo na função lucro, mostra-se que o lucro da firma é dado por:

$$\pi = r_1 + \alpha(2r_2 - r_1). \tag{4.120}$$

Se estes são os lucros máximos que a empresa pode fazer, eles devem dominar a política de vender apenas no período 1 a um preço de r_1 e apenas satisfazer os clientes de alta demanda ou vender em r_2 e satisfazer ambos os clientes. Por isso, devemos satisfazer as duas seguintes desigualdades:

$$r_1 + \alpha(2r_2 - r_1) \geq r_1, \tag{4.121}$$
$$r_1 + \alpha(2r_2 - r_1) \geq 2r_2. \tag{4.122}$$

No entanto, é facilmente verificado que essas desigualdades implicam $r_1 = 2r_2$. Substituindo de volta para a função lucro (4.120), encontramos que $\pi = r_1 = 2r_2$, de modo que os lucros da discriminação de preços sejam iguais aos de uma precificação uniforme.

Em resumo, a política ótima da empresa envolve cobrar um preço constante em cada período ou vender a um preço de r_1 ou r_2 no primeiro período e não vender de todo no segundo período. Nenhuma alternativa envolve discriminação intertemporal de preços.

Salant (1987) pergunta por que essa solução extrema surge no caso da discriminação intertemporal de preços, mas não no caso geral da precificação não linear. Ele ressalta que a análise geral de precificação não linear tipicamente assume funções e restrições objetivas devidamente curvadas e se limita a examinar apenas soluções interiores. No caso intertemporal, a linearidade da função objetivo e das restrições é uma suposição muito natural, e não devemos nos surpreender que as soluções de contorno possam ser ótimas.

No entanto, o resultado que a discriminação intertemporal de preços não maximiza o lucro é preocupante, uma vez que as empresas parecem se envolver em tal comportamento. Existem várias maneiras de relaxar as suposições do modelo para permitir a discriminação intertemporal de preços. Por exemplo, se as taxas de desconto diferirem entre os consumidores, a discriminação de preços pode ser facilmente otimizada. Da mesma forma, se as taxas de desconto dos consumidores forem diferentes das do monopolista, a discriminação de preços pode ser ótima.

É fácil ver quando isso pode ocorrer em nosso modelo. O único caso envolvendo a função de lucro da firma é o caso 3. Sendo β o fator de desconto do monopolista, uma condição suficiente para a discriminação de preço ser ótima é que os lucros descontados da discriminação de preço dominam a venda apenas no primeiro período:

$$(1-\alpha)r_1 + (\alpha+\beta)r_2 \leq \max\{r_1, 2r_2\}. \tag{4.123}$$

Segue-se que a discriminação intertemporal de preços só pode ser ótima quando $\beta > \alpha$, ou seja, quando o monopolista é menos impaciente que os consumidores. Landsberger e Meilijson (1985) examinam a rentabilidade da discriminação de preços quando as taxas de desconto diferem e derivam uma condição similar usando a formulação original de tempo contínuo de Stokey.

Também assumimos que o bem foi produzido sem custo. É trivial generalizar nosso argumento para o caso de custos marginais constantes, mas se os custos marginais estão aumentando, o argumento pode falhar. No caso de aumentar os custos marginais, pode ser lucrativo distribuir as vendas ao longo do tempo, de modo a manter os custos de produção baixos.

Uma suposição implícita em nossa análise é que a empresa pode se comprometer com credibilidade a cobrar um preço constante a cada período. Para ver que isso pode ser um problema, consideremos o caso em que é lucrativo atender apenas aos consumidores do primeiro período. Nesse caso, os preços são definidos em $p_1 = p_2 = r_1$, mas nenhum consumidor compra no segundo período.

No entanto, quando o monopolista satisfaz as necessidades da alta demanda (consumidores do primeiro período) eles ficam com apenas os consumidores de baixa demanda. Neste subjogo, a política ótima do monopolista é cobrar dos consumidores de baixa demanda seu preço de reserva

r_2 no segundo período. Mas os consumidores do primeiro período devem ser capazes de perceber que o monopolista cobrará esse preço mais baixo do segundo período - e, portanto, se recusam a comprar no primeiro período.

O problema é que a solução de cobrar um preço constante a cada período não é perfeita para o subjogo - o comportamento do monopolista não é o ideal para cada subjogo em que ele possa se encontrar. Sem a capacidade de se comprometer com o esquema de preços constante, o monopolista pode ser incapaz de impor a solução de preço constante, sem discriminação.

Em nosso modelo, o único equilíbrio perfeito no subjogo é cobrar $p_1 = r_2$ e vender para ambos os grupos de consumidores no primeiro período, independentemente do tamanho dos dois grupos de consumidores. Qualquer apólice em que a firma vende apenas para os consumidores de alta demanda não é crível, pois a empresa sempre será tentada a vender a um preço menor mais tarde. Observe que esse resultado segue, não importa quão pequeno seja o r_2, ou não importa quantos períodos ou grupos de consumidores estejam envolvidos. A incapacidade de firmar um compromisso praticamente destruiu o poder de monopólio da empresa.

4.6.4.3 Discrimanção de Preços e Integração Vertical

Para entender como a integração vertical pode ajudar a impor a discriminação de preços, consideremos um modelo em que um produtor de um produto primário, como o alumínio, vende para duas indústrias competitivas que produzem produtos finais distintos. Para assumir um caso extremo, suponha que cada um dos produtores de bens finais use uma unidade de insumo para produzir uma unidade de seu produto. Seja $p_1(x_1)$ e $p_2(x_2)$ as funções de demanda inversa pelos produtos das duas indústrias. Para simplificar, suponha que essas funções de demanda tenham elasticidades constantes de ε_1 e ε_2, respectivamente.

Efetivamente, o produtor primário controla a produção em cada setor e definirá o preço do produto primário para maximizar seus lucros. Isso leva à solução convencional:

$$p_1^* \left[1 - \frac{1}{\varepsilon_1}\right] = c, \tag{4.124}$$

$$p_2^* \left[1 - \frac{1}{\varepsilon_2}\right] = c. \tag{4.125}$$

A solução só é viável se o monopolista for capaz de impedir a arbitragem. Se a indústria que recebe o preço mais baixo puder revender para o setor de preços altos, a política de discriminação de preços do monopolista não poderá ser implementada. No entanto, existe uma estratégia que pode ser utilizada para contornar tal situação. Suponha que a indústria 2 tenha maior elasticidade de demanda e, portanto, menor preço. O monopolista pode operar uma empresa na indústria 2, vendendo sua produção ao preço p_2^* e vendendo o restante do produto primário a um preço uniforme mais alto de p_1^*.

Neste caso, as empresas da indústria 1 ainda estarão dispostas a pagar o preço mais alto, mas as empresas na indústria 2 serão pressionadas pelo preço de transferência do monopolista para sua subsidiária. Se nada for feito, as empresas do setor 2 serão expulsas dos negócios e o monopolista poderá impor o resultado da discriminação de preços através do mecanismo de integração vertical. Houve várias tentativas de resolver o problema de como detectar preços de transferência subsidiados desse tipo. Ver Perry (1978, 1980) e Joskow (1985) para uma discussão mais detalhada da discriminação de preços como um motivo para a integração vertical.

4.6.4.4 Discrimanção de Preços e Informação Imperfeita

Existem intuições conflitantes sobre o efeito da informação imperfeita em mercados monopolizados. Por um lado, as atividades de busca pelos consumidores representam um preço pago pelos consumidores que não é capturado pela empresa, portanto, é do interesse da empresa "internalizar" esse custo, eliminando a dispersão de preços. Por outro lado, se os consumidores diferem em seus custos de pesquisa, a dispersão de preços pode ser um meio eficaz de separar os consumidores e dividir o mercado, permitindo assim a discriminação de preços em equilíbrio.

Classificar os consumidores com base em seu custo de busca é especialmente conveniente, já que é natural supor que os consumidores que estão bem informados sobre os preços oferecidos em outros lugares tenham demandas mais elásticas do que os consumidores que estão mal informados. Ou, de forma mais geral, os consumidores com baixos custos de pesquisa terão demandas mais elásticas do que os consumidores com altos custos de pesquisa. Essa observação sugere que pode ser lucrativo para as lojas usar "preços confusos" como um dispositivo de seleção para discriminar os consumidores.

Esse fenômeno foi examinado pela primeira vez por Salop (1977) e posteriormente estendido por Berninghaus e Ramser (1980) e Wiesmeth (1982). No modelo de Salop, existe um único monopolista com vários produtos. Os consumidores sabem a distribuição dos preços cobrados nos vários pontos de venda, mas não sabem exatamente quais lojas cobram quais preços. Assim, eles se envolvem em pesquisas dispendiosas antes de comprar o bem. Os consumidores têm diferentes custos de pesquisa e preços de reserva diferentes para o bem. Salop (1977) e Wiesmeth (1982) mostram que, sob certas suposições sobre a distribuição conjunta de custos de busca e preços de reserva, o monopolista pode usar a dispersão de preços para classificar os consumidores de uma maneira que aumente seus lucros.

A classificação por custos de informação é uma idéia importante, mas o caso particular examinado por Salop não parece muito plausível. O empirismo causal sugere que as cadeias de lojas, como o McDonald's, normalmente tentam cobrar preços uniformes em seus pontos de venda, em vez de randomizar seus preços. Parece que a primeira intuição - a de minimizar os custos de busca do consumidor - é uma consideração mais importante para as cadeias de lojas e franquias do que a discriminação de preços.

No entanto, em vez de dispersão de preços no espaço, podemos considerar a discriminação de preços ao longo do tempo. Suponhamos que consideremos um modelo em que existem várias empresas imperfeitamente competitivas, cada uma com um produto, que cobra aleatoriamente preços diferentes em semanas diferentes. Ao randomizar seus preços, as lojas são capazes de competir pelos consumidores sensíveis a preços quando seus preços são baixos, mas ainda cobram preços altos para consumidores insensíveis a preços em média. Pode-se interpretar esse comportamento como lojas envolvidas em vendas aleatórias. Varian (1981) construiu um modelo formal desse comportamento para determinar o padrão de equilíbrio de vendas. Aqui nós brevemente consideramos o modelo de Varian.

Suponha que existam n lojas vendendo um produto idêntico que tenham curvas de custo médio idênticas, estritamente decrescentes. Cada loja escolhe a frequência $f(p)$ com a qual anuncia cada preço. Alguma fração dos consumidores lê os anúncios e aprende toda a distribuição de preços; eles, portanto, só fazem compras na loja de preço mais baixo. O resto dos consumidores faz compras aleatoriamente. Cada consumidor compra no máximo uma unidade do bem.

Procuramos caracterizar um equilíbrio de Nash simétrico neste modelo. A primeira observação é que, uma vez que os custos médios estão sempre em declínio, não pode haver um equilíbrio estratégico puro em que todas as empresas cobrem um único preço. O único equilíbrio possível envolve, portanto, uma estratégia mista.

Acontece que é possível mostrar que a distribuição da frequência de equilíbrio deve ser tal que não pode haver preços que sejam carregados com probabilidade estritamente positiva. A intuição não é difícil: suponha que houvesse tal situação - algum preço que todas as lojas carregassem de probabilidade positiva. Então haveria uma probabilidade positiva de empate a um preço e então algumas lojas dividiriam os consumidores informados. Ao escolher uma distribuição de frequência que cobrava um preço ligeiramente mais baixo com probabilidade positiva, uma loja capturaria todo o mercado de consumidores informados. Portanto, cobrar um preço com probabilidade positiva não pode ser um lucro maximizador que configure um equilíbrio de Nash simétrico.

Dada essa observação, não é difícil calcular os lucros esperados de uma empresa. Se $F(p)$ é a função de distribuição de preços de equilíbrio, então exatamente dois eventos são relevantes. Ou a empresa em questão está cobrando o preço mais baixo, um evento que acontece com a probabilidade $(1-F(p))^{n-1}$, ou não tem o preço mais baixo, um evento que tem a probabilidade complementar. Se ela cobra o preço mais baixo, obtém $I+U$ clientes, em que I é o número total de clientes informados e U é o número de clientes desinformados por loja. Suponha, por simplicidade, que a empresa tenha custos marginais constantes de zero e custos fixos de k. Então o lucro de uma empresa representativa é

$$\pi = \int_0^\infty \left\{[1-F(p)]^{n-1}[pI+pU-k] + [1-(1-F(p))^{n-1}](pU-k)\right\} f(p)dp. \tag{4.126}$$

Se a loja está escolhendo a função de densidade ótima $f(p)$, então o integrando deve ser constante. Assumindo entrada livre, esse nível constante de lucros deve ser zero. Isso nos dá a condição de equilíbrio:

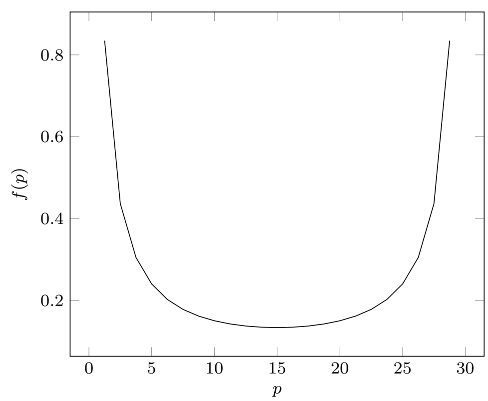
$$\left\{[1-F(p)]^{n-1}[pI+pU-k] + [1-(1-F(p))^{n-1}](pU-k)\right\} = 0. \tag{4.127}$$

Resolvendo para $F(p)$, temos:

$$F(p) = 1 - \left(\frac{k - pU}{pI}\right)^{\frac{1}{n-1}}. \tag{4.128}$$

Essa função de distribuição é o único equilíbrio de Nash simétrico nesse modelo. A função de densidade associada, $f(p)$, é simplesmente a derivada dessa função de distribuição. Em vez de ter a forma de sino normal que esperamos de uma distribuição de probabilidade, a densidade de equilíbrio $f(p)$ tem uma forma de U - ou seja, cada loja cobra preços altos e baixos com mais frequência do que os preços intermediários. Isso parece bastante intuitivo: uma loja quer cobrar preços altos para explorar os consumidores desinformados e baixos para competir pelos informados. Os preços intermediários não atendem a nenhum objetivo e, portanto, são cobrados com menos frequência. No entanto, eles ainda são cobrados às vezes, já que se ninguém cobrasse um preço intermediário, alguma loja faria isso e teria lucro.

Figura 4.18 – Distribuição dos preços com $r = 30$



O preço médio pago pelos consumidores será de:

$$\overline{p} = \frac{k}{I + \frac{k}{r}} + \left(\frac{k}{I}\right)^{\frac{1}{n-1}} \int_{p^*}^{r} \left(\frac{1}{p} - \frac{1}{r}\right)^{\frac{1}{n-1}} dp, \tag{4.129}$$

e o preço médio pago pelos consumidores desinformados é de:

$$p_U = \left(\frac{k}{I}\right)\left(1 - \frac{\overline{p}}{r}\right). \tag{4.130}$$

O número de firmas pode ser calculdado como:

$$n = \frac{rM}{k}, \tag{4.131}$$

em que r é o preço reserva e M é o número de consumidores desinformados.

4.6.4.5 Diferenças de Qualidade

Há muito se reconhece que um monopolista pode usar diferenças de qualidade para discriminar os consumidores. Testemunhe, por exemplo, as observações perspicazes de Dupuit:

> Não é por causa dos poucos milhares de francos que teriam que ser gastos para colocar um teto sobre as carruagens de terceira classe ou para estofar os assentos da terceira classe que algumas empresas ou outras têm carruagens abertas com bancos de madeira [...] O que a empresa está tentando fazer é impedir que os passageiros que podem pagar a tarifa de segunda classe de viajar de terceira classe; atinge os pobres, não porque quer prejudicá-los, mas para assustar os ricos [...] E é novamente pela mesma razão que as empresas, tendo se provado quase cruéis para os passageiros de terceira classe e para as de segunda classe, tornar-se generoso ao lidar com passageiros de primeira classe. Tendo recusado aos pobres o que é necessário, eles dão aos ricos o que é supérfluo (Ekelund, 1970).

Esta passagem afirma claramente as considerações que o monopolista enfrenta: ao exagerar a diferença de qualidade nas classes de serviço, ele pode efetivamente discriminar preços entre clientes com diferentes disposições a pagar pelo serviço básico de transporte. Este fenômeno foi modelado por Mussa e Rosen (1978), Maskin e Riley (1984) e vários outros. Em um nível formal, a diferença de qualidade pode ser analisada usando técnicas de precificação não linear. Usamos $u(x, t)$ para representar a utilidade de um consumidor do tipo t que consumiu uma quantidade x do bem em questão, e usou $c(x)$ para denotar o custo de produzir a quantidade x. A função de precificação, $r(\cdot)$, mediu o custo de compra de uma quantidade x. Mas suponha que, em vez disso, seja x a qualidade de um bem, $c(x)$ o custo de produção, e $r(x)$ o preço de comprar uma unidade com nível de qualidade x. Dadas essas substituições, o problema de precificação é isomorfo ao problema de precificação de qualidade de um monopolista. Todas as análises e resultados passam praticamente inalterados.

A restrição fundamental no problema de precificação de qualidade é a mesma que na precificação quantitativa, ou seja, a restrição de autosseleção: a escolha de um esquema de preços que induz os consumidores de cada nível de qualidade a preferir sua própria qualidade a qualquer outra qualidade. Esta é a ênfase da passagem do Dupuit citada acima: a natureza das escolhas de qualidade feitas é exagerada de modo a satisfazer as restrições de auto-seleção. Em geral, esperamos que o monopolista amplie o espectro de escolha de qualidade para discriminar de forma mais eficaz os consumidores que enfrenta. Os compradores de itens de menor qualidade geralmente pagam mais do que o custo marginal da qualidade que escolhem. Para mais informações sobre escolha de qualidade, ver Mussa e Rosen (1978), Maskin e Riley (1984), Milgrom (1987), Gabszewicz, Shaked, Sutton e Thisse (1986), e Oren, Smith e Wilson (1982).

4.7 Competição Monopolística

Lembre-se de que assumimos que a curva de demanda do produto do monopolista dependia apenas do preço definido pelo monopolista. No entanto, este é um caso extremo. A maioria dos bens possuem alguns substitutos e os preços desses substitutos afetarão a demanda pelo bem original. O monopolista define seu preço assumindo que todos os produtores de outros bens manterão seus preços, mas é claro que isso não será verdade. Os preços estabelecidos pelas outras empresas responderão - talvez indiretamente - ao preço estabelecido pelo monopolista em questão. Nesta seção, vamos considerar o que acontece quando vários monopolistas "competem" na definição de seus preços e níveis de produção.

Nós imaginamos um grupo de "monopolistas" que vendem produtos similares, mas não idênticos. O preço que os consumidores estão dispostos a pagar pela produção da empresa i depende do nível de produção da empresa i, mas também dos níveis de produção das outras empresas: escrevemos essa função de demanda inversa como $p_i(y_i, y)$, em que $y = (y_1, \dots, y_n)$.

Cada empresa está interessada em maximizar os lucros, isto é, cada firma i quer escolher seu nível de produção y_i de modo a maximizar:

$$p_i(y_i, y)y_i - c(y_i). \tag{4.132}$$

Infelizmente, a demanda enfrentada pela empresa depende do que as outras firmas fazem. Como a empresa deveria prever o comportamento das outras empresas?

Adotaremos uma hipótese comportamental muito simples: a de que a empresa assume que o comportamento das outras empresas será constante. Assim, cada firma i escolherá seu nível de produção y_i^* de modo a satisfazer:

$$p_i(y_i^*, y) + \frac{\partial p_i(y_i^*, y)}{\partial y_i} y_i^* - c_i'(y_i^*) \leq 0 \quad \text{[com igualdade se } y_i^* > 0\text{]}. \tag{4.133}$$

Para cada combinação de níveis de operação para as empresas $y_1, \dots, y_n$, haverá algum nível operacional ótimo para a empresa i. Nós denotaremos esta escolha ótima de produto por $Y_i(y_1, \dots, y_n)$. É claro que o produto da empresa i não é um argumento dessa função, mas parece desnecessário criar uma nova notação apenas para refletir esse fato. Para que o mercado esteja em equilíbrio, a previsão de cada empresa sobre o comportamento da outra as empresas devem ser compatíveis com as outras firmas. Assim, se $(y_1^*, \dots, y_n^*)$ é para ser um equilíbrio, deve satisfazer:

$$y_1^* = Y_1(y_1^*, \dots, y_n^*) \tag{4.134}$$

$$\vdots \tag{4.135}$$

$$y_n^* = Y_n(y_1^*, \dots, y_n^*), \tag{4.136}$$

isto é, y_1^* deve ser a melhor escolha para a empresa i se ela assumir que as outras firmas vão produzir

$y_2^*, \ldots, y_n^*$, e assim por diante. Assim, um equilíbrio de competição monopolista $(y_1^*, \ldots, y_n^*)$ deve satisfazer:

$$p_i(y_i^*, y^*) + \frac{\partial p_i(y_i^*, y^*)}{\partial y_i} y_i^* - c_i'(y_i^*) \leq 0 \quad [\text{com igualdade se } y_i^* > 0]. \tag{4.137}$$

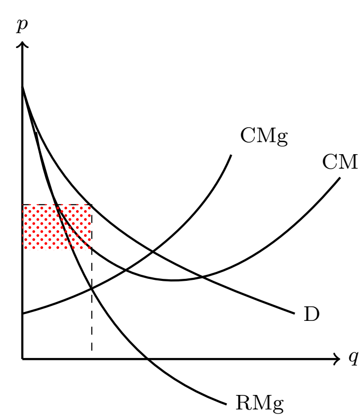
Para cada empresa, sua receita marginal é igual ao seu custo marginal, dadas as ações de todas as outras empresas. Isso é ilustrado na Figura 4.19. Agora, no equilíbrio de competição monopolista representado na Figura 4.19, a empresa i está obtendo lucros positivos. Portanto, esperamos que outras empresas entrem no setor e compartilhem o mercado com a empresa, para que o lucro da empresa diminua devido a produtos substitutos próximos. Assim, a longo prazo, as empresas entrariam no setor até que os lucros de cada empresa fossem levados a zero. Isto significa que a empresa i deve cobrar um preço p_i^* e produzir um produto y_i^* tal que:

$$p_i^* y_i^* - c_i(y^*) \leq 0 \quad [\text{com igualdade se } y_i^* > 0], \tag{4.138}$$

ou então

$$p_i^* - \frac{c_i(y^*)}{y_i^*} \leq 0 \quad [\text{com igualdade se } y_i^* > 0]. \tag{4.139}$$

Figura 4.19 – Equilíbrio de curto prazo em competição monopolística



Assim, o preço deve ser igual ao custo médio e à curva de demanda que a empresa enfrenta. Como resultado, enquanto a curva de demanda enfrentada por cada empresa tiver algum declive negativo, cada empresa produzirá em um ponto em que o custo médio seja maior do que os custos médios mínimos. Assim, como as firmas competitivas puras, os lucros obtidos por cada empresa são zero e muito próximos do equilíbrio competitivo de longo prazo. Por outro lado, como um monopolista puro, ainda resulta em alocação ineficiente, desde que a curva de demanda que a empresa enfrenta tenha uma inclinação negativa.

4.8 Oligopólio

A teoria da concorrência monopolística assume que as firmas tomam decisões sem explicitamente levar em consideração as reações das concorrentes. Mas quando as ações de uma firma produzem, de fato, reações por parte das competidoras, está-se numa situação de oligopólio.

As características do oligopólio são:

- Pequeno número de empresas
- Interdependência entre as firmas
- Consideráveis obstáculos à entrada de novas firmas
- Produto, em geral, diferenciado
- Concorrência extra-preço (diferenciação do produto, propaganda, serviços especiais)

São exemplos de indústrias oligopolísticas no Brasil: alumínio, automóveis, equipamentos elétricos, aço e petróleo, etc. Em cada uma das indústrias, um pequeno número de firmas produzem pelo menos uma grande porcentagem da produção total.

A diferenciação é o principal meio de competição numa estrutura de mercado oligopolizado, e tanto pode ser facilmente perceptível no produto, com pode ser conseguida através de propaganda maciça. A competição de preço não é normalmente uma parte da estratégia de mercado para uma empresa oligopolística. A firma não pode estar segura da localização ou do nível da sua curva de demanda porque ela nunca pode estar certa da intensidade da reação das firmas competitivas a uma mudança de preço. Se uma determinada firma estava produzindo a quantidade q_0, a qual vinha sendo vendida ao preço p_0 durante algum tempo, a firma pode geralmente esperar as seguintes reações a uma mudança de preço:

- Se ela aumentar o preço de seu produto, as firmas concorrentes não elevarão seus preços. O decréscimo na quantidade vendida devido ao preço majorado será uma função do grau de diferenciação do seu produto. A menos que os esforços à diferenciação tenham tido um grande sucesso, a porção da curva de demanda para preços acima de p_0 será elástica (o que significa um decréscimo percentual maior na quantidade do que o aumento relativo em preço e, portanto, uma redução na receita total).
- Se a firma baixar o preço do seu produto, as firmas concorrentes tenderão também a reduzir os seus preços. Isto fará com que não haja um aumento expressivo na quantidade vendida desta firma devido ao decréscimo de preço, tornando a porção da curva de demanda, abaixo de p_0, inelástica (o que significa uma redução na receita total, se o preço cair). Por causa desta interdependência, os preços numa estrutura oligopolista tendem a ser estáveis, ou seja, uma vez estabelecido o preço p_0, este tende a ser mantido. A diferença entre preços de várias firmas é uma função do sucesso de diferenciação do produto. Como consequência tem-se a chamada

curva de demanda quebrada, em que cada firma tem uma curva de demanda semelhante à esta. Portanto, a quantidade e o preço que a firma seleciona é uma função da sua estrutura de custo e da sua habilidade em diferenciar seu produto.

Na agroindústria brasileira, temos pelo lado do produto agrícola processado, alguns casos de oligopólios nas indústrias de: óleos vegetais, café solúvel, chocolate, cigarros, frutas, sucos, etc. Já pelo lado dos fatores de produção vendidos à agricultura, temos os exemplos das indústrias de: rações, fertilizantes, defensivos e tratores.

Dada a posição dentro da qual a firma oligopolista opera, há uma forte tendência para a liderança de preço, caso haja uma firma dominante no mercado. Uma firma é dominante ou por sua relativamente grande participação no mercado e/ou por vantagens em termos de custos menores de produção. A firma dominante determina seu lucro produzindo determinada quantidade de produto, e as outras firmas vendem o restante para completar o mercado. A líder permite às seguidoras vender tudo o que elas desejam pelo preço por ela estabelecido, e isto não representa um problema para ela, porque o nível de produção das firmas seguidoras é limitado por suas respectivas curvas de custo marginal.

Na agricultura ou no setor agroindustrial, a liderança de preço pela firma dominante é um pouco mais difícil de ocorrer, mas também existe. Não fosse o preço administrado pelo governo, a "Souza Cruz", que controla dois terços do subsetor cigarros e fumos, poderia ser considerada um exemplo típico de empresa dominante.

4.8.1 Competição em Cournot

Considere que haja n firmas. Cada firma i produz $q_i \geq 0$ unidades de um bem ao custo marginal $c \geq 0$ e vende ao preço

$$P = \max\{1 - Q, 0\}, \tag{4.140}$$

em que

$$Q = q_1 + \ldots + q_n, \tag{4.141}$$

é a oferta total. Cada firma maximiza o lucro esperado. Portanto, o payoff da firma i é

$$\pi_i = q_i(P - C). \tag{4.142}$$

Assumindo que as informações acima são de conhecimento comum, podemos escrever o jogo na forma normal, em que:

- $N = \{1, 2, \ldots, n\}$ é o conjunto de jogadores.

- $S_=[0, \infty)$ é o espaço de estratégias do jogador i, em que uma estratégia típica é a quantidade q_i a ser produzida pela firma i.

- $\pi_i = S_1 \times S_2 \times \ldots \times S_n \to \mathbb{R}$ é a função payoff.

É útil conhecer a melhor resposta de uma empresa aos níveis de produção das outras empresas. Seja

$$Q_{-i} = \sum_{j \neq i} q_j, \tag{4.143}$$

a oferta total das empresas que não a empresa i. Se $Q_{-i} > 1$, então o preço é $P = 0$ e a melhor resposta da empresa i é produzir zero e obter lucro zero. Agora assuma $Q_{-i} < 1$. Para qualquer $q_i \in (0, 1 - Q_{-i})$, a função lucro da firma i é

$$q_i^B(Q_{-i}) = \frac{1 - Q_{-i} - c}{2}. \tag{4.144}$$

A função lucro é plotada na Figura 4.20. A função melhor resposta é plotada na Figura 4.21.

Figura 4.20 – Função lucro

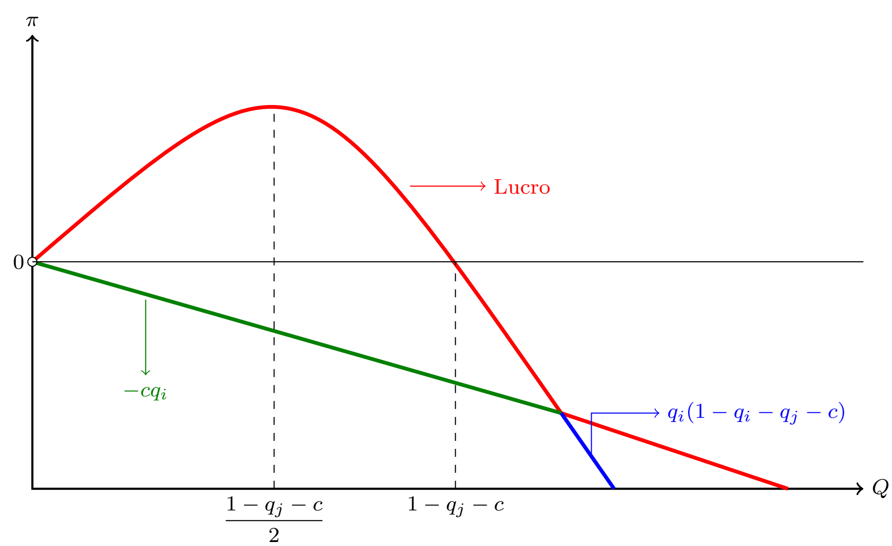
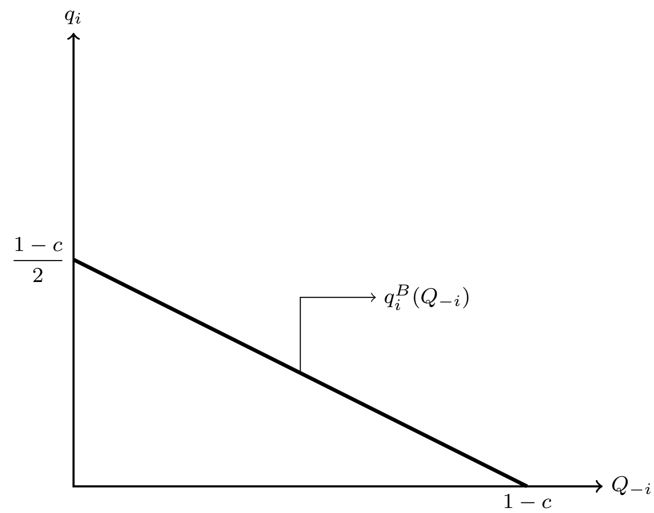


Figura 4.21 – Função melhor resposta



Agora considere o caso de duas firmas, duopólio de Cournot. Nesse caso, $Q_{-i} = q_j$ para $i \neq j$.

Um equilíbrio de Nash deve (q_1, q_2) deve satisfazer

$$q_1 = q_1^B(q_2) \equiv \frac{1 - q_2 - c}{2}, \tag{4.145}$$

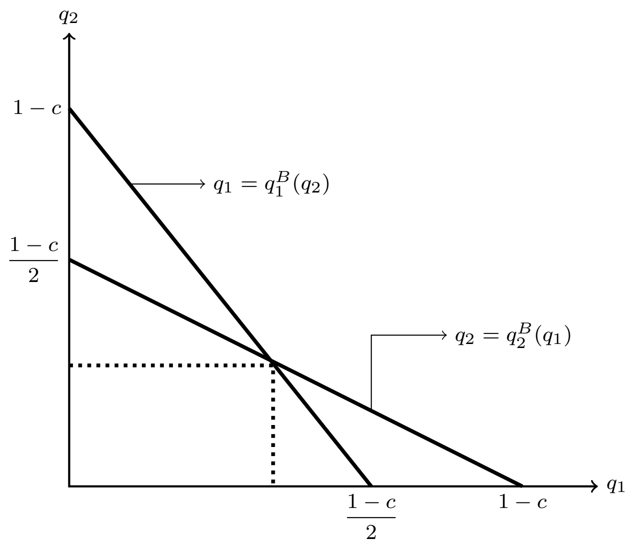
$$q_2 = q_2^B(q_1) \equiv \frac{1 - q_1 - c}{2}. \tag{4.146}$$

Resolvendo essas duas equações simultaneamente, podemos obter

$$q_1^* = q_2^* = \frac{1 - c}{3}, \tag{4.147}$$

como o único equilíbrio de Nash. Graficamente, Figura 4.22, podemos plotar a função melhor resposta de ambas as firmas e a intersecção corresponde ao equilíbrio de Nash.

Figura 4.22 – Função melhor resposta em duopólio de Cournot



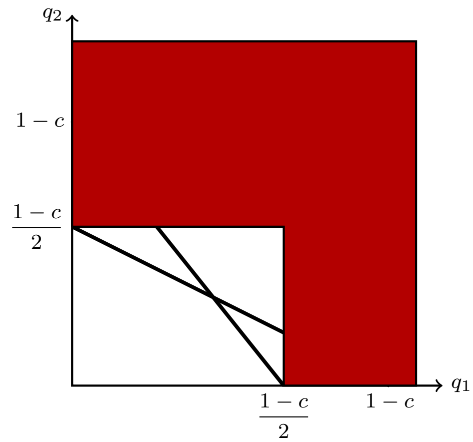
O jogo do duopólio de Cournot (linear) considerado aqui é de domínio solucionável. Isto é, há uma estratégia racionalizável única. Vamos primeiro considerar as primeiras rodadas de eliminação para ver isso intuitivamente. Então mostrarei matematicamente que esse é realmente o caso.

- Round 1

  Observe que uma estratégia $\widehat{q}_i > \dfrac{1-c}{2}$ é estritamente dominada por $\dfrac{1-c}{2}$. Para ver isso, considere qualquer q_j. Como na Figura 4.20, $\pi_i(q_i, q_j)$ é estritamente crescente até $q_i = \dfrac{1-c-q_j}{2}$ e decrescente a partir desse ponto. Em particular,

$$\pi_i\left(\frac{1-c-q_j}{2}, q_j\right) \geq \pi_i\left(\frac{1-c}{2}, q_j\right) > \pi_i(\widehat{q}_i, q_j). \tag{4.148}$$

  mostrando que $\widehat{q}_i$ é estritamente dominada por $\dfrac{1-c}{2}$. Portanto, eliminamos todos $\widehat{q}_i > \dfrac{1-c}{2}$ para todo jogador i. As estratégias resultantes são as seguintes, em que a área pintada foi eliminada.

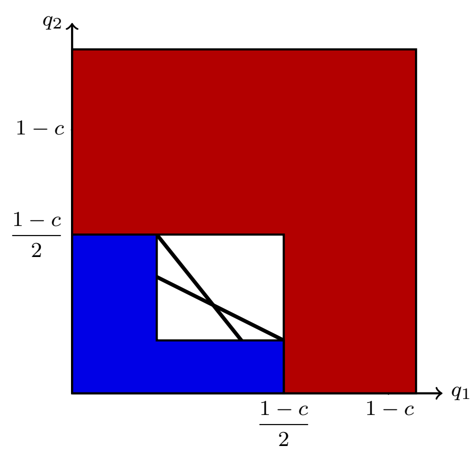


- Round 2

  No restante do jogo, $q_j \leq \frac{1-c}{2}$. Consequentemente, qualquer estratégia $\widehat{q}_i < \frac{1-c}{4}$ é estritamente dominada por $\frac{1-c}{4}$. Para ver isso, considere qualquer $q_j \leq \frac{1-c}{2}$. Como na Figura 4.20, $\pi_i(q_i, q_j)$ é estritamente crescente até $q_i = \frac{1-c-q_j}{2}$, o que é maior ou igual a $\frac{1-c}{4}$. Em particular,

$$\pi_i(\widehat{q}_i, q_j) < \pi_i\left(\frac{1-c}{4}, q_j\right) \leq \pi_i\left(\frac{1-c-q_j}{2}, q_j\right), \tag{4.149}$$

mostrando que $\widehat{q}_i$ é estritamente dominada por $\frac{1-c}{4}$. Portanto, eliminamos todos $\widehat{q}_i < \frac{1-c}{4}$ para todo jogador i. As estratégias resultantes são as seguintes, em que a área pintada foi eliminada.



Observe que o jogo restante é uma réplica menor do jogo original. Aplicando o mesmo procedimento repetidamente, pode-se eliminar todas as estratégias, exceto o equilíbrio de

Nash. (Após cada duas rodadas, uma réplica menor é obtida.) Portanto, a única estratégia racionalizável é a estratégia única do equilíbrio de Nash.

Lema 4.8.1. *Dado que* $q_j \leq \overline{q}$, *toda estratégia* $\widehat{q}_i$ *com* $\widehat{q}_i < q_i^B(\overline{q})$ *é estritamente dominada por* $q_i^B(\overline{q}) = \dfrac{1-\overline{q}-c}{2}$. *Dado que* $q_j \geq \overline{q}$, *toda estratégia* $\widehat{q}_i$ *com* $\widehat{q}_i > q_i^B(\overline{q})$ *é estritamente dominada por* $q_i^B(\overline{q}) = \dfrac{1-\overline{q}-c}{2}$

Demonstração. Para provar a primeira parte do lema, tome um q_j qualquer tal que $q_j \leq \overline{q}$. Note que $\pi_i(q_i; q_j)$ é estritamente crescente em q_i para qualquer $q_i < q_i^B(q_j)$. Dado que $\widehat{q}_i < q_i^B(\overline{q}) < q_i^B(q_j)$, uma vez que q_i^B é decrescente, isso implica que

$$\pi(\widehat{q}_i, q_j) < \pi_i(q_i^B(\overline{q}), q_j). \tag{4.150}$$

Isto é, $\widehat{q}_i$ é estritamente dominado por $q_i^B(\overline{q})$.

Para provar a segunda parte do lema, tome um q_j qualquer tal que $q_j \leq \overline{q}$. Note que $\pi_i(q_i; q_j)$ é estritamente decrescente em q_i para qualquer $q_i > q_i^B(q_j)$. Dado que $q_i^B(q_j) < q_i^B(\overline{q}) < \widehat{q}_i$, isso implica que

$$\pi(\widehat{q}_i, q_j) < \pi_i(q_i^B(\overline{q}), q_j). \tag{4.151}$$

Isto é, $\widehat{q}_i$ é estritamente dominado por $q_i^B(\overline{q})$. ■

Considere a sequência $q^0, q^1, q^2, \ldots$, com $q^0 = 0$ e

$$q^m = q_i^B(q^{m-1}) \equiv \frac{1-q^{m-1}-c}{2} = \frac{1-c}{2} - \frac{q^{m-1}}{2}, \quad \forall m > 0. \tag{4.152}$$

Isto é,

$$\begin{aligned}
q^0 &= 0 && (4.153)\\
q^1 &= \frac{1-c}{2} && (4.154)\\
q^2 &= \frac{1-c}{2} - \frac{1-c}{4} && (4.155)\\
q^2 &= \frac{1-c}{2} - \frac{1-c}{4} + \frac{1-c}{8} && (4.156)\\
&\vdots\\
q^m &= \frac{1-c}{2} - \frac{1-c}{4} + \frac{1-c}{8} - \ldots - (-1)^m \frac{1-c}{2^m} && (4.157)\\
&\vdots
\end{aligned}$$

Teorema 4.8.1. *O conjunto de estratégias restantes após qualquer rodada ímpar m, $m = (1, 3, \ldots)$, é $[q^{m-1}, q^m]$. O conjunto de estratégias restantes após qualquer rodada par m, $m = (2, 4, \ldots)$, é $[q^m, q^{m-1}]$. O conjunto de estratégias racionalizáveis é* $\left\{\dfrac{1-c}{3}\right\}$.

Vamos agora considerar o caso de três ou mais empresas. Quando existem três ou mais empresas, a racionalização não ajuda: não se pode eliminar qualquer estratégia menos do que a produção monopolista $q = \dfrac{1-c}{2}$.

Na primeira rodada, pode-se eliminar qualquer estratégia $q_i > \dfrac{1-c}{2}$ usando o mesmo argumento no caso do duopólio. Mas na segunda rodada, a oferta total máxima possível pelas outras empresas é

$$(n-1)\frac{1-c}{2} \geq 1-c, \tag{4.158}$$

em que n é o número de firmas. A melhor resposta para este nível de oferta agregada é 0. Portanto, não é possível eliminar qualquer estratégia na rodada 2. O processo de eliminação é interrompido, gerando $\left[0, \dfrac{1-c}{2}\right]$ como o conjunto de estratégias racionalizáveis. Como o conjunto de estratégias racionalizáveis é grande, a racionalização tem um poder preditivo fraco neste jogo.

Embora a racionalização tenha um fraco poder preditivo, na medida em que o conjunto de estratégias racionalizáveis é grande, o equilíbrio de Nash continua a ter um forte poder preditivo. Existe um equilíbrio único de Nash. Lembre-se que $q^* = (q_1^*, q_2^*, \ldots, q_n^*)$ é um equilíbrio de Nash se e somente se

$$q_i^* = q_i^B\left(\sum_{j\neq i} q_j^*\right) = \frac{1 - \sum_{j\neq i} q_j^* - c}{2}, \quad \forall i \tag{4.159}$$

Reescrevendo este sistema mais explicitamente:

$$2q_1^* + q_2^* + \ldots + q_n^* = 1 - c \tag{4.160}$$
$$q_1^* + 2q_2^* + \ldots + q_n^* = 1 - c \tag{4.161}$$
$$\vdots \tag{4.162}$$
$$q_1^* + q_2^* + \ldots + 2q_n^* = 1 - c \tag{4.163}$$

Para qualquer i e j, subtraindo a j-ésima equação da i-ésima equação, obtemos

$$q_i^* - q_j^* = 0. \tag{4.164}$$

Portanto,

$$q_1^* = q_2^* = \ldots = q_n^*. \tag{4.165}$$

Substituindo dentro da primeira equação, encontramos:

$$(n+1)q_1^* = 1 - c, \tag{4.166}$$

ou seja,

$$q_1^* = q_2^* = \ldots = q_n^* = \frac{1-c}{n+1}. \tag{4.167}$$

Portanto, existe um único equilíbrio de Nash, no qual cada firma produz $\dfrac{1-c}{n+1}$.
No único equilíbrio, a oferta total é

$$Q = \frac{n}{n+1}(1-c), \tag{4.168}$$

e o preço é dado por

$$P = c + \frac{1-c}{n+1}. \tag{4.169}$$

O lucro por firma é

$$\pi = \left(\frac{1-c}{n+1}\right)^2. \tag{4.170}$$

À medida que n vai para o infinito, a oferta total Q converge para $1-c$ e o preço P converge para c. Estes são os valores nos quais a demanda ($P = \max\{1-Q, 0\}$) é igual a oferta ($P = c$), que é chamado de equilíbrio perfeitamente competitivo. Quando há poucas empresas, no entanto, o preço é significativamente maior do que o preço competitivo c, e a oferta total é significativamente menor do que a oferta competitiva $1-c$.

4.8.2 Competição em Bertrand

Considere duas empresas. Simultaneamente, cada empresa define um preço p_i. A empresa i com o preço mais baixo $p_i < p_j$ vende $1 - p_i$ unidades e a outra empresa não vende nenhuma. Se as firmas definem o mesmo preço, a demanda é dividida igualmente entre elas. Ou seja, a quantidade de vendas para a empresa i é

$$Q_i(p_1,p_2) = \begin{cases} 1-p_i & \text{se } p_i < p_j \\ \dfrac{1-p_i}{2} & \text{se } p_i = p_j \\ 0 & \text{caso contrário} \end{cases} \tag{4.171}$$

Suponha que não custa nada produzir o bem (ou seja, $c = 0$). Portanto, o lucro de uma empresa i é

$$\pi_i = p_i Q_i(p_1,p_2) = \begin{cases} (1-p_i)p_i & \text{se } p_i < p_j \\ \dfrac{(1-p_i)p_i}{2} & \text{se } p_i = p_j \\ 0 & \text{caso contrário} \end{cases} \tag{4.172}$$

Assumindo que as informações acima são de conhecimento comum, podemos escrever o jogo na forma normal, em que:

- $N = \{1, 2\}$ é o conjunto de jogadores.
- $S_= [0, \infty)$ é o espaço de estratégias do jogador i, em que uma estratégia típica é o preço p_i a ser cobrado pela firma i.
- π_i é a função payoff.

Observe que quando $p_j = 0$, $\pi_i = (p_1, p_2) = 0$ para todo p_i e, portanto, todo p_i é a melhor resposta para $p_j = 0$. Isso tem duas implicações importantes:

1. Toda estratégia é racionalizável (não se pode eliminar nenhuma estratégia, porque cada uma delas é a melhor resposta a zero).
2. $p_1^* = p_2^* = 0$ é um equilíbrio de Nash.

Este é de fato o único equilíbrio de Nash. Em outras palavras, mesmo com duas firmas, quando as firmas competem definindo preços, o equilíbrio competitivo vai emergir. Se modificarmos o jogo ligeiramente, discretizando o conjunto de preços permitidos e colocando um preço mínimo, então o jogo se torna solucionável por dominância, ou seja, apenas uma estratégia permanece racionalizável. No jogo modificado, o preço mínimo é a única estratégia racionalizável, como no equilíbrio competitivo.

Teorema 4.8.2. *No modelo de Bertrand, o único equilíbrio de Nash é* $p^* = (0, 0)$.

Demonstração. Já vimos que $p^* = (0, 0)$ é um equilíbrio de Nash. Eu mostrarei aqui que se (p_1, p_2) é um equilíbrio de Nash, então $p_1 = p_2 = 0$. Para fazer isso, tome qualquer equilíbrio de Nash (p_1, p_2). Por contradição, suponha que $p_i > p_j$. Se $p_j = 0$, então $\pi_i(p_i, p_j) = 0$ enquanto $\pi(p_i, p_j) = \dfrac{(1-p_i)p_i}{2}$, ou seja, escolher p_i é um desvio lucrativo para a empresa j, mostrando que $p_i > p_j = 0$

não é um equilíbrio de Nash. Portanto, para $p_i > p_j$ ser um equilíbrio, deve ser que $p_j > 0$. Mas, então, a empresa i tem um desvio lucrativo: $\pi_i(p_i, p_j) = 0$ enquanto $\pi(p_i, p_j) = \frac{(1-p_i)p_i}{2}$. Em suma, isso mostra que não se pode ter $p_i > p_j$ em equilíbrio. Portanto, $p_1 = p_2$. Mas se $p_1 = p_2$ é um equilíbrio de Nash, então deve ser que $p_1 = p_2 = 0$. Isto ocorre porque se $p_1 > p_2 = 0$, então a firma 1 teria um desvio lucrativo e esse desvio tenderia a voltar ao equilíbrio. ■

4.8.3 Duopólio de Stackelberger

No duopólio de Cournot, assumimos que as empresas definem as quantidades simultaneamente. Isso reflete a suposição de que nenhuma empresa pode se comprometer com um nível de quantidade. Às vezes, uma empresa pode se comprometer com um nível de quantidade. Por exemplo, uma empresa pode já estar no mercado e construir sua fábrica e armazéns, e seu nível de produção é fixo. A outra firma entra no mercado sabendo mais tarde o nível de produção da primeira firma. Vamos considerar tal situação, que é chamada de duopólio de Stackelberg. Existem duas empresas. A primeira firma é chamada de líder, e a segunda firma é chamada de seguidora. Como antes, tomamos o custo marginal constante.

- A firma líder escolhe seu nível de produção q_1.
- Então, conhecendo q_2, a seguidora escolhe seu nível de produção q_2.
- Cada firma i escolhe sua quantidade q_i aos preços de mercado,

$$P(q_1 + q_2) = \max\{1 - (q_1 + q_2), 0\}, \tag{4.173}$$

  resultando em um lucro

$$\pi_i(q_1, q_2) = q_i(P(q_1 + q_2) - c). \tag{4.174}$$

- No nó inicial, a firma 1 escolhe uma ação q_1, em que o conjunto de ações disponíveis é $[0, \infty)$.
- Depois da ação da firma 1, a firma 2 se move e escolhe uma ação q_2, em que o conjunto de ações disponíveis é $[0, \infty)$.
- Cada uma das ações é levada ao nó terminal, no qual o vetor de payoffs é $(\pi_1(q_1, q_2), \pi_2(q_1, q_2))$.

Observe que a estratégia da empresa 1 é um número real q_1 definido no intervalo $[0, \infty)$ e, mais importante, a estratégia da empresa 2 é uma função definido no intervalo $[0, \infty) \times [0, \infty)$, que atribui um nível de produção $q_2(q_1)$ para cada q_1. Essas estratégias com a função utilidade $u_i(q_1, q_2) = \pi_i(q_1, q_2(q_1))$ nos dão a forma normal do jogo.

Vamos resolver por indução retroativa. Dado que $q_1 \leq 1 - c$, o melhor nível de produção para a seguidora é

$$q_2^*(q_1) = \frac{1 - q_1 - c}{2}, \tag{4.175}$$

resultando no seguinte vetor de payoffs[7]:

$$\begin{pmatrix} \pi_1(q_1, q_2^*(q_1)) \\ \pi_2(q_1, q_2^*(q_1)) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} q_1(1 - q_1 - c) \\ \frac{1}{4}(1 - q_1 - c)^2 \end{pmatrix}. \tag{4.176}$$

Substituindo os movimentos da empresa 2 com os payoffs associados, obtemos um jogo no qual a empresa escolhe um nível de produção q_1 que leva ao vetor de payoffs acima. Neste jogo, a empresa 1 maximiza seu lucro ao escolher

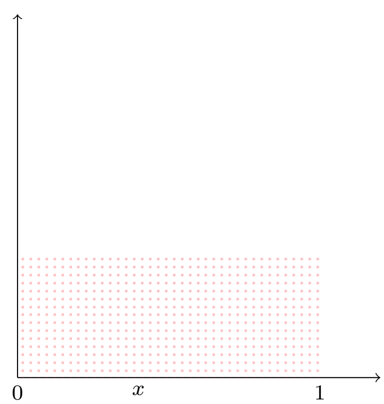
$$q_1^* = \frac{1 - c}{2}. \tag{4.177}$$

Também devemos verificar se há um equilíbrio de Nash desse jogo no qual o seguidor produz a quantidade de Cournot independentemente do que o líder produz, e o líder produz a quantidade de Cournot. É claro que isso não é consistente com a indução retroativa: quando o seguidor sabe que o líder produziu a quantidade de Stackelberg, ele mudará de ideia e produzirá uma quantidade menor, a quantidade computada durante a indução retroativa.

4.8.4 Modelo de Hotelling

Os produtos fornecidos em um mercado não regulamentado seriam muito semelhantes ou muito diferentes em relação ao ótimo social?

Figura 4.23 – Distribuição dos consumidores

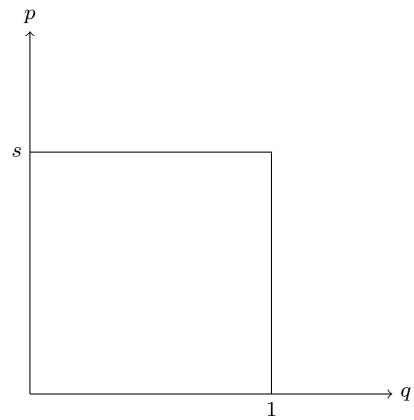


Os consumidores são distribuídos uniformemente ao longo do intervalo $[0, 1]$. Um consumidor

[7] Note que $q_1 \left(1 - q_1 - \frac{1 - q_1 - c}{2} - c\right) = \frac{1}{2} q_1(1 - q_1 - c)$.

prefere a variedade de produtos x. Por fim, assumimos que o consumidor tem demanda unitária:

Figura 4.24 – Demanda unitária



A desutilidade de consumir uma variedade de produtos y é dada por $t(|y-x|)$. Pode ser visto como um custo de transporte. Vamos assumir um custo de transporte linear, isto é, $t(d) = td$. Vamos assumir que a firma 1 localizada em 0 tem preço $p_1 + tx$ e a a firma 2 localizada em 1 tem preço $p_2 + t(1-x)$.

O consumidor é indiferente entre as firmas, tal que:

$$s - p_1 - t\widetilde{x} = s - p_2 - t(1-\widetilde{x}) \quad \Longrightarrow \quad \widetilde{x}(p_1, p_2) = \frac{1}{2} + \frac{p_2 - p_1}{2t}. \tag{4.178}$$

Assim,

$$D_1(p_1, p_2) = \widetilde{x}(p_1, p_2) = \frac{1}{2} + \frac{p_2 - p_1}{2t}, \tag{4.179}$$

$$D_2(p_1, p_2) = 1 - \widetilde{x}(p_1, p_2) = \frac{1}{2} + \frac{p_1 - p_2}{2t}. \tag{4.180}$$

Assumindo que o custo de produção é c, o lucro é dado por:

$$\pi_1(p_1, p_2) = (p_1 - c)\left[\frac{1}{2} + \frac{p_2 - p_1}{2t}\right], \tag{4.181}$$

$$\pi_2(p_1, p_2) = (p_2 - c)\left[\frac{1}{2} + \frac{p_1 - p_2}{2t}\right]. \tag{4.182}$$

As condições de primeira ordem são:

$$\frac{\partial \pi_1(p_1, p_2)}{\partial p_1} = 0 \quad \Longrightarrow \quad (p_1 - c)\left(-\frac{1}{2t}\right) + \frac{1}{2} + \frac{p_2 - p_1}{2t} = 0, \tag{4.183}$$

$$\frac{\partial \pi_2(p_1,p_2)}{\partial p_2} = 0 \implies (p_2 - c)\left(-\frac{1}{2t}\right) + \frac{1}{2} + \frac{p_1 - p_2}{2t} = 0, \tag{4.184}$$

o que implica que $p_1^* = p_2^* = c + t$.

O consumidor é indiferente a comprar se

$$\begin{aligned} s - p - t\widetilde{x} &\geq 0 \\ s &\geq p + \frac{t}{2} \\ s &\geq p - t + \frac{t}{2} + t \\ s &\geq c + \frac{3}{2}t. \end{aligned} \tag{4.185}$$

Da condição de primeira ordem, temos a função melhor resposta:

$$p_1 = \frac{1}{2}(p_2 + c + t), \tag{4.186}$$

$$p_2 = \frac{1}{2}(p_1 + c + t). \tag{4.187}$$

O grau de diferenciação do produto é t. A diferenciação do produto torna as empresas menos agressivas em seus preços.

Mas são 0 e 1 as variações de produto de equilíbrio das firmas? Vamos ver por meio de um jogo de diferenciação de produto em dois estágios: na etapa 1 as empresas escolhem locais em $[0,1]$ e na etapa 2 as empresas escolhem os preços. Vamos assumir que os custos de transporte são quadráticos.

Começando pelo estágio 2, as firmas se localizam em a e $1-b$ com $a \geq 0$, $b \geq 0$ e $a + b \leq 1$.

O consumidor é indiferente entre as firmas, tal que:

$$p_1 + t(\widetilde{x} - a)^2 = p_2 + t(1 - b - \widetilde{x})^2 \implies \widetilde{x}(p_1,p_2) = a + \frac{1}{2}(1 - a - b) + \frac{p_2 - p_1}{2t(1-a-b)}. \tag{4.188}$$

Assim,

$$\begin{aligned} D_1(p_1,p_2) = \widetilde{x}(p_1,p_2) &= a + \frac{1}{2}(1-a-b) + \frac{p_2 - p_1}{2t(1-a-b)} \\ , &= \frac{1}{2}(1+a-b) + \frac{p_2 - p_1}{2t(1-a-b)} \end{aligned} \tag{4.189}$$

$$\begin{aligned} D_2(p_1,p_2) = 1 - \widetilde{x}(p_1,p_2) &= 1 - \left[a + \frac{1}{2}(1-a-b) + \frac{p_2 - p_1}{2t(1-a-b)}\right] \\ &= \frac{1}{2}(1-a+b) + \frac{p_1 - p_2}{2t(1-a-b)}. \end{aligned} \tag{4.190}$$

Assumindo que o custo de produção é c, o lucro é dado por:

$$\pi_1(p_1,p_2) = (p_1 - c)\left[\frac{1}{2}(1+a-b) + \frac{p_2 - p_1}{2t(1-a-b)}\right], \tag{4.191}$$

$$\pi_2(p_1,p_2) = (p_2 - c)\left[\frac{1}{2}(1-a+b) + \frac{p_1 - p_2}{2t(1-a-b)}\right]. \tag{4.192}$$

As condições de primeira ordem são:

$$\frac{\partial \pi_1(p_1,p_2)}{\partial p_1} = 0 \quad \Longrightarrow \quad 2p_1 - p_2 = c + t(1-a-b)(1+a-b), \tag{4.193}$$

$$\frac{\partial \pi_2(p_1,p_2)}{\partial p_2} = 0 \quad \Longrightarrow \quad 2p_2 - p_1 = c + t(1-a-b)(1-a+b). \tag{4.194}$$

Em equilíbrio,

$$p_1 = c + t(1-a-b)\left(1 + \frac{a-b}{3}\right), \tag{4.195}$$

$$p_2 = c + t(1-a-b)\left(1 + \frac{b-a}{3}\right). \tag{4.196}$$

Se a localização for a mesma, isto é, $a = b$, então $p_1 = p_2 = c + t(1-2a)$.

O preço de uma empresa diminui quando a outra empresa se aproxima, isto é, $\frac{\partial p_1}{\partial b} < 0$ e $\frac{\partial p_2}{\partial a} < 0$.

Assim, no estágio 1, temos:

$$\pi_1(a,b) = (p_1(a,b) - c)D_1(a,b,p_1(a,b),p_2(a,b)). \tag{4.197}$$

Logo,

$$\begin{aligned}\frac{\partial \pi_1(a,b)}{\partial a} &= D_1\frac{\partial p_1}{\partial a} + (p_1 - c)\left[\frac{\partial D_1}{\partial a} + \frac{\partial D_1}{\partial p_1}\frac{\partial p_1}{\partial a} + \frac{\partial D_1}{\partial p_2}\frac{\partial p_2}{\partial a}\right]\\ &= \frac{\partial p_1}{\partial a}\left[D_1 + (p_1 - c)\frac{\partial D_1}{\partial p_1}\right] + (p_1 - c)\left[\frac{\partial D_1}{\partial a} + \frac{\partial D_1}{\partial p_2}\frac{\partial p_2}{\partial a}\right]\\ &= (p_1 - c)\left[\frac{\partial D_1}{\partial a} + \frac{\partial D_1}{\partial p_2}\frac{\partial p_2}{\partial a}\right],\end{aligned} \tag{4.198}$$

em que $\frac{\partial D_1}{\partial a} > 0$ é o efeito direto e $\frac{\partial D_1}{\partial p_2}\frac{\partial p_2}{\partial a} < 0$ é o efeito estratégico com $\frac{\partial D_1}{\partial p_2} > 0$ e $\frac{\partial p_2}{\partial a} < 0$.

Mover para o meio tem um efeito direto positivo vs. um efeito estratégico negativo. Veja que:

$$
\begin{aligned}
\frac{\partial D_1}{\partial a} + \frac{\partial D_1}{\partial p_2}\frac{\partial p_2}{\partial a} &= \frac{3-5a-b}{6(1-a-b)} + \frac{1}{2t(1-a-b)}\frac{2}{3}t(a-2) \\
&= -\frac{3a+b+1}{6(1-a-b)} < 0. \qquad (4.199)
\end{aligned}
$$

O efeito estratégico é mais forte que o efeito direto. Assim, temos diferenciação máxima em equilíbrio.

4.8.5 Modelo de Salop

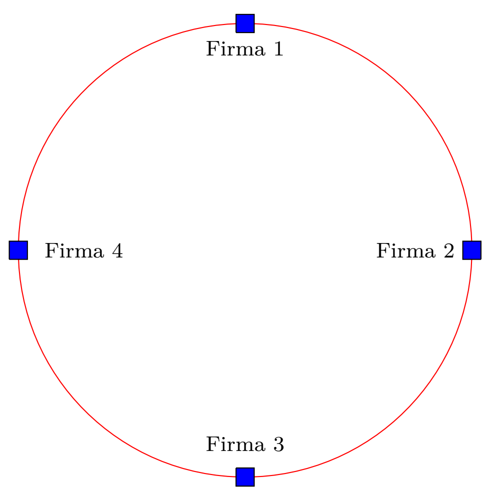
Para evitar o problema da não existência de equilíbrio, Salop (1979) desenvolveu um modelo circular que introduz duas grandes mudanças no modelo de Hotelling.

1. As empresas estão localizadas em torno de um círculo em vez de uma linha.
2. Consideração de um bem externo (segundo), não diferenciado e competitivo.

As firmas estão localizadas em torno de um círculo (circunferência = 1) com igual distância ($1/N$) uma da outra. O custo fixo é f e o custo marginal é c. O lucro é dado por:

$$
\pi_i(q_i) = (p_i - c)q_i - f \qquad (4.200)
$$

Figura 4.25 – Modelo circular de Salop



Uniformemente localizado ao redor do círculo (por exemplo, serviços de linha aérea, ônibus e trem, etc). A localização de um consumidor x^* representa o tipo de produto mais preferido do consumidor. Cada consumidor compra uma unidade. o custo de transporte por unidade de distância é t. Dado o preço, p, cobrado pelas empresas adjacentes (esquerda e direita) e p_1 cobrado pela empresa 1, podemos derivar a localização do consumidor indiferente localizado à distância $x \in (0, 1/N)$.

$$s - p_1 - tx = s - p - t\left(\frac{1}{N} - x\right) \implies D^1(p_1, p) = 2x = \frac{p - p_1}{t} + \frac{1}{N} \tag{4.201}$$

Portanto,

$$\pi_1 = (p_1 - c)\left(\frac{p - p_1}{t} + \frac{1}{N}\right) - f \tag{4.202}$$

Logo,

$$\frac{\partial \pi}{\partial p_1} = 0 \implies p_1 = \frac{t}{2N} + \frac{p + c}{2} \tag{4.203}$$

Por simetria temos que $p_1 = p$. Portanto,

$$p = c + \frac{t}{N} \implies p - c = \frac{t}{N} \tag{4.204}$$

Assim como no modelo de Hotelling, o preço e a margem de lucro aumentam com o custo de transporte t e diminuem com N. Suponha que a entrada de novas empresas seja possível (entrada livre): a entrada ocorrerá até que o lucro seja totalmente dissipado. Assim,

$$\pi_i = (p - c)\frac{1}{N} - f = \frac{t}{N^2} - f = 0 \implies N^* = \sqrt{\frac{t}{f}}, p^* = c + \sqrt{tf} \tag{4.205}$$

O preço da empresa está acima do custo marginal, mas ainda assim não obtém lucro.

Sob a entrada livre, um aumento no custo fixo (f) causa uma diminuição no número de empresas (N) e um aumento na margem de lucro ($p - c$).

Sob a entrada livre, um aumento no custo de transporte (t) aumenta tanto a margem de lucro ($p - c$) quanto o número de empresas (N).

Quando o custo fixo (f) cai para zero (0), o número de empresas tende a ser muito grande (infinito): $\lim_{f \to 0} N^* \to \infty$.

Até agora, temos discutido o caso em que as empresas estão localizadas suficientemente próximas umas das outras e competem pelos mesmos consumidores. Uma empresa deve levar em conta o preço dos concorrentes da região competitiva. Se existem poucas empresas que não competem pelo mesmo consumidores, cada empresa é um monopólio local.

4.8.6 Modelo de Edgeworth

Uma variante do modelo de Bertrand foi proposto por Edgeworth (1925). Como Bertrand ele assumiu que as firmas apresentam variações conjecturais iguais a zero (ela mede a expectativa da firma 1 da reação da produção da firma 2 para uma mudança em sua própria produção e vice-versa), mas questionava a condição de um equilíbrio estável e de que as firmas não teriam qualquer limitação de produção. Baseou-se nas premissas de Cournot, mas incorporava a idéia de Bertrand de uma certa disputa de preços.

Tem-se as curvas de demanda simétricas para as firmas 1 e 2. Tendo em vista a limitação da capacidade produtiva das firmas, há uma produção máxima para as firmas. Se a firma 1 entrar primeiro no mercado, tem-se que esta irá ofertar uma menor quantidade do que uma situação de concorrência a um preço de monopólio (onde a elasticidade-preço da demanda é unitária). A firma 2 entra no mercado prevendo que a sua rival irá manter o seu preço fixo e, observando que o mercado está insatisfeito, procura vender toda a sua produção a um preço menor, retirando parcela do mercado da firma 1. Se 1 quiser manter o preço de monopólio irá ter prejuízo e, portanto, faz uma revisão da sua posição reduzindo o seu nível de preço e aumentando a quantidade vendida. Uma guerra de preços do tipo de Bertrand ocorre neste momento até um determinado ponto entre o preço de monopólio e o preço de concorrência perfeita, onde torna-se lucrativo para as firmas aumentar os preços em direção do preço de monopólio a baixar mais os preços.

Não existe nenhuma solução de equilíbrio no modelo de Edgeworth, pois os preços variam em torno dos níveis de monopólio por um período indefinido.

4.8.7 Modelo da Firma Dominante

Em contraste com o modelo de Stackelberg, onde as firmas seguidoras adotam uma variação conjectural de Cournot, no modelo da firma dominante as firmas seguidoras agem como perfeitos competidores. Neste modelo é assumido que as seguidoras tomam o preço de mercado dado, de modo a maximizarem o lucro igualando este preço à sua curva de custo marginal. A firma dominante, por sua vez, estabelece seu preço considerando a curva de oferta dada pelas firmas competitivas, agindo como uma monopolista parcial, ficando condicionada à oferta das seguidoras para obtenção dos lucros de monopolista da indústria.

De modo a maximizar o lucro, a firma dominante iguala a receita marginal com a sua curva de custo marginal, determinando um nível de produção a um preço. Como o esperado, a introdução de uma oferta competitiva enfraquece o poder de monopólio da firma dominante. Da mesma forma, uma expansão da oferta tende a reduzir o preço de mercado. Em geral, aumentos na participação das firmas competitivas no mercado podem servir como um indicador da extensão pela qual o preço excede o custo marginal neste modelo.

4.8.8 Modelo do Comportamento Coordenado de Chamberlin

Assumindo que as firmas em uma indústria oligopolística objetivam a maximização de lucro, existe um óbvio incentivo à cooperação entre as mesmas ao invés da competição. Tal cooperação pode ter o formato de cartel ou somente nas combinações de preços, produção, etc. Chamberlin

(1966) originalmente sugeriu que as firmas oligopolísticas podem coordenar suas ações para este fim. Especificamente, ele sugeriu que grupos de oligopólio poderiam reconhecer sua mútua interdependência e então deixar de executarem competição. Na visão de Chamberlin, geralmente a coordenação oligopolística pode variar sistematicamente com o número de firmas ou com o grau de concentração de mercado, onde alta concentração leva a um maior comportamento coordenado. Desta forma pode-se imaginar um amplo espectro de possibilidades de comportamento indo da competição à coalizão, dependendo do grau de concentração do mercado.

Seguindo Cowling & Waterson (1976), podemos considerar um mercado de n firmas produzindo um produto homogêneo. Também assume-se que cada firma maximiza lucro adotando variações conjecturais de produção arbitrárias, ou seja:

$$\lambda_i = \frac{d\sum_{j\neq i} x_j}{dx_i}. \tag{4.206}$$

Sob estas condições, Cowling e Waterson mostram que o equilíbrio de mercado implica que:

$$\frac{p - \sum_i CMg_i\left(\frac{x_i}{x}\right)}{p} = \frac{H}{\eta}(1+\mu), \tag{4.207}$$

em que H é o índice de concentração de mercado de Hirschman-Herfindahl, η é a elasticidade-preço da demanda de mercado e μ é a média ponderada dos valores de λ_i. Assim, em um modelo generalizado, diferentes conjecturas sobre as respostas na produção das firmas rivais, refletidas em μ, influenciam a média ponderada dos preços em comparação com o custo marginal na indústria. Neste caso firmas que combinam maiores preços antecipados das rivais tendem a aumentar preços e extrair maiores lucros de monopólio.

4.8.9 Modelo de Equilíbrio de Conjecturas Consistentes

Consideraremos brevemente alguns trabalhos recentes do tão chamado equilíbrio de conjecturas consistentes (ECC) como desenvolvido por Bresnahan (1981) e outros. Bresnahan enfatiza que a inconsistência dos muitos modelos de oligopólio, tais como o de Cournot, que adotam hipóteses de variação conjectural pelos oligopolistas que não são consistentes com as reações atuais das firmas no modelo. No caso de Cournot, as firmas assumem que as rivais mantém seus níveis de produção quando, no mínimo os custos são constantes, tal hipótese implica que as funções de reação atuais não são constantes. Assim, o modelo tem inconsistência interna, e isto se manifesta pelas reações não constantes das rivais em uma versão dinâmica do modelo.

Proponentes de modelos de conjecturas consistentes têm sugerido que, sendo as firmas racionais, espera-se que as firmas adotem variações conjecturais consistentes. A idéia é análoga à idéia da obtenção de solução no caso de expectativas racionais em modelos macroeconômicos. Como desenvolvido até aqui, entretanto, têm sido aplicado somente no contexto estático, e não está claro, em geral, como funcionaria em condições dinâmicas. Como os trabalhos nesta área estão em andamento, não se tem ainda grandes e fortes conclusões a respeito dos valores das conjecturas consistentes.

4.8.10 Modelo da Curva de Demanda Quebrada

Este modelo foi introduzido por Paul Sweezy em 1939, nos Estados Unidos, e por Hall e Taylor também em 1939, no Reino Unido. Em contraste aos modelos anteriores, o foco inicial da análise não está no preço de equilíbrio mas no movimento dos preços oligopolísticos. Especificamente, a teoria sugere que os preços oligopolísticos podem tender a serem rígidos, não respondendo à mudanças no custo e na demanda como se esperaria normalmente pela teoria de maximização de lucro.

Assume-se um mercado oligopolístico com produtos diferenciados. Tem-se a situação para um oligopolista típico com o corrente preço, p_0. Há a curva que representa a demanda do oligopolista na hipótese de que as rivais combinam suas variações nos preços e a curva de demanda mais elástica que seria aplicada se as rivais mantivessem o preço constante quando a firma tivesse seu preço alterado. A teoria da curva quebrada assume que as firmas oligopolistas esperam que as rivais combinem reduções de preços, mas não aumentos nos mesmos. Assim, a demanda para a firma é dada por uma çurva abaixo do preço p_0, e por uma outra curva acima deste preço, implicando uma curva de demanda quebrada ao preço corrente. Isto significa que a curva de receita marginal é descontínua a um determinado nível de produção. Dadas estas variações conjecturais, se o custo marginal corta a receita marginal em algum ponto nesta faixa de descontinuidade, então o preço e a produção são aqueles que maximizam o lucro.

A novidade desta teoria é que ela prevê que os preços oligopolísticos tenderão a serem rígidos à mudanças nos custos e nas condições de demanda. Grandes aumentos ou reduções no custo marginal podem ocorrer com a receita marginal se mantendo igual ao custo marginal a um determinando nível de produção, fazendo com que as firmas não tenham nenhum incentivo a mudarem os preços. Da mesma forma, deslocamentos na curva de demanda de mercado, enquanto elas induzem a alterações na quantidade produzida, não necessitam de alterações nos preços desde que o custo marginal continue interceptando a porção descontínua da receita marginal.

Alguma evidência a favor da hipótese da curva de demanda quebrada foi apresentada por Hall e Hitch. Seu estudo aborda o comportamento do preço baseado na interpretação de entrevistas com 38 empresários do Reino Unido. Eles encontraram que estes empresários tenderam a estabelecer o preço a cobrir os custos unitários, não variando, estes preços, devido à curva da demanda quebrada.

Em 1947, Stigler examinou a evidência empírica da hipótese da curva da demanda quebrada, além de predições da teoria usando dados de preços e produção dos Estados Unidos para 19 oligopólios e 2 monopólios no período de 1929–1937. Entre outras coisas, ele encontrou que: (i) os preços foram muito mais rígidos nestes dois monopólios (alumínio e níquel) do que nos oligopólios neste período; (ii) os preços foram mais rígidos nas indústrias com mais produtos heterogêneos, mesmo em mercados onde se esperaria uma menor quebra. Estes resultados são contrários à visão de que a curva de demanda quebrada é uma importante característica dos mercados oligopolísticos. Sendo assim, Stigler foi incapaz de encontrar alguma evidência sistemática da teoria da curva de demanda quebrada.

Também há o modelo de cartel dominante e oligopólio generalizado. Neste tópico estudamos a formação de preços em oligopólio e as possibilidades de aumentos de preços acima dos níveis competitivos. Analisamos muitos modelos e abordagens teóricas, apontando uma grande variedade de soluções possíveis para o problema do oligopólio. Pelo que foi visto podemos concluir que a formação de preços em mercados oligopolísticos é um fenômeno complexo, e que nenhuma teoria é suficiente para explicar o funcionamento de toda e qualquer indústria oligopolística. Apesar disso, analisou-se, como uma ampla generalização, a influência de fatores tais como alta concentração de mercado, condições de mercado estáveis, demandas homogêneas e condições de custos na tendência à coordenação oligopolística e aos altos preços.

4.9 Monopsônio

O monopsônio[8] se caracteriza pela existência de um único comprador para o produto embora, do outro lado, possa haver um grande número de vendedores, como normalmente acontece na agricultura, onde há muitos produtores. Por exemplo, suponha que o monopsônio é uma empresa que processa tomate e tem controle da indústria de processamento deste produto. Portanto, os produtores de tomate são forçados a negociar com este processador.

Para o monopsonista existem duas situações:

1. por um lado, a sua curva de demanda é a própria curva de demanda de mercado. No caso da firma processadora, esta curva de demanda é a curva de demanda derivada;

2. por outro lado, o monopsonista se defronta com a curva de oferta do porduto no mercado.

Vejamos separadamente estas duas situações.

1. Curva de demanda do monopsônio

   O monopsonista compra tomates e os utiliza como principal insumo de um produto processado (extrato, por exemplo). Sua função de produção caracteriza o produto processado Q como uma função da quantidade de X (tomate) utilizada.

   Portanto:

$$Q = f(X). \tag{4.208}$$

   Sua receita total é $P_Q \cdot Q = P_Q \cdot f(X)$. Deve-se, agora, definir o conceito de valor do produto marginal (VPmg), a fim de melhor entender o significado da curva de demanda do monopsonista. O VPmg , devido à utilização de uma unidade adicional de X, é igual ao produto físico marginal multiplicado pelo preço do produto:

$$VPMg = P_Q \cdot \frac{dQ}{dX} = P_Q \cdot PFMg. \tag{4.209}$$

[8] O termo "monopsônio" foi introduzido por Joan Robinson em seu clássico de 1932, "The Economics of Imperfect Competition", embora ela dê crédito ao estudioso B. L. Hallward de Cambridge pela cunhagem real do termo.

O valor do produto físico marginal (que corresponde à receita marginal que o monopsonista obterá por adquirir, digamos, uma tonelada a mais de tomate "in natura") decresce com o aumento no emprego de X, em virtude da lei dos rendimentos decrescentes, ou seja, o produto físico marginal decresce à medida que se aumenta o emprego de X, mantidos os demais recursos constantes. O preço P_Q não se altera, porque está se admitindo que o monopsonista vende o seu produto num mercado competitivo. Portanto, a sua curva de $VPMg$ é negativamente inclinada e representa a curva de demanda do monopsônio (análogo com o monopólio, quanto à curva de oferta).

2. Curva de oferta do insumo

Neste caso, a curva S, representa a função de oferta de milhares de produtores agrícolas. A curva de oferta mostra a quantidade de produto (tomate) que o monopsonista pode adquirir para cada nível de preço. Em outras palavras, esta curva de oferta positivamente inclinada mostra que o comprador deve reconhecer que quanto maior a quantidade de tomate que ele deseja adquirir, mais elevado deve ser o preço que ele terá de pagar. O custo variável total na aquisição do insumo pode ser expresso como:

$$CVT = P_X \cdot X. \tag{4.210}$$

Tendo em vista que a curva de oferta que o monopsonista se defronta é positivamente inclinada, o preço do tomate é uma função crescente da quantidade utilizada do insumos (tomate):

$$P_X = g(X). \tag{4.211}$$

É importante, agora definirmos o conceito de custo marginal do fator (CMgF), o qual é a variação no custo total devido ao emprego de uma unidade adicional do fator:

$$CMgF = \frac{dCVT}{dX} = P_X \frac{dX}{dX} + X \frac{dP_X}{dX} = P_X + X \frac{dP_X}{dX}. \tag{4.212}$$

Como $\frac{dP_X}{dX} > 0$, o custo marginal do fator excede seu preço (P_X) (se estivesse em concorrência perfeita, o $CMgF$ seria o próprio preço). Desse modo a curva de $CMgF$ se encontra acima da curva de oferta S. Isto implica que se o monopsonista desejar mais da matéria-prima, ele deve pagar um preço mais elevado não só para as unidades adicionais, mas para todas as demais unidades que ele comprar. Portanto, a inclinação positiva da curva de oferta do produto para o monopsonista é a característica básica que o distingue da firma compradora em concorrência pura.

A maximização do lucro referente ao emprego do recurso, pelo monopsonista, ocorre no nível em que o valor do produto marginal se iguala ao custo marginal do fator, ou seja:

$$VPMG = CMgF. \tag{4.213}$$

Isto significa que o monopsonista terá interesse em comprar mais tomate “in natura”, enquanto o custo marginal de, por exemplo, uma tonelada a mais de tomate for menor que o correspondente acréscimo na receita da firma. Com base nesta igualdade, determina-se a quantidade X_0 de insumo que maximiza o lucro. Esta quantidade determina o preço através da curva de oferta do produto.

Se o mercado fosse competitivo, em ambos os lados, o equilíbrio ocorreria onde o preço do insumo P_X fosse igual ao valor do produto marginal, ou seja, resultando em maior preço para o produtor e maior quantidade. Portanto, o monopsonista emprega menor quantidade de insumo e paga menor preço.

Referências Bibliográficas

[Bain1956] Bain, J. S. (1956). *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries.* Harvard University Press, Cambridge, MA.

[Belleflamme and Peitz2015] Belleflamme, P. and Peitz, M. (2015). *Industrial Organization: Markets and Strategies.* Cambridge University Press.

[Cabral2017] Cabral, L. M. (2017). *Introduction to Industrial Organization.* MIT press.

[Debreu1959] Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* Yale University Press.

[Dixit and Stiglitz1977] Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297–308.

[Jehle2001] Jehle, G. A. (2001). *Advanced Microeconomic Theory.* Pearson Education India.

[Kreps and Scheinkman1983] Kreps, D. M. and Scheinkman, J. A. (1983). Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes. *The Bell Journal of Economics*, pages 326–337.

[Milgrom and Roberts1987] Milgrom, P. and Roberts, J. (1987). Informational Asymmetries, Strategic Behavior, and Industrial Organization. *American Economic Review*, 77(2):184–193.

[Schmalensee et al.1989] Schmalensee, R., Armstrong, M., Willig, R. D., and Porter, R. H. (1989). *Handbook of Industrial Organization.* Elsevier.

[Shaked and Sutton1982] Shaked, A. and Sutton, J. (1982). Relaxing Price Competition through Product Differentiation. *The Review of Economic Studies*, pages 3–13.

[Shy1995] Shy, O. (1995). *Industrial Organization: Theory and Applications.* MIT press.

[Tirole1988] Tirole, J. (1988). *The Theory of Industrial Organization.* MIT press.

[Varian and Varian1992] Varian, H. R. and Varian, H. R. (1992). *Microeconomic Aanalysis.* Norton New York.

Capítulo 5

Teoria da Informação

Contents

5.1 Informação

Costuma-se dizer que a informação é um bem público: acesso fácil a previsões do tempo, informações de tráfego, preços de ações, etc, ajuda-nos todos os dias. Isso sugeriria que (a) quanto mais informação, melhor - afinal, no pior dos casos, podemos ignorá-lo; (b) muito pouca informação será fornecida em um equilíbrio de mercado livre, uma vez que os bens públicos são subutilizados pelo mercado. O artigo de 1973 de Michael Spence "Job Market Signaling" demonstra que, em alguns casos, pode haver muita informação em um equilíbrio de livre mercado. Por quê? Não é que divulgar informações seja per se prejudicial. Mas o valor social da informação divulgada pode não valer o custo de transmiti-la. E ainda, as pessoas podem ter um incentivo para produzir e divulgar informações que geram benefícios privados não-zero e ainda geram valor social líquido zero (ou negativo). No modelo de sinalização no artigo de Spence, de 1973, os incentivos para a divulgação ou não divulgação são puramente privados, e esses incentivos privados podem ou não gerar resultados desejáveis medidos em termos de eficiência social.

Os economistas historicamente conjecturaram que os mercados de informação eram bem comportados, assim como os mercados de outros bens e serviços. Poder-se-ia decidir da melhor maneira a quantidade de informação a ser comprada e, assim, equacionar os retornos marginais às compras de informação com os retornos marginais a todos os outros bens.

Na década de 1970, os economistas tiveram de reavaliar essa crença por uma série de artigos de George Akerlof, Micheal Rothschild, Joseph Stiglitz e Michael Spence. Muitos (não todos) desses economistas passaram a dividir o Nobel de 2001 por seu trabalho na economia da informação.

A informação não é um bem padrão de mercado. Não é rival, já que não há custo marginal para uma pessoa adicional. É extremamente durável, pois não desaparece quando é consumido. Não é um bem típico que você pode "experimentar antes de comprar". Um vendedor não pode permitir que você "prove" informações sem realmente fornecer as informações. Ao contrário de outros bens (ou seus atributos), a informação é extremamente difícil de medir, observar e verificar.

Outra propriedade importante da informação é que ela pode ser assimétrica. Alguns agentes em um mercado podem estar mais bem informados do que outros sobre os atributos de um produto ou transação. Esse recurso de informação é diferente da maioria das outras propriedades de bens e serviços. Nas trocas convencionais, não há incerteza sobre a natureza do bem ou serviço em questão; as únicas características relevantes para o mercado são quantidade e preço. Onde a informação está envolvida, no entanto, diferentes partes de uma transação podem não ter as mesmas informações sobre os atributos da mercadoria. Essa assimetria de informação (e o conhecimento dessa assimetria) frequentemente afetará o preço e a quantidade em que as partes estão dispostas a negociar.

Os economistas frequentemente usam os termos informação assimétrica e informações privadas de maneira intercambiável. Ambos os termos significam que um ator tem informações diferentes - geralmente mais - do que outro ator, que ambos os atores tipicamente reconhecem essa assimetria, e que eles podem se comportar estrategicamente como consequência.

A maneira mais natural (e certamente onipresente) em que isso ocorre é que os compradores podem ter informações gerais sobre as características médias de um produto que desejam comprar, ao passo que os vendedores terão informações específicas sobre os produtos individuais que estão vendendo.

Quando os compradores e vendedores têm informações assimétricas sobre transações de mercado, os negócios que realmente ocorrem provavelmente serão um subconjunto dos negócios viáveis e que melhoram o bem-estar. Os vendedores querem vender certos itens cujos atributos eles conhecem e os compradores podem ser cautelosos se entenderem os incentivos do vendedor. Muitos negócios que ocorreriam voluntariamente se todas as partes tivessem informações completas não ocorreriam.

Os modelos econômicos de informação concentram-se no ambiente da informação - isto é, quem sabe o que quando. A especificação cuidadosa desses recursos no modelo é fundamental para entender o que segue.

As próximas seções irão mergulhar em múltiplos fenômenos que surgem quando a informação é assimétrica. Esta nota de aula discute nosso primeiro insight principal da literatura sobre informação assimétrica: é possível que os agentes se envolvam em sinalização socialmente improdutiva em um equilíbrio de livre mercado.

A seleção adversa é um conceito econômico que frequentemente aparece na literatura para o gerenciamento de seguros e riscos. A ideia principal é que os compradores e vendedores tenham informações diferentes ou assimétricas sobre bens, instrumentos financeiros ou produtos, e aqueles comerciantes com melhores informações sobre a qualidade do produto participarão seletivamente de negócios que mais os beneficiam - às custas de suas contrapartes. O conceito também existe em um sentido mais geral de informação imperfeita entre as partes em qualquer troca e em qualquer ação humana empreendida. Vários laureados do Nobel publicaram trabalhos relacionados à informação assimétrica, incluindo Friedrich A. von Hayek, George Akerlof, Douglass North e Michael Spence.

Compradores com baixa preferência pelo tempo podem aproveitar as vendas de supermercados indo de loja em loja na cidade e comprando virtualmente apenas esses itens de venda em cada um deles - mesmo que leve muitas horas para fazê-lo. Os supermercados, é claro, geralmente não conhecem as preferências pelo tempo dos clientes e dificilmente podem excluí-los da compra de itens de venda. Assim, nos cálculos do negócio, eles devem considerar as perdas incorridas por esses clientes. Outra empresa que pode perder negócios na preferência do tempo são as casas de câmbio, especialmente se elas estiverem próximas à concorrência. Pessoas com o tempo estarão dispostas a andar de loja em loja até encontrarem os melhores preços. Consumidores de combustível de baixo preço fazem a mesma coisa.

O mesmo se aplica às churrascarias que oferecem um bife gratuito de quatro quilos, batata e salada para aqueles que podem limpar os pratos (caso contrário, pagam caro pela refeição). O restaurante espera ganhar mais dinheiro naqueles que tentam e fracassam do que aqueles que se elevam ao feito. Buffets livres enfrentam assimetrias similares, assim como as seguradoras de vida e de saúde, já que pessoas insalubres provavelmente terão mais probabilidade de comprar esses produtos. Cada indústria tem seus próprios métodos testados e comprovados para evitar ser "pega" por clientes "ruins".

Por exemplo, o seguro de vida é a maior categoria de seguro vendido nos Estados Unidos, e pesquisas mostram que as empresas têm tido sucesso em superar assimetrias de informação (Cawley & Philipson, 1999), assim como seguradoras de automóveis na França (Chiappori & Salanie 2000)) e seguradoras de saúde em grupo nos Estados Unidos (Cardon & Hendel 2001; Cutler & Zeckhauser

1998). No entanto, o mesmo não é verdade para o seguro de saúde no Reino Unido (Olivella & Vera-Hernández 2013), seguro de assistência de longo prazo nos Estados Unidos (Finkelstein & McGarry 2006), certas anuidades (Finkelstein & Poterba 2004) ou seguro residencial na Flórida (Dionne, Gouriéroux & Vanasse 2001).

Muitos outros estudos no setor de seguros têm sido feitos para sugerir que as assimetrias informacionais não são um grande problema no setor de seguros (de Meza & Webb 2001), e às vezes tendem a ser mitigadas pela "seleção propícia" (Hemenway 1990). De fato, alguns estudiosos concluíram que a presença de pequenas doses de informação assimétrica é realmente benéfica para o mercado de seguros (Thomas, 2008) - gerando até mesmo um "efeito de melhoria da informação privada" como uma "consequência do fato de que o comportamento dos vendedores está em linha com a teoria da seleção adversa" (Hoppe & Schmitz 2015).

Essa teoria está ligada à crítica mais ampla e devastadora do planejamento central, que não pode funcionar simplesmente porque nenhuma mente ou comitê único possui a informação social ou econômica "fragmentada" e "dispersa" necessária para planejar corretamente (Hayek, 1945; Sowell, 1996; Holcombe, 1995; Cobin, 2009). Alguns têm sido levados a considerar a informação assimétrica como uma "falha de mercado". O princípio básico dessa premissa é que, como os mercados não informam perfeitamente compradores e vendedores sobre qualquer transação ou bem negociado, eles falham.

O exemplo clássico de informação assimétrica foi apresentado no artigo "Market for Lemons" de Akerlof (1970), que, curiosamente, foi rejeitado por três grandes revistas antes de finalmente ser publicado. Ele argumenta que a qualidade das mercadorias pode declinar sob condições de assimetria de informação entre compradores e vendedores, de tal modo que apenas "limões" permanecem, por exemplo, um carro usado ou outro bem que é considerado defeituoso depois de ter sido comprado. Por um lado, os compradores informados de forma imperfeita muitas vezes não conseguem distinguir entre um carro usado de alta qualidade ou outro bem e um "limão", pois não podem conhecer facilmente a história do bem usado em questão. Sabendo deste fato, eles escolhem pagar um preço por um carro que é fixo entre um bem e um limão.

Por outro lado, os vendedores sabem com frequência se o produto à venda é um limão, e só o venderá aos compradores com um preço intermediário, nunca oferecendo produtos de maior qualidade com esse desconto. Por exemplo, o dono de um bom carro usado cuidadosamente mantido, nunca abusado, será incapaz de obter um preço alto o suficiente para fazer a venda desse carro valer a pena. No entanto, o proprietário do carro usado de qualidade oposta terá prazer em fazê-lo.

Eventualmente, à medida que vendedores não-limão suficientes deixam o mercado, a disposição média a pagar dos compradores também diminuirá, já que a qualidade média dos produtos restantes no mercado diminuiu, de tal forma que a qualidade do mercado continua sua espiral descendente. Como resultado, a estratégia de preços de compradores desinformados cria um problema de seleção adversa que impulsiona produtos de alta qualidade do mercado, o que talvez resulte em um eventual colapso. O resultado é que, em um mercado com informações assimétricas com relação à qualidade, terá características semelhantes àquelas descritas pela Lei de Gresham: o mal expulsa o bem. Para resolver esse problema, as empresas se adaptam variando os preços de acordo com diferentes qualidades (Phlips, 1983).

O problema do conhecimento fornece uma nova crítica da economia do bem-estar. De fato, "a soma total do conhecimento disponível em uma economia" nunca existe na forma concentrada ou integrada, mas apenas como os fragmentos dispersos de conhecimento incompleto e frequentemente contraditório que todos os indivíduos separados possuem. Informação imperfeita - a mesma coisa presumida em modelos estáticos - é precisamente o objeto que permite o lucro do empreendedor. "Um estado de desequilíbrio do mercado é caracterizado por uma ignorância generalizada. Os participantes do mercado desconhecem as oportunidades reais de troca benéfica que estão disponíveis no mercado. O resultado desse estado de ignorância é que inúmeras oportunidades são perdidas" (Kirzner, 1973). Então, as assimetrias informacionais são realmente "falhas"?

A ignorância do consumidor certamente os leva a fazer escolhas subótimas às vezes. O mercado torna-se o culpado por produzir resultados socialmente inaceitáveis, particularmente quando uma empresa ou pessoa engana outra a sua mágoa (por exemplo, comprar produtos defeituosos, ou componentes potencialmente perigosos conhecidos apenas pelo vendedor). Da mesma forma, terceiros podem sofrer danos devido a falhas de produção geradas pelas ações de terceiros.

Para evitar essas armadilhas, os consumidores geralmente exigem serviços de seguro, classificação, certificação e outros serviços informativos para garantir - ou pelo menos assegurar - a qualidade dos bens adquiridos, ou para minimizar os riscos associados a uma transação. A teoria da falha de mercado sustenta que a existência de assimetrias de informação implica uma falha de mercado que deve ser parcial ou totalmente aliviada por políticas públicas proativas ou políticas públicas de fornecimento ineficiente de bens e serviços normalmente fornecidos pelos mercados (por exemplo, correios, mecânica de automóveis).

No entanto, há um crescente corpo de evidências (Cobin, 2014a; Holcombe, 1995) de que empresas especializadas em serviços de informação se desenvolverão espontaneamente nos mercados. Eles vão garantir a disciplina no mercado", que incentiva empresas e indivíduos a construir e manter reputações íntegras.

Juntamente com os fortes direitos de propriedade, a instituição legal de contratação, que define as "regras do jogo" para os agentes econômicos (North 1992), facilita a troca e, quando bem feita, reduz drasticamente os custos de transação. No entanto, nenhum contrato elimina totalmente os problemas de assimetria informacional, embora os contratos tenham melhorado ao longo do tempo por meio de tentativa e erro legais. "Na realidade, é geralmente impossível estabelecer as obrigações de cada parte de forma completa e inequívoca com antecedência, e assim a maioria dos contratos efetivos está seriamente incompleta" (Hart & Holmstrom 1987: 112).

A literatura sobre contratos sugere que as formas contratuais têm grandes efeitos sobre o comportamento. Incentivos importam. Curiosamente, o mercado está cheio de "contratos bastante simples" (Chiappori & Salanie 2002: 34), que deixam em aberto o potencial para maiores custos de transação no futuro. Evidentemente, os custos esperados de tais ocorrências são menores do que os benefícios esperados resultantes da formação de melhores contratos. O campo do direito e da economia tem como objetivo central incorporar conceitos econômicos como assimetria informacional, risco moral e adaptação às mudanças de circunstâncias, na interpretação e análise de contratos e disputas, bem como desenvolver uma doutrina contratual (Goldberg 2012).

Alguns dos fascinantes tópicos de enquadramento de contrato estudados pelos economistas incluem triagem, sinalização e risco moral. Além disso, contratação multilateral, contratação de longo prazo e comércio bilateral com informação privada (ação oculta), teoria de leilão, mais a organização interna de empresas, contratos incompletos, teoria da propriedade e controle, e lidar com externalidades (Bolton & Dewatripont 2005).

Os vendedores sujeitos à seleção adversa tentarão proteger-se rastreando clientes ou identificando "sinais" de qualidade credível. Na teoria do contrato, a sinalização significa que um agente transmite, com credibilidade, informações sobre si mesmo a um principal. Por exemplo, empregados em potencial enviam sinais a potenciais empregadores sobre seus níveis de habilidade obtendo credenciais educacionais (Spence, 1973; Weiss, 1995; Hungerford e Solon, 1987). O valor de fazê-lo reside no fato de que os empregadores acreditam que essa educação está positivamente correlacionada com uma maior capacidade. Assim, ele permite que os empregadores diferenciem, de maneira confiável, trabalhadores com habilidades mais baixas de trabalhadores com maior capacidade.

A sinalização floresceu ao lado da teoria da informação assimétrica nas transações econômicas. As desigualdades de acesso à informação distorcem a troca de mercado "normal". No entanto, as partes envolvidas no comércio podem contornar os problemas de assimetria se uma parte enviar um sinal que revele um pouco da informação relevante para a outra (Spence 1973). Essa parte interpreta o sinal e ajusta as decisões de compra de acordo - talvez aumentando o preço pré-sinal e, assim, resultando em muitos problemas auxiliares.

A quantidade de tempo, energia e dinheiro que o remetente (ou agente) deve gastar para enviar um sinal dependerá de quanto o receptor (ou principal, geralmente o comprador) pode confiar no sinal para ser uma revelação honesta da informação. Em muitos casos, as condições de equilíbrio surgirão até que a confiança seja quebrada.

Mesmo nos mercados on-line, onde a informação é bastante assimétrica, especialmente para veículos, os mercados se desenvolveram para ajudar os consumidores a confiar nas informações que veem na internet e dar aos vendedores a confiança de que sua venda será realizada. Na verdade, "os sinais de informação (descrições de produtos de diagnóstico e garantias de produtos de terceiros) reduzem a incerteza do produto" (Dimoka, Hong & Pavlou 2012). Além disso, "o volume de comércio nesses mercados mostrou-se tão robusto para a seleção adversa" (Lewis, 2011, p. 1535). Além disso, os sites da Internet que não são especializados em produtos de marca podem atrair clientes com sucesso, aumentando a reputação através da participação em "sites de comparação de preços" (Waldfogel & Chen 2006). Assim, à medida que a tecnologia aumenta, o mesmo acontece com o impacto da sinalização.

Em economia, um risco moral ocorre quando uma pessoa assume um risco maior (por exemplo, não tranca as portas) porque outra pessoa (geralmente uma seguradora) arca com o custo do risco aumentado. Situações de risco moral são mais freqüentemente vistas após as transações financeiras ou de seguros serem concluídas, em que as ações de uma parte mudam em detrimento de outra (por exemplo, uma reivindicação deve ser paga).

O risco moral surge especialmente quando ocorrem assimetrias de informação entre compradores e vendedores, onde a parte prestes a assumir riscos de uma transação sabe mais sobre suas intenções do que a parte que paga as consequências de seu comportamento. Em tais situações, a parte com mais informações sobre as ações pretendidas tem um incentivo para se comportar de forma inadequada do ponto de vista da parte com menos informações.

A teoria do risco moral implica que pagamentos de seguro social, seguro-desemprego e compensação dos trabalhadores não devem ser muito generosos, uma vez que os beneficiários tenderão a abusar do sistema e permanecerão desempregados em vez de trabalhar (Dewan, 2012). Eles também podem trabalhar em mercados mais informais para complementar os folhetos ou pagamentos de "seguro" que recebem. O risco moral tem seus limites.

O mesmo deve acontecer com o seguro de incêndio. As pessoas provavelmente não arriscariam suas vidas "não se preocupando muito" se a casa deles queimar, já que é segurada. É mais provável que as casas se queimem devido a problemas auxiliares decorrentes da busca de aluguel e captura regulatória que tornem os riscos maiores ou materiais de construção menos seguros ou menos eficientemente instalados, do que se desgastarem por negligência. Além da política, a lógica sugere que o risco moral deve ter seus limites naturais.

Comprar um seguro de vida antes de cometer suicídio, a fim de beneficiar seus entes queridos, é um exemplo de tal comportamento. Por essa razão, os contratos de seguro de vida geralmente têm uma cláusula contratual que exclui pagamentos de pedidos de morte resultantes de suicídio por dois anos após a emissão da apólice. O mesmo pode ser dito em relação a condições preexistentes e cláusulas de exclusão em seguros de saúde e invalidez, levando alguns a criticar o uso da teoria do risco moral na prática do setor de seguros (Baker, 1996: 291-292).

O risco moral também surge dentro do problema de principal-agente, em que o agente (ou gerente) age em nome de outro (o principal). O agente geralmente tem maior informação sobre suas ações e intenções do que o principal, já que é difícil para ele monitorar o agente (Laffont & Martimort, 2002). Consequentemente, o agente pode ter um incentivo para agir de forma inadequada (do ponto de vista do principal) se os interesses de ambos não estiverem alinhados. Isso é especialmente verdadeiro entre acionistas e administradores, onde os últimos tendem a ser avessos ao risco e fazem o que é necessário para manter seus cargos com salários mais altos, em vez de assumir maiores riscos para a empresa que tenderia a aumentar os retornos dos acionistas. Supõe-se também que os médicos tenham uma vantagem sobre os pacientes, que só podem com grande dificuldade determinar seu nível de conhecimento e habilidade (Arrow, 1963 e 1971).

"O risco moral descreve uma situação em que uma parte é isolada das consequências de suas ações. Assim protegida, não tem incentivo para se comportar de maneira diferente" (Ahrens, 2008). O mesmo ocorre quando empresas financeiras ou de montadoras (e seus acionistas) sabem que o governo irá socorrê-la se enfrentarem consequências irrecuperáveis devido a decisões arriscadas. O governo é, na verdade, uma seguradora. No entanto, a existência de risco moral pode realmente melhorar as disposições contratuais e, assim, reduzir os custos de incerteza e transações (Holmstrom, 1979: 89-90).

Por que é tão difícil conseguir um bom preço em um carro usado? Por que as empresas oferecem prêmios mais baixos por unidade de cobertura para segurados que aceitam franquias? Por que as empresas continuam pagando dividendos, embora os acionistas estejam sujeitos à dupla tributação? Em que condições um monopólio pode impedir a entrada lucrativamente estabelecendo um "preço limite" baixo em vez do preço de monopólio? Estas são apenas algumas das inúmeras questões que a teoria econômica não estava em condições de responder até o desenvolvimento de novos fundamentos teóricos.

Com o benefício da retrospectiva, houve quatro trabalhos pioneiros que prepararam o terreno para um esforço de pesquisa sem precedentes que continua até hoje. O notável artigo de William Vickrey (1961) examinou uma série de questões na provisão de incentivos quando os agentes têm informações privadas. No apêndice, Vickrey até forneceu uma base para a moderna teoria dos leilões. Dez anos mais tarde, James Mirrlees (1971), em sua análise da tributação ótima da renda, forneceu insights sutis sobre o tradeoff entre eficiência (o incentivo ao trabalho) e a redistribuição. Na mesma época, George Akerlof (1970) mostrou como o comércio pode quase entrar em colapso quando os agentes de um lado de um mercado conhecem apenas a distribuição da qualidade do produto, e não a qualidade de cada item negociado. Finalmente, Michael Spence (1973) perguntou se, em um mercado competitivo, os vendedores de produtos de qualidade acima da média poderiam "sinalizar" esse fato, tomando algumas medidas dispendiosas. Do outro lado do mercado, os compradores desinformados poderiam usar a ação dispendiosa como uma maneira de "rastrear" (screen) a qualidade?

Na terminologia de hoje, os artigos de Vickrey e Mirrlees centraram-se no desenho de um esquema de incentivos por um monopolista mal informado. Apesar de ter uma desvantagem informacional, é tipicamente o caso de o monopolista ter um incentivo para oferecer um conjunto de alternativas que (pelo menos parcialmente) separam agentes com características diferentes. No caso de Mirrlees, trabalhadores mais capazes optam por ganhar uma renda maior, mesmo sabendo que pagarão impostos mais altos. E em um leilão de Vickrey, compradores com avaliações maiores têm um incentivo para fazer lances mais altos. Akerlof e Spence examinaram modelos com uma estrutura formal muito semelhante. Agentes mal informados fazem ofertas, levando em conta a heterogeneidade do outro lado do mercado. No entanto, existe uma diferença fundamental. Em vez de um único agente desinformado, há agora competição entre esses agentes. Assim, enquanto Vickrey e Mirrlees estudam os esquemas ideais de incentivo, Akerlof e Spence examinam esquemas de incentivo que sobreviverão às forças competitivas do mercado, isto é, esquemas de incentivo de equilíbrio.

Central para a teoria do equilíbrio tradicional é a ideia de que uma economia guiada pelos preços economiza na informação. Em um mundo de bens privados, os agentes individuais não precisam saber nada sobre os outros agentes no mercado e, no entanto, os preços de equilíbrio walrasianos resultam em uma alocação Pareto eficiente. Crítico para o funcionamento ideal da mão invisível, no entanto, é a exigência de que os agentes tenham as mesmas informações sobre as características das mercadorias que estão sendo negociadas. Quando essa suposição falha, como aproveitar ao máximo os ganhos potenciais para o comércio se torna uma questão muito mais sutil. Akerlof (1970) fornece o primeiro modelo formal, ilustrando como a informação dramaticamente assimétrica pode afetar as negociações de equilíbrio. Considere uma população de proprietários de carros, cada um dos quais deve escolher entre comprar um carro novo ou permanecer com o antigo. Enquanto os carros usados se parecem com os compradores em potencial, a qualidade real varia. Apenas o atual proprietário conhece a qualidade real do seu carro. A partir da experiência de períodos anteriores, os compradores antecipam corretamente a qualidade média dos carros usados que são negociados. O preço de mercado de um carro usado reflete, assim, a qualidade média de um carro no mercado. Se as diferenças de qualidade são grandes, aqueles com carros de qualidade suficientemente alta acham melhor manter seus carros antigos em vez de vender, diminuindo assim a qualidade média dos carros que estão sendo negociados. Como Akerlof enfatizou, esse incentivo para se retirar do mercado pode levar a um equilíbrio no qual ninguém, a não ser os piores carros (os “limões”) são negociados.

5.2 Falhas de Mercado e Sinalização

Em muitas situações econômicas há incerteza sobre a qualidade dos produtos ou outras características importantes. Frequentemente, tal situação é caracterizada por informações ocultas: um indivíduo ou um grupo de indivíduos possui informações privadas, ou seja, está melhor informado sobre aspectos economicamente relevantes do que outros. Exemplos são:

1. Em um mercado, geralmente, a parte vendedora tem melhores informações sobre o produto do que a parte compradora.
2. Um banco que decide se concede ou não um empréstimo comercial a um cliente sabe menos do que o cliente sobre o negócio do cliente. Em particular, isso significa que o cliente tem uma ideia melhor do que o banco sobre o risco do empréstimo, tendo informações mais precisas sobre a probabilidade de o negócio falir e o empréstimo não ser (totalmente) recuperável.
3. Uma companhia de seguros que oferece apólices normalmente sabe menos sobre as habilidades e hábitos de direção de seus clientes potenciais do que eles.
4. Os trabalhadores que se candidatam a um emprego normalmente têm uma ideia mais clara de suas habilidades e hábitos de trabalho do que os possíveis empregadores.

No mundo real observamos muitos mercados nos quais os preços de bens usados são “muito baixos”. Embora seja fácil atribuir o fenômeno ao comportamento irracional dos participantes do mercado, é difícil explicar em bases racionais. Assim, a pergunta de pesquisa que Akerlof pergunta em seu artigo é: por que os mercados em presença de incerteza qualitativa exibem preços de equilíbrio baixos? Uma breve resposta da Akerlof é que os vendedores estão mais bem informados sobre a qualidade e se comportam estrategicamente em escolhas de qualidade, e que a interação entre a existência de informações assimétricas e os compradores com expectativas racionais leva ao fracasso do mercado.

Essas assimetrias de informação podem levar a um problema chamado seleção adversa, que pode fazer com que os mercados não funcionem como deveriam, apesar dos ganhos potenciais do comércio. Usaremos o famoso exemplo de Akerlof (1970) do mercado de carros usados para ilustrar o problema da seleção adversa. Suponha que existam apenas dois tipos de carros, carros bons e carros ruins (“limões”), onde a fração de todos os carros que são limões é $\pi = 1/2$. Suponha que compradores e vendedores nesse mercado sejam neutros ao risco e que haja mais compradores em potencial do que carros usados. Um comprador estaria disposto a comprar um bom carro por \$10.000, mas pagaria apenas \$5.000 por um limão. Suponha que um vendedor se desfizesse de um bom carro por \$9.000 e um limão por \$3.000.

Primeiro, considere a situação em que ninguém sabe se um carro é um limão ou não. Então, a disposição do comprador em pagar por um carro de qualidade média é

$$\mathbb{E} = \frac{1}{2} \cdot 5000 + \frac{1}{2} \cdot 10000 = 7500. \tag{5.1}$$

O preço de reserva de um vendedor é o valor que ele ou ela atribui a um carro de qualidade média,

$$\mathbb{E} = \frac{1}{2} \cdot 3000 + \frac{1}{2} \cdot 9000 = 6000. \tag{5.2}$$

Como há uma competição de Bertrand entre os compradores de carros usados, o preço de mercado é de \$7.500 e todos os vendedores comercializam seus carros.

Agora, considere a situação em que o vendedor conhece a qualidade de seu carro, enquanto os compradores não conseguem distinguir a qualidade quando compram um carro usado. Assim, o comprador deve perguntar a si mesmo: "Por que essa pessoa está disposta a vender o carro para mim?" Suponha que um carro fosse oferecido pelo preço que o comprador estivesse disposto a pagar por um carro de qualidade média, \$7.500. Todos os vendedores com limão venderiam a esse preço. No entanto, nenhum vendedor com um bom carro estaria disposto a vender uma vez que eles gostariam de cobrar pelo menos \$9.000. Assim, o comércio de carros usados só pode ocorrer a um preço de \$5000 e apenas limões são negociados. Isto leva a uma perda de bem-estar, uma vez que os ganhos potenciais do comércio de \$1.000 por um bom carro usado não são realizados. Akerlof argumenta que a seleção adversa pode explicar por que há uma diferença maior nos preços entre carros novos e usados do que poderia ser justificado pelo declínio no valor de uso devido à idade.

Com essas ideias em mente, vamos nos voltar agora para um modelo que coloca um pouco mais de estrutura no problema da seleção adversa. Como veremos, com mais níveis de qualidade, a seleção adversa pode levar a um colapso parcial dos mercados, apesar dos ganhos potenciais do comércio.

5.2.1 Modelo de Akerlof (1970)

Considere um mercado com m vendedores e $n > m$ compradores. Suponha que cada vendedor tenha uma unidade de um bem. A qualidade dos produtos dos vendedores é distribuída de acordo com a distribuição F no intervalo $[\underline{q}, \overline{q}]$, $0 < \underline{q} < \overline{q}$. Suponha que F segue uma distribuição uniforme, ou seja,

$$F(q) = \begin{cases} 0 & \text{se } q < \underline{q} \\ \dfrac{q - \underline{q}}{\overline{q} - \underline{q}} & \text{se } \underline{q} < q < \overline{q} \\ 1 & \text{se } q > \overline{q}, \end{cases} \tag{5.3}$$

e a densidade é

$$f(q) = \begin{cases} 0 & \text{se } q < \underline{q} \\ \dfrac{1}{\overline{q} - \underline{q}} & \text{se } \underline{q} < q < \overline{q} \\ 0 & \text{se } q > \overline{q}. \end{cases} \tag{5.4}$$

Os vendedores maximizam a utilidade esperada com a função de utilidade de Bernoulli

$$v(q, t) = \alpha q x \quad 0 < \alpha < 1, \tag{5.5}$$

em que $x \in \{0, 1\}$ é o valor do bem possuído, q é a qualidade do bem e t é a quantidade de dinheiro que o vendedor tem. Ou seja, um vendedor que possui uma bem de qualidade q tem um preço de reserva de αq.

Cada comprador está interessado em comprar no máximo uma unidade do bem. Os compradores maximizam a utilidade esperada com a função de utilidade de Bernoulli

$$u(q, t) = qx + t. \tag{5.6}$$

Ou seja, cada comprador está disposto a pagar até q para uma adquirir um bem de qualidade q. Assumimos que os compradores não têm restrições de crédito para que possam pagar qualquer preço p para adquirir o bem, diminuindo assim sua utilidade por p.

5.2.1.1 Informação Completa

Começamos analisando o caso de benchmark em que compradores e vendedores podem observar a qualidade dos produtos (informações completas). Cada nível de qualidade q é visível para os compradores e, portanto, pode ser visto como um mercado separado. Considere um vendedor que possui uma unidade do bem de qualidade q. Ele valoriza o bem menos que os compradores uma vez que $\alpha q < q$. Assim, há um ganho potencial do comércio de $(1 - \alpha)q$.

Vamos primeiro esclarecer o que queremos dizer quando falamos de um equilíbrio neste contexto.

Definição 5.2.1 (Equilíbrio de mercado sob informação completa)**.** *Em um equilíbrio de mercado as seguintes condições devem ser satisfeitas:*

1. *Todos os participantes do mercado maximizam sua utilidade, assumindo os preços como determinados.*
2. *Condição de market-clearing:*

$$X(p^*, q) = X^s(p^*, q) = X^d(p^*, q). \tag{5.7}$$

A primeira condição afirma que todos os participantes do mercado são racionais, além de auto-interessados, e operam sob a suposição de que seu comportamento não afeta os preços. A segunda condição é uma suposição padrão de compensação de mercado: o número de bens negociados no mercado para o nível de qualidade q ao preço de equilíbrio p^*, $X(p^*, q)$, deve igualar a demanda por mercadorias a esse preço, $X^d(p^*, q)$, com o fornecimento de bens a esse preço, $X^s(p^*, q)$.

Como há mais compradores em potencial do que mercadorias ($n > m$), cada vendedor pode extrair o excedente completo cobrando um preço de q para o bem. Os compradores estão competindo entre si pelo bem, já que nem todas as demandas dos compradores podem ser atendidas por outros vendedores. Assim, em equilíbrio, o mercado irá ajustar cada nível de qualidade q a um preço $p^*(q) = q$, de tal forma que:

$$\int_{\underline{q}}^{\overline{q}} X(p^*(q), q) f(q) dq = m. \tag{5.8}$$

Note que a alocação resultante é Pareto-ótima, já que todos os ganhos do comércio são realizados, e os vendedores estão estritamente melhores do que antes, enquanto os compradores não estão em situação pior do que se nenhum negócio tivesse ocorrido. Isso também decorre do Primeiro Teorema Fundamental do Bem-Estar, que afirma que um equilíbrio competitivo é necessariamente Pareto ótimo se os mercados forem perfeitamente competitivos e se houver informação completa. O excedente criado pelo comércio sob informação completa (S^{CI})

$$S^{CI} = \int_{\underline{q}}^{\overline{q}} m(1-\alpha) q f(q) dq = m(1-\alpha)\mathbb{E}[q] = m(1-\alpha)\frac{\overline{q} + \underline{q}}{2}. \tag{5.9}$$

Proposição 5.2.1. *Sob informação completa, em um equilíbrio de mercado*

1. *todos os bens são comercializados*
2. $p^*(q) = q$
3. $S^{CI} = m(1-\alpha)\dfrac{\overline{q} + \underline{q}}{2}$

5.2.1.2 Informação Incompleta e Simétrica

Agora, considere a situação em que compradores e vendedores não conhecem a qualidade de um determinado bem, informação incompleta, a ser negociado. Uma vez que a qualidade dos bens não é distinguível por qualquer participante do mercado no momento do comércio, eles são substitutos perfeitos e haverá apenas um preço para todos os bens comercializados no mercado (ou seja, há apenas um mercado em vez de $q \in [\underline{q}, \overline{q}]$ mercados sob informação completa). A qualidade esperada de um bem de propriedade de um vendedor é

$$\begin{aligned}
\mathbb{E}[q] &= \int_{\underline{q}}^{\overline{q}} q f(q) dq \\
&= \int_{\underline{q}}^{\overline{q}} \frac{q}{\overline{q} - \underline{q}} dq \\
&= \left. \frac{q^2}{2(\overline{q} - \underline{q})} \right|_{\underline{q}}^{\overline{q}} \\
&= \frac{\overline{q}^2 - \underline{q}^2}{2(\overline{q} - \underline{q})} \\
&= \frac{(\overline{q} - \underline{q})(\overline{q} + \underline{q})}{2(\overline{q} - \underline{q})} \\
&= \frac{\overline{q} + \underline{q}}{2}.
\end{aligned} \tag{5.10}$$

Vamos assumir que o preço de reserva dos vendedores é $\alpha \mathbb{E}[q] = \alpha \dfrac{\overline{q} + \underline{q}}{2}$.

Enquanto o preço de mercado está aquém do preço de reserva do vendedor, $\alpha \mathbb{E}[q]$, ninguém estará disposto a vender. Em contraste, todos os vendedores terão prazer em vender se o preço exceder o preço de reserva, levando a fornecer m. Se o preço de mercado for igual ao preço de reserva do vendedor, eles serão indiferentes entre vender e não vender, de modo que a oferta será qualquer valor entre $[0, m]$. Assim, a oferta de bens é

$$X^s(p) = \begin{cases} 0 & \text{se } p < \alpha \mathbb{E}[q] \\ [0, m] & \text{se } p = \alpha \mathbb{E}[q] \\ m & \text{se } p > \alpha \mathbb{E}[q]. \end{cases} \tag{5.11}$$

Qual é a qualidade esperada dos bens ofertados? Como a informação é simétrica, a qualidade esperada das mercadorias ofertadas ao preço p, denotada por $Q(p)$, é igual à qualidade esperada de todos os bens possuídos pelos vendedores, $\mathbb{E}[q]$ para qualquer preço p em que os vendedores estejam dispostos vender. Se nenhuma mercadoria for vendida, a qualidade não é definida, pois não é possível calcular o valor esperado de "nada":

$$Q(p) = \begin{cases} 0 & \text{se } p < \alpha \mathbb{E}[q] \\ \mathbb{E}[q] & \text{se } p \geq \alpha \mathbb{E}[q]. \end{cases} \tag{5.12}$$

Os compradores não são capazes de discriminar entre os produtos oferecidos pelos vendedores. No entanto, eles sabem que os vendedores também não conseguem distinguir a qualidade dos bens.

Além disso, os compradores conhecem a distribuição de qualidades entre os vendedores, F, e estão cientes das preferências dos vendedores. Com base nessas informações, os compradores formam crenças sobre a qualidade esperada dos produtos vendidos no mercado ao preço p, que denotamos por $Q^e(p)$. Se $Q^e(p) > p$ todos os n compradores desejarem comprar uma unidade do bem, uma vez que cada um estaria disposto a pagar até $Q^e(p)$ por isso. Assim, o montante exigido a esses preços é $X^d(p) = n$. Em contraste, se $Q^e(p) < p$ nenhum comprador quer comprar um bem, levando a $X^d(p) = 0$. Finalmente, se $Q^e(p) = p$ os compradores são indiferentes entre comprar o bem ou não de modo que qualquer nível de demanda $X^d(p) \in [0, n]$ é consistente com o comportamento de maximização pelos compradores. Para resumir,

$$X^d(p) = \begin{cases} n & \text{se } Q^e(p) > p \\ [0, n] & \text{se } Q^e(p) = p \\ 0 & \text{se } Q^e(p) < p. \end{cases} \tag{5.13}$$

Precisamos enriquecer nosso conceito de equilíbrio para acomodar a incompletude da informação. Esta definição ignora algumas tecnicalidades. Para nossos propósitos, essa definição será suficiente. Uma definição rigorosa de equilíbrio requer uma base teórica de jogos. O modelo que analisamos pertence à classe de jogos de informação incompleta à qual normalmente se aplica o conceito de Equilíbrio Bayesiano. Em equilíbrio, as ações de cada indivíduo são as melhores respostas às ações de equilíbrio dos outros atores (ou seja, o equilíbrio é um equilíbrio de Nash). Como a informação é incompleta, o conceito também exige que as crenças dos jogadores sobre os aspectos do jogo para os quais a informação está incompleta sejam logicamente consistentes (o que significa que os jogadores são bons estatísticos que podem aplicar a regra de Bayes).

Proposição 5.2.2. *Sob informação incompleta, em um equilíbrio de mercado*

1. *todos os participantes maximizam sua utilidade esperada, assumindo os preços como dados*
2. *a condição de market clearing implica* $X(p^*) = X^s(p^*) = X^d(p^*)$
3. *os agentes têm expectativa racional:*

$$Q^e(p) = \begin{cases} Q(p) & \textit{para todo } p \textit{ para o qual } X^s(p) > 0 \\ \underline{q} & \textit{caso contrário.} \end{cases} \tag{5.14}$$

A última condição afirma que os compradores antecipam corretamente o comportamento do vendedor e, portanto, podem calcular a qualidade esperada dos produtos oferecidos no mercado a qualquer preço. Isso significa que $Q^e(p) = Q(p)$ para qualquer preço para o qual os vendedores estão oferecendo uma quantidade positiva das mercadorias. Para os preços p para os quais nenhum bem é oferecido, $Q(p)$ não está definido: não é possível calcular o valor esperado de "nada". Assim, as expectativas "racionais" não fixam valores para $Q(p)$ neste caso e temos um grau de liberdade.

Nosso modelo permite qualquer preço $p \geq 0$. Como vimos, o comportamento do vendedor pode ser perfeitamente caracterizado por qualquer um desses preços dados por (5.11). No entanto, para caracterizar o comportamento do comprador usando (5.13), precisamos atribuir um valor a $Q^e(p)$ para preços em que $X^s(p) = 0$. Como queremos um modelo "completo" que forneça previsões sobre o comportamento de compradores e vendedores para todos os níveis de preços possíveis, resolvemos esse problema técnico adicionando uma suposição. Para os preços em que nenhuma mercadoria é oferecida, as crenças são simplesmente definidas para o menor nível de qualidade possível

$$Q^e(p) = \underline{q} \text{ para todo } p \text{ para o qual } X^s(p) = 0. \tag{5.15}$$

Há ganhos esperados do comércio para todas as unidades do bem, uma vez que os preços de reserva dos vendedores são mais baixos do que a utilidade esperada dos compradores de possuir o bem: $\alpha\mathbb{E}[q] < \mathbb{E}[q]$. Como sob informação completa, desde que $n > m$, cada vendedor pode extrair o excedente total esperado cobrando um preço $p^* = Q^e(p^*) = Q(p^*) = \mathbb{E}[q]$ para o bem. Isso satisfaz a condição de expectativa racional. Todos os vendedores venderão seu bem pelo preço de equilíbrio, de modo que $X^s(p^*) = m$. Como $n > m > 0$, a única maneira de equalizar a oferta e a demanda é na quantidade $X(p^*) = X^s(p^*) = X^d(p^*) = m$.

Essa condição satisfaz o market clearing: qualquer demanda entre 0 e n - isto é, demanda igual a $m < n$ - é consistente com o comportamento de maximização porque os compradores são indiferentes entre comprar e vender em $p^* = \mathbb{E}[q]$. O excedente criado pelo comércio sob informação incompleta mas simétrica (IS) é igual àquele sob informação completa. Em nosso modelo, estamos adotando um ponto de vista "ex ante", no qual as utilidades são avaliadas antes que a qualidade se torne conhecida. Assim, temos uma alocação Pareto-ótima ex ante. Claramente, ex post, um comprador que adquiriu um bem que acaba por ter qualidade $q < p^*$ estará pior do que se nenhuma negociação tivesse ocorrido (enquanto o vendedor estivesse em melhor situação) para que a alocação de equilíbrio não fosse Pareto ótima ex post também.

Proposição 5.2.3. *Sob informação incompleta, mas simétrica, existe um equilíbrio de mercado no qual*

1. *todos os bens são comercializados*
2. $p^* = \mathbb{E}[q] = \dfrac{\overline{q} + \underline{q}}{2}$
3. $S^{IS} = S^{CI} = m(1-\alpha)\dfrac{\overline{q} + \underline{q}}{2}$

Este equilíbrio de mercado leva a uma alocação Pareto-ótima ex ante.

Fazendo um balanço dos nossos resultados até agora, vemos que a introdução de informações imperfeitas, por si só, não leva a resultados ineficientes. Todos os ganhos do comércio são realizados apesar da qualidade incerta. O que importa é se a informação é simétrica ou não. Na próxima seção, veremos que ineficiências podem surgir quando a informação é assimétrica.

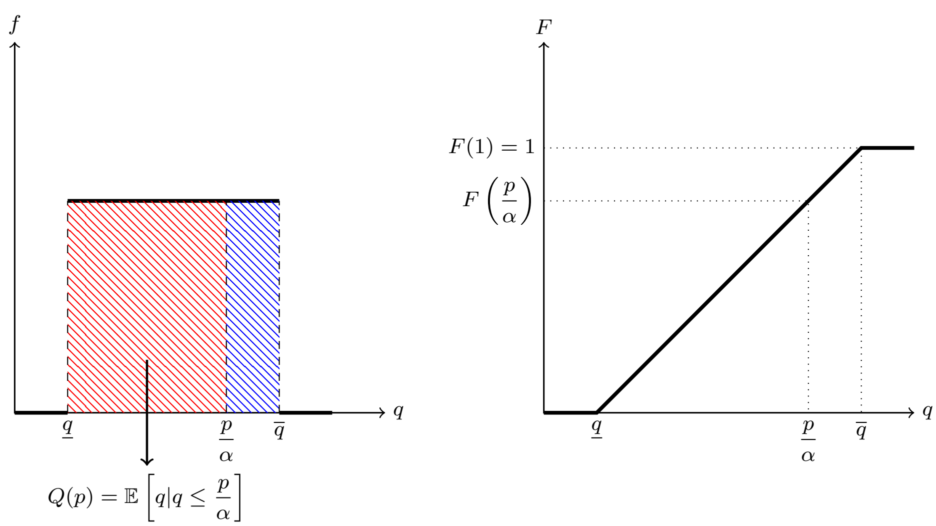
5.2.1.3 Informação Incompleta e Assimétrica

Nesse cenário, o vendedor sabe a qualidade do bem que possui, enquanto os potenciais compradores não (informação assimétrica). Como a qualidade dos produtos não é distinguível pelos compradores, todos os produtos dos vendedores são substitutos perfeitos para os compradores. Assim, como na última seção, haverá um preço de mercado único para todos os bens comercializados no mercado.

Lembre-se que o preço de reserva de um vendedor que possui um bem de qualidade conhecida q é αq, $0 < \alpha < 1$. A um determinado preço de mercado p apenas vendedores com um bem para o qual $p \geq \alpha q$ estão dispostos a vender. Isso é equivalente a $q \leq \dfrac{p}{\alpha}$. Assim, ao preço p, todos os vendedores que possuem uma bem de qualidade não maior que $\dfrac{p}{\alpha}$ vendem, enquanto todos os outros vendedores não vendem seus bens. A Figura 5.1 ilustra como obter a parcela de vendedores dispostos a vender a preço p na população de vendedores.

O painel esquerdo representa a função de densidade f. Todos os vendedores que possuem um bem com qualidade que não exceda $\dfrac{p}{\alpha}$ constituem uma fração da população total m. A superfície de cada uma das áreas hachuradas pode ser lida da função de distribuição F. Por exemplo, a área hachurada maior tem uma superfície total igual a 1, $F(\overline{q}) = 1$, uma vez que todos os vendedores (massa m) possuem um bem de qualidade que não excede a qualidade máxima possível $\overline{q}$.

Figura 5.1 – Fração de vendedores que ofertam bens aos preços p



Assim, a função de distribuição F permite medir diretamente a proporção da população de vendedores para os quais $q \leq \dfrac{p}{\alpha}$, como segue:

$$F\left(\frac{p}{\alpha}\right) = \int_{\underline{q}}^{p/\alpha} f(q)dq. \tag{5.16}$$

Para a distribuição uniforme podemos calcular a quantidade de bens ofertados ao preço p

como:

$$X^s(p) = mF\left(\frac{p}{\alpha}\right) = \begin{cases} 0 & \text{se } p < \alpha \underline{q} \\ \dfrac{m(p-\alpha\underline{q})}{\alpha(\overline{q}-\underline{q})} & \text{se } \alpha\underline{q} \leq p \leq \alpha\overline{q} \\ m & \text{se } p > \alpha\underline{q}. \end{cases} \tag{5.17}$$

Para calcular a qualidade esperada dos bens que são fornecidos a um preço p, precisamos calcular a qualidade esperada do bem, dado que $q \leq \dfrac{p}{\alpha}$: $Q(p) = \mathbb{E}\left[q|q \leq \dfrac{p}{\alpha}\right]$. Como primeiro passo, calculamos a densidade condicional

$$g\left(q|q \leq \frac{p}{\alpha}\right) = \begin{cases} \dfrac{f(q)}{F\left(p/\alpha\right)} = \dfrac{\alpha}{p-\alpha\underline{q}} & \text{se } q \leq \dfrac{p}{\alpha} \\ 0 & \text{se } q > \dfrac{p}{\alpha}. \end{cases} \tag{5.18}$$

Já que a preços $p > \alpha\overline{q}$ todos os bens são oferecidos, um bem negociado então espera qualidade igual à expectativa incondicional $Q(p) = \mathbb{E}[q] = \dfrac{\underline{q}+\overline{q}}{2}$. Nenhum bem é oferecido para $p < \alpha\underline{q}$. A qualidade esperada dos bens oferecidos pelo preço $p \in [\alpha\underline{q}, \alpha\overline{q}]$ é

$$\begin{aligned} Q(p) = \mathbb{E}\left[q|q \leq \frac{p}{\alpha}\right] &= \int_{\underline{q}}^{p/\alpha} qg\left(q|q \leq \frac{p}{\alpha}\right) dq \\ &= \int_{\underline{q}}^{p/\alpha} \frac{\alpha q}{p-\alpha\underline{q}} dq \\ &= \frac{\alpha\left(\dfrac{p^2}{\alpha^2} - \underline{q}^2\right)}{2(p-\alpha\underline{q})} = \frac{p+\alpha\underline{q}}{2\alpha},\ p \in [\alpha\underline{q}, \alpha\overline{q}]. \end{aligned} \tag{5.19}$$

Portanto, a oferta é

$$Q(p) = \begin{cases} \text{não definida} & \text{se } p < \alpha\underline{q} \\ \min\left\{\dfrac{p+\alpha\underline{q}}{2\alpha}, \dfrac{\underline{q}+\overline{q}}{2}\right\} & \text{se } \alpha\underline{q}. \end{cases} \tag{5.20}$$

Vemos que sob informação assimétrica não apenas a quantidade ofertada no mercado, $X^s(p)$, mas também a qualidade média dos produtos oferecidos, $Q(p)$, dependem do preço de mercado. A um preço p, o vendedor marginal está oferecendo um nível de qualidade $\dfrac{p}{\alpha}$. Se o preço aumenta, tanto a quantidade quanto a qualidade média aumentam, porque os vendedores com produtos de maior qualidade agora acham que vale a pena também vender o seu bem, aumentando a qualidade do vendedor marginal e aumentando a quantidade no mercado. Além disso, quanto menor o gosto dos vendedores pelo seu bem, α, maior a qualidade média oferecida no mercado para qualquer preço (e maiores os ganhos potenciais do comércio). A qualidade média oferecida no mercado atinge seu nível máximo - a qualidade média de todos os bens na economia - somente se o preço de mercado exceder o preço de reserva do dono do item de maior qualidade, $\alpha\bar{q}$.

Os compradores não são capazes de discriminar entre os produtos oferecidos pelos vendedores, mas sabem que os vendedores estão cientes da qualidade de seu próprio bem. Infelizmente, não há como verificar a alegação que os vendedores fazem sobre a qualidade de seu bem, e os compradores precisam confiar na distribuição de qualidades entre os vendedores, F, e em seu conhecimento das preferências dos vendedores. Com base nesta informação, os compradores formam crenças, $Q^e(p)$, sobre a qualidade esperada dos produtos vendidos no mercado ao preço p. Como na seção anterior, a demanda é dada por

$$X^d(p) = \begin{cases} n & \text{se } Q^e(p) > p \\ [0, n] & \text{se } Q^e(p) = p \\ 0 & \text{se } Q^e(p) < p. \end{cases} \tag{5.21}$$

Proposição 5.2.4. *Sob informação incompleta, em um equilíbrio de mercado*

1. *todos os participantes maximizam sua utilidade esperada, assumindo os preços como dados*
2. *a condição de market clearing implica* $X(p^*) = X^s(p^*) = X^d(p^*)$
3. *os agentes têm expectativa racional:*

$$Q^e(p) = \begin{cases} Q(p) & \textit{para todo } p \textit{ para o qual } X^s(p) > 0 \\ \underline{q} & \textit{caso contrário.} \end{cases} \tag{5.22}$$

Para ver se existe um equilíbrio de mercado no qual uma quantidade positiva do bem deve ser negociada, examinamos se $X(p^*) > 0$ pode ser compatibilizado com as condições de equilíbrio.

A condição de market clearing nos diz que em equilíbrio $X(p^*) = X^s(p^*) = X^d(p^*)$. Sabemos que $X^s(p) > 0$ se e somente se $p > \alpha\underline{q}$, caso contrário, nenhum vendedor está disposto a oferecer o bem. Assim, em equilíbrio

$$p^* > \alpha\underline{q}. \tag{5.23}$$

A partir do esquema de demanda em (5.21), vemos que o mercado pode se ajustar em uma quantidade positiva se e somente se $Q^e(p^*) = p^*$:

1. Se $Q^e(p^*) > p^*$ temos excesso de demanda, isto é, $X^d(p^*) = n > m \geq X^s(p^*)$.
2. Se $Q^e(p^*) < p^*$ temos demanda igual a zero, isto é, $X^d(p^*) = 0$.
3. Se $Q^e(p^*) = p^*$ a condição de market clearing pode ser satisfeita dado que para qualquer valor de p^* temos $X^d(p^*) \in [0, n], n < m$, enquanto $X^s(p^*) \in (0, m]$ depende de p^*.

Portanto, em equilíbrio

$$Q^e(p^*) = p^*. \tag{5.24}$$

Esta condição junto com a condição de expectativas racionais implica que em equilíbrio

$$Q(p^*) = Q^e(p^*) = p^*. \tag{5.25}$$

Logo,

$$\min\left\{\frac{p + \alpha \underline{q}}{2\alpha}, \frac{\underline{q} + \overline{q}}{2}\right\} = p^*. \tag{5.26}$$

Podemos distinguir os seguintes casos:

1. $p^* = \dfrac{\underline{q} + \overline{q}}{2} = \min\left\{\dfrac{p + \alpha \underline{q}}{2\alpha}, \dfrac{\underline{q} + \overline{q}}{2}\right\}$

   O caso acima ocorre se e somente se $p^* \geq \alpha \overline{q}$. Juntamente com $p^* = \dfrac{\underline{q} + \overline{q}}{2}$ isto implica que o parâmetro α deve assumir o valor $\alpha \leq \dfrac{\underline{q} + \overline{q}}{2\overline{q}}$. Lembre-se que, em equilíbrio, devemos ter $p^* > \alpha \underline{q}$. Então,

$$p^* = \frac{\underline{q} + \overline{q}}{2} > \frac{\underline{q} + \underline{q}}{2} > \alpha \underline{q}, \quad \text{dado que } 0 < \alpha < 1. \tag{5.27}$$

2. Se $\alpha > \dfrac{\underline{q} + \overline{q}}{2\overline{q}}$, temos que $p^* = \dfrac{p^* + \alpha \underline{q}}{2\alpha} = \min\left\{\dfrac{p + \alpha \underline{q}}{2\alpha}, \dfrac{\underline{q} + \overline{q}}{2}\right\}$. Isso implica que $p^* = \dfrac{\alpha \underline{q}}{2\alpha - 1}$. Lembre-se que, em equilíbrio, devemos ter $p^* > \alpha \underline{q}$. Então,

$$p^* = \frac{\alpha}{2\alpha - 1}\underline{q} > \alpha \underline{q}, \quad \text{dado que } p^* > 0 \Rightarrow 1 > \alpha. \tag{5.28}$$

Resumindo,

$$p^* = \begin{cases} \dfrac{\underline{q} + \overline{q}}{2} & \text{se } \alpha \leq \dfrac{\underline{q} + \overline{q}}{2\overline{q}} \\ \dfrac{\alpha \underline{q}}{2\alpha - 1} & \text{se } \alpha > \dfrac{\underline{q} + \overline{q}}{2\overline{q}} \end{cases} \tag{5.29}$$

Substituindo p^* na equação que define a oferta de bens, o caso 1 leva à comercialização de todos os bens, uma vez que $p^* \geq \alpha \overline{q} > \alpha \underline{q}$. Assim:

$$X(p^*) = X^s(p^*) = X^d(p^*) = m\frac{2(1-\alpha)\underline{q}}{(2\alpha-1)(\overline{q}-\underline{q})} < m, \quad \text{para } \alpha > \frac{\underline{q}+\overline{q}}{2\overline{q}} \tag{5.30}$$

O excedente criado é menor que no caso de informações simétricas:

$$S^{AI} = \int_{\underline{q}}^{p^*/\alpha} (1-\alpha)qf(q)dq = \frac{p+\alpha\underline{q}}{2\alpha} < S^{IS} = m(1-\alpha)\frac{\overline{q}+\underline{q}}{2}. \tag{5.31}$$

No caso 2, os vendedores valorizam suficientemente seus próprios bens (α alto), de tal forma que os proprietários de produtos de alta qualidade não estejam dispostos a negociar com os preços pelos quais os produtos de menor qualidade são comercializados. Como os vendedores têm informações privadas sobre a qualidade de seus produtos, surge um problema de seleção adversa, em que produtos de maior qualidade não são comercializados.

Proposição 5.2.5. *Sob informação assimétrica, existe um equilíbrio de mercado no qual os bens são comercializados se e somente se* $\alpha \leq \frac{\overline{q}+\underline{q}}{2}$. *Então,*

$$p^* \geq \alpha\overline{q} > \alpha\underline{q} \tag{5.32}$$

e

$$S^{AI} = S^{IS} \tag{5.33}$$

Se $\alpha > \frac{\overline{q}+\underline{q}}{2}$, *então*

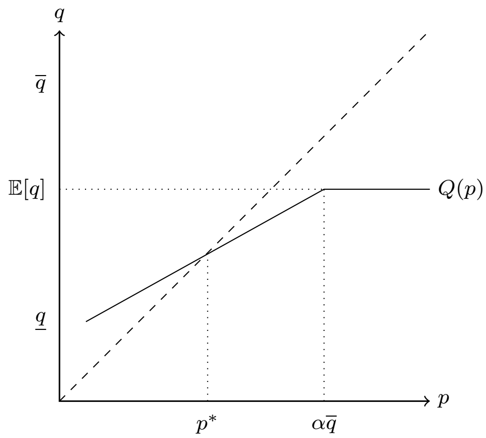
$$X(p^*) = m\frac{2(1-\alpha)\underline{q}}{(2\alpha-1)(\overline{q}-\underline{q})} < m \tag{5.34}$$

$$p^* = \frac{\alpha\underline{q}}{2\alpha-1} \tag{5.35}$$

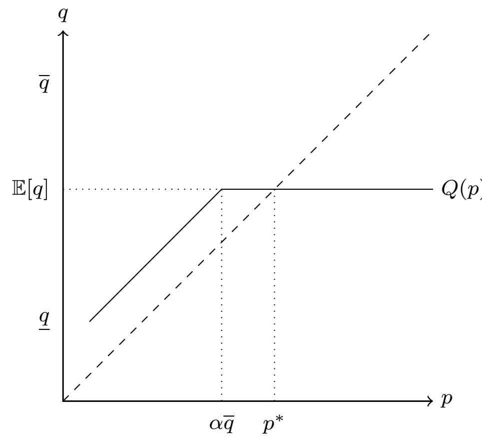
$$S^{AI} = \frac{p+\alpha\underline{q}}{2\alpha} < S^{IS} \tag{5.36}$$

Assim, em contraste com as configurações de informação completa e informação incompleta mas simétrica, um equilíbrio de mercado no qual todos os bens são comercializados só existe se $\alpha \leq \frac{\overline{q}+\underline{q}}{2}$ (painel (b) da Figura 5.2). Sempre que o gosto pela qualidade dos compradores e vendedores não diverge tanto, a seleção adversa expulsa produtos de maior qualidade (painel (a) da Figura 5.2). Isso impede que todos os ganhos potenciais do comércio sejam realizados, diminuindo o excedente realizado de S^{IS} para a S^{AI}. A seleção adversa torna-se mais severa, quanto mais alinhados os gostos de compradores e vendedores, ou seja, quanto mais próximo α é da unidade.

Figura 5.2 – Equilíbrio de mercado sob informação asismétrica



(a) Equilíbrio: comércio parcial



(b) Equilíbrio: comércio de todos os bens

De nossos resultados até agora, podemos ver que a seleção adversa tende a expulsar produtos de maior qualidade. Para todos os preços $p < \alpha\overline{q}$ a qualidade média oferecida no mercado, $q(p)$ é inferior à qualidade média de todos os bens potencialmente disponíveis para venda, $\mathbb{E}[q]$. Devido à incerteza de qualidade para os compradores, um preço deve "encaixar" todos os bens. No entanto, a preços inferiores a $\alpha\overline{q}$ os proprietários de bens de maior qualidade preferem não vender os seus bens, diminuindo a qualidade média dos bens que estão disponíveis para venda. Isto confirma muito bem a intuição descrita por Akerlof de que "os bens [ruins] expulsam os não tão ruins que expulsam os médios que expulsam os não-tão-bons". Uma possível solução para situações em que a seleção adversa surgiria é que os vendedores oferecem garantias aos compradores. Isso permite que os compradores inspecionem o bem e o devolvam se a qualidade não corresponder àquela indicada pelo vendedor.

5.2.2 Modelo de Capital Humano de Becker (1964)

A educação é talvez a decisão de investimento mais significativa que você fará. A maioria dos cidadãos passa de 12 a 20 anos de vida estudando. A oferta de ensino envolve dois tipos de custos: custos diretos como edifícios, professores, livros didáticos, etc.; custos indiretos como custos de oportunidade de frequentar a escola em vez de trabalhar ou se divertir. Esses custos certamente afetam os custos diretos da educação. Esse enorme investimento é um uso socialmente eficiente de recursos? Ou é o resultado de equilíbrio de um processo que não necessariamente maximiza o bem-estar social?

Os economistas historicamente usaram o Modelo de Capital Humano criado por Gary S. Becker (1964) e estendido por Jacob Mincer. Este modelo considera a educação como um investimento inicial que aumenta a produtividade futura. Spence sugeriu um segundo modelo: o modelo de sinalização. Esse modelo sugere que indivíduos que passam décadas na escola podem estar sinalizando que são produtivos. Esse modelo sugere conclusões muito diferentes sobre o tamanho e a forma ideal de nossas instituições educacionais.

Defina $w(s)$ como o salário de um indivíduo com s anos de escolaridade. Suponha que $w'(s) > 0$: a produtividade aumenta com a escolaridade. Também modelaremos o mercado de trabalho como competitivo, de modo que os lucros reflitam a produtividade (e também aumentem com a escolaridade).

Suponha que os custos diretos da escolaridade, c, são zero por enquanto. Claro que eles não são zero no mundo real, mas não precisamos deles para ilustrar as principais lições desse modelo. Defina $r > 0$ como a taxa de juros. Se eu decidisse não colocar \$1 em educação e em vez disso colocar \$1 no banco, eu receberia $\$(1 + r)$ no próximo período. Por simplicidade, suponha que as pessoas vivam para sempre. Neste modelo, qual é o valor presente dos ganhos vitalícios de alguém com 1 ano de escolaridade? É o Valor Presente Descontado (VPD) de receber $w(1)$ anualmente em cada ano subsequente:

$$VPD[w(1)] = w(1) + \frac{w(1)}{(1+r)} + \frac{w(1)}{(1+r)^2} + \ldots + \frac{w(1)}{(1+r)^\infty}, \tag{5.37}$$

que pode ser simplificado para

$$\begin{aligned}
VPD[w(1)]\left(\frac{1}{1+r}\right) &= \frac{w(1)}{(1+r)} + \frac{w(1)}{(1+r)^2} + \frac{w(1)}{(1+r)^3} + \ldots + \frac{w(1)}{(1+r)^\infty} \\
VPD[w(1)] - VPD[w(1)]\left(\frac{1}{1+r}\right) &= w(1) \\
VPD[w(1)] &= w(1)\left[\frac{1}{1-\dfrac{1}{1+r}}\right] \\
VPD[w(1)] &= w(1)\left(\frac{1+r}{r}\right). \qquad (5.38)
\end{aligned}$$

Note, no entanto, que se você frequentar a escola por um ano para obter $w(1)$, você não receberá o primeiro pagamento de $w(1)$ até que seu primeiro ano de escolaridade esteja completo. Então o VPD de um ano de educação é

$$w(1)\frac{1+r}{r}\frac{1}{1+r} = \frac{w(1)}{r}. \tag{5.39}$$

Por outro lado, uma pessoa que não tem um ano adicional de escolaridade recebe

$$VPD[w(0)] = w(0) + \frac{w(0)}{1+r} + \frac{w(0)}{(1+r)^2} + \ldots + \frac{w(0)}{(1+r)^\infty} = w(0)\left(\frac{1+r}{r}\right). \tag{5.40}$$

Assim, o benefício líquido do VPD de obter mais um ano de escolaridade é

$$VPD(+1) = w(1)\frac{1}{r} - w(0)\left(\frac{1+r}{r}\right) = \frac{1}{r}(w(1) - w(0)) - w(0). \tag{5.41}$$

Agora, vamos assumir:

- Um mercado competitivo para o trabalho.
- Mercados de capital perfeitos (podem sempre emprestar o custo total da escolaridade).
- Agentes racionais, cada um com o mesmo potencial de ganho.

Em equilíbrio, deve ser o caso que os custos e benefícios de um ano adicional de escolaridade são equacionados. Se os custos fossem menores que os benefícios, ninguém receberia educação. Se os custos fossem maiores, todos receberiam educação. Então, o equilíbrio deve tornar todos indiferentes. Isso implica que

$$\begin{aligned} w(1)\frac{1}{r} &= w(0)\left(\frac{1+r}{r}\right) \\ \frac{w(1)}{w(0)} &= 1+r \\ \ln w(1) - \ln w(0) &= \ln(1+r) \approx r. \end{aligned} \tag{5.42}$$

Em outras palavras, o incremento salarial de mais um ano de escolaridade deve ser aproximadamente igual à taxa de juros. Se isso não fosse verdade, então as pessoas mudariam seus investimentos em educação versus outras oportunidades de igualar esses retornos marginais.

Por mais simples que seja este modelo, ele faz um ótimo trabalho ao capturar uma notável regularidade empírica. Nos últimos 101 anos (o que é tão longo quanto podemos medir para os EUA), a taxa estimada de retorno a um ano de escolaridade foi de cerca de 5 a 10% - aproximadamente igual à taxa real de juros mais inflação.

O modelo de diferenças equalizadoras de investimento em capital humano de Mincer tem quatro implicações testáveis:

1. Pessoas que frequentam anos adicionais de escolaridade são mais produtivas.
2. Pessoas que frequentam anos adicionais de escolaridade recebem salários mais altos.
3. As pessoas frequentam a escola enquanto são jovens, ou seja, antes de ingressarem no mercado de trabalho. Por quê? Porque os custos da escola são os mesmos sempre que você a frequenta, mas os benefícios não começam a acumular até que você os complete. Você deve, portanto, obter sua educação antes de começar a trabalhar.
4. A taxa de retorno à escolaridade deve ser aproximadamente igual à taxa de juros.

5.2.3 Modelo de Sinalização de Spence (1973)

O modelo Mincer pressupõe que a educação aumenta a produtividade. Mas se a educação fosse improdutiva, alguma das coisas acima seria verdadeira? Antes do artigo de Spence de 1973, a maioria dos economistas teria dito "não" reflexivamente. Se a educação é improdutiva, por que as pessoas gastariam tempo para adquiri-la? E por que os empregadores pagariam salários mais altos aos trabalhadores instruídos?

A surpresa do modelo de Spence é que, mesmo que a educação seja improdutiva, pode haver uma demanda de empregados e empregadores em equilíbrio. A principal lição do modelo de Spence é que as pessoas podem usar a educação para sinalizar que são produtivas, em vez de se tornarem mais produtivas.

Enquanto Akerlof observa a quebra do mercado como uma consequência muito séria da informação assimétrica, não observamos muitas falhas de mercado na realidade, apesar do fato de informações assimétricas. A questão de pesquisa que Spence nos propõe é: podemos construir um modelo no qual a quebra do mercado como em Akerlof pode ser evitada mesmo quando há informação assimétrica? A breve resposta aqui é positiva. Em um cenário de sinalização com custos específicos de atributos, os trabalhadores conhecem melhor sua própria produtividade do que os empregadores, e podem efetivamente sinalizar suas informações para alcançar um equilíbrio.

Considere os seguintes pressupostos:

1. As pessoas são de habilidade "alta" ou "baixa": H, L.

2. Pessoas de alta capacidade são inerentemente mais produtivas do que pessoas de baixa capacidade. Não há como uma pessoa de alta capacidade se tornar uma pessoa de baixa habilidade, ou vice-versa.

3. A capacidade de um indivíduo é conhecida para ele ou ela, mas não para potenciais empregadores.

4. A educação não afeta a capacidade/produtividade.

5. Pessoas de alta capacidade têm menor custo de frequentar a escola do que outras. Por que isso seria assim? Custos psíquicos mais baixos para estudar tópicos em que você é naturalmente bom, os subsídios para a educação são maiores para pessoas de alta capacidade por meio de bolsas de estudo, por exemplo.

Tabela 5.1 – Parâmetros do modelo

Grupo	Produtividade	% na população	Custo da educação
L	$Y_L = 1$	λ	S
H	$Y_L = 2$	$1 - \lambda$	$\frac{S}{2}$

Assim, a produtividade média da população é $2 \cdot (1 - \lambda) + 1 \cdot \lambda = 2 - \lambda$. Observe que a produtividade de um trabalhador não depende da quantidade de escola que ele consegue. Quais são os possíveis equilíbrios desse modelo especificamente, que salários os empregadores devem oferecer aos trabalhadores com diferentes níveis de escolaridade S, e quanto de escolaridade os trabalhadores H e L devem obter?

Uma característica que torna os modelos com informações assimétricas um pouco diferentes dos modelos que estudamos anteriormente é que o equilíbrio não é determinado primariamente por um conjunto de condições marginais (por exemplo, lucro marginal é zero ou a taxa marginal de substituição é igual à razão de preço) . Em vez disso, o equilíbrio depende de encontrar um conjunto de estratégias compatíveis, nas quais o comportamento de cada jogador faz sentido, considerando os comportamentos dos outros jogadores. Pensamos em partidos nos diferentes lados do mercado (por exemplo, compradores vs. vendedores) como estratégias de escolha (ações viáveis) que maximizem seus pagamentos dadas as estratégias escolhidas dos jogadores do outro lado do mercado. Mas, é claro, os jogadores do outro lado do mercado também estão escolhendo estratégias para maximizar seus retornos, dadas as ações (ou ações antecipadas) dos outros jogadores.

Um equilíbrio nesse cenário é um conjunto de estratégias complementares de tal forma que nenhum dos lados quer mudar unilateralmente sua estratégia, dada a estratégia do outro lado. Esta noção é o que é chamado de Equilíbrio de Nash depois de John Forbes Nash, que desenvolveu a idéia e provou sua existência em uma dissertação de doutorado de Princeton em 28 páginas. Para solidificar ainda mais sua compreensão, aqui está uma definição informal do Equilíbrio de Nash: Um Equilíbrio de Nash é um conceito de solução para um jogo envolvendo dois ou mais jogadores, no qual cada jogador presume conhecer as estratégias de equilíbrio dos outros jogadores, e nenhum jogador tem nada a ganhar mudando apenas sua própria estratégia unilateralmente. Se cada jogador tiver escolhido uma estratégia e nenhum jogador puder se beneficiar mudando sua estratégia enquanto os outros jogadores mantiverem a mesma inalterada, então o conjunto atual de escolhas de estratégia e os payoffs correspondentes constituem um equilíbrio de Nash.

5.2.3.1 Equilíbrio Separador

Em um modelo com dois tipos de alunos (aqui H e L), há duas classes potenciais de equilíbrio estratégico puro: um equilíbrio separador em que os dois tipos seguem diferentes estratégias (ou seja, buscam diferentes níveis de educação) e realizam resultados diferentes; e um equilíbrio agregador onde ambos os tipos seguem a mesma estratégia (ou seja, seguem o mesmo nível de educação) e apresentam os mesmos resultados. Também pode haver um equilíbrio de estratégia mista onde os agentes randomizam entre as ações de modo a fazer com que outros jogadores desempenhem estratégias compatíveis. Não consideraremos aqui os equilíbrios de estratégia mista. Consideraremos em primeiro lugar a possibilidade de separar os equilíbrios.

Suponha que as empresas ofereçam o seguinte esquema salarial:

$$w(S) = 1 + I[S \leq S^*], \tag{5.43}$$

em qeu $I[\cdot]$ é uma função indicadora. Um trabalhador com $S \leq S^*$ anos de escolaridade recebe 2 e 1 caso contrário. Isso pode fazer sentido na perspectiva das empresas se elas acreditarem que apenas trabalhadores do tipo H estão dispostos a estudar S^* anos de escolaridade.

Quanta educação os trabalhadores obterão? O problema do trabalhador é

$$\max_S w(S) - c(S). \tag{5.44}$$

Para um trabalhador do tipo H, o custo de S^* anos de educação é de $0.5S^*$ e o benefício salarial é 1. Assim, trabalhadores do tipo H irão frequentar a escola por S^* anos se: $w(S \leq S^*) - 0.5S^* > w(S < S^*) \Rightarrow\ 2 - 0.5S^* > 1$.

Para um trabalhador do tipo L, o custo de S^* anos de educação é de S^* e o benefício salarial é 1. Assim, trabalhadores do tipo L irão frequentar a escola por S^* anos se: $w(S \leq S^*) - S^* > w(S < S^*) \Rightarrow\ 2 - 1 > 1$.

Considere se o empregador fixa $S^* = 1 + \varepsilon$, em que ε é um número muito pequeno e positivo. De acordo com esta política salarial, os trabalhadores do tipo H preferirão estritamente a obtenção de S anos de educação (uma vez que $2 - 0.5(1+\varepsilon) > 1$), enquanto que os trabalhadores do tipo L não acharão que vale a pena obter S^* anos de educação (uma vez que $2 - (1+\varepsilon) < 1$).

Agora vamos analisar o lado da firma. O cronograma salarial do empregador representado por (5.44) é um cronograma salarial de equilíbrio? Ou os empregadores querem construir um cronograma salarial diferente dado o comportamento dos trabalhadores?

Para as empresas oferecerem esse cronograma salarial, deve ser o caso que

$$\mathbb{E}[Y(S)|w(S)] \leq w(S), \tag{5.45}$$

isto é, a produtividade esperada dos trabalhadores que se qualificam para um nível salarial baseado em sua escolaridade deve ser pelo menos igual ao salário. Aqui, a função $Y(S)$ fornece a produtividade de trabalhadores que fornecem mão-de-obra com nível de escolaridade S. Estamos condicionando $Y(S)$ à função $w(S)$ porque os empregadores oferecem um plano salarial em vez de um único salário. A escolha do trabalhador acerca de S_i depende, portanto, do plano salarial $w(S)$ e, assim, a produtividade $Y(S)$ muda com o plano salarial $w(S)$ quando diferentes tipos de trabalhadores selecionam diferentes quantidades de escolaridade S.

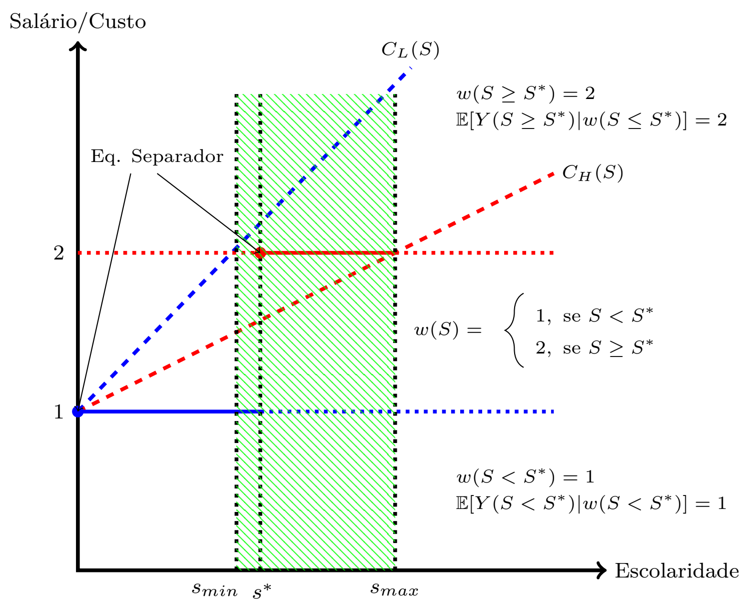
No nosso exemplo, os trabalhadores do tipo L não estão dispostos a passar por $1 + \varepsilon$ anos de escolaridade, então eles podem também escolher $S = 0$. Esses trabalhadores têm produtividade 1 e recebem o salário 1, então o plano salarial do empregador é racional para esses trabalhadores ($\mathbb{E}[Y(0)|w(S)] = w(0) = 1$).

Da mesma forma, neste exemplo, os trabalhadores do tipo H escolheriam a escolaridade $S = 1 + \varepsilon$. Eles têm produtividade igual a 2 e recebem um salário igual 2, e assim o esquema salarial do empregador também é racional para esses trabalhadores: ($\mathbb{E}[Y(1)|w(S)] = w(1) = 2$).

Então, este é um equilíbrio: trabalhadores de alta habilidade obterão educação $S = 1 + \varepsilon$, enquanto trabalhadores com baixa habilidade obterão educação $S = 0$. Os empregadores podem perfeitamente determinar o tipo de cada trabalhador com base em seu nível de escolaridade e definir o cronograma salarial de acordo, de modo que a empresa está bem com esse arranjo. Note-se que o mercado de trabalho é considerado perfeitamente competitivo, portanto não há desvio que possa dar à empresa lucros estritamente positivos. Nem os trabalhadores nem os empregadores terão incentivo para se desviar do esquema de pagamento.

Vamos desenhar as curvas de indiferença de ambos os tipos de trabalhadores no espaço (w, S), como na Figura 5.3 abaixo. Observe que, nessa figura, as curvas de custo $C_L(S)$ e $C_H(S)$ servem como curvas de indiferença: os trabalhadores estão dispostos a ter uma certa quantidade de escolaridade apenas se forem compensados com um salário mais alto. As curvas de indiferença originam-se da oferta inicial de salário para $w(S = 0)$, e inclinam-se para cima com o custo de educação do trabalhador. Ao longo dessas curvas de custo, os trabalhadores de cada tipo são indiferentes entre todos os pacotes em suas respectivas curvas de custo, oferecendo salários mais altos e maior escolaridade em relação a $w(S = 0)$. Os trabalhadores preferem estritamente estar acima (a noroeste) dessas curvas em relação a sua cesta padrão de $w(S = 0)$. Os trabalhadores preferem, estritamente, não ficar abaixo (sudeste) dessas curvas, ou seja, pior do que em sua cesta padrão de $w(S = 0)$.

Figura 5.3 – Potencial equilíbrio separador com $\lambda = 0.5$



Observe que o equilíbrio separador requer que o esquema salarial induza a auto-seleção: os trabalhadores de alta produtividade optam por obter dois anos de escolaridade e os trabalhadores de baixa produtividade optam por obter apenas um. Em equilíbrio, os empregadores têm o prazer de pagar aos trabalhadores com S^* anos de escolaridade um salário de 2 anos e aos trabalhadores com menos de um ano de escolaridade um salário de 1, e nem os trabalhadores nem os empregadores têm um incentivo para se desviar deste equilíbrio.

O aspecto lamentável desse modelo é que a educação é completamente improdutiva, por isso esses investimentos são socialmente inúteis. Ao obter educação, os trabalhadores do tipo "sinalizam" que merecem um salário alto - mas esse é um benefício privado puro. Do ponto de vista social, essa sinalização não é útil, pois não aumenta a produção total.

Importa para este modelo se os empregadores acreditam que a educação é produtiva? Na verdade, isso não acontece. Enquanto as pessoas que possuem escolaridade $S \geq S^*$ têm produtividade 2 e as que possuem escolaridade $S < S^*$ têm produtividade 1, os empregadores não têm incentivo para se desviar do esquema salarial.

5.2.3.2 Equilíbrio Agregador com Educação Positiva

O exemplo acima é um equilíbrio separador: os tipos L e H obtêm diferentes níveis de educação em equilíbrio. Há também uma multiplicidade de possíveis equilíbrios agregadores, isto é, casos em que os tipos L e H recebem educação idêntica.

Imagine a seguinte política salarial:

$$w(S) = 1 + I[S \geq S^*] \cdot (1 - \lambda), \tag{5.46}$$

assim, os trabalhadores com escolaridade inferior a S^* recebem um salário de 1 e aqueles com escolaridade maior ou igual a S^* recebem salário de $2 - \lambda$. Quem investiria em educação?

Os tipos H adquirem $S = S^*$ ao custo $0.5S^*$ se

$$2 - \lambda - 0.5S^* > 1 \ \Rightarrow \ S^* < 2(1 - \lambda), \tag{5.47}$$

e os trabalhadores do tipo L adquirem $S = S^*$ ao custo S^* se

$$2 - \lambda - S^* > 1 \ \Rightarrow \ S^* < (1 - \lambda). \tag{5.48}$$

Então, se $S^* < 1 - \lambda$, todos os trabalhadores adquirem educação igual a S^*.

Os salários dos empregadores são racionais devido a esse fato? Sim. Porque a produtividade esperada da população trabalhadora é

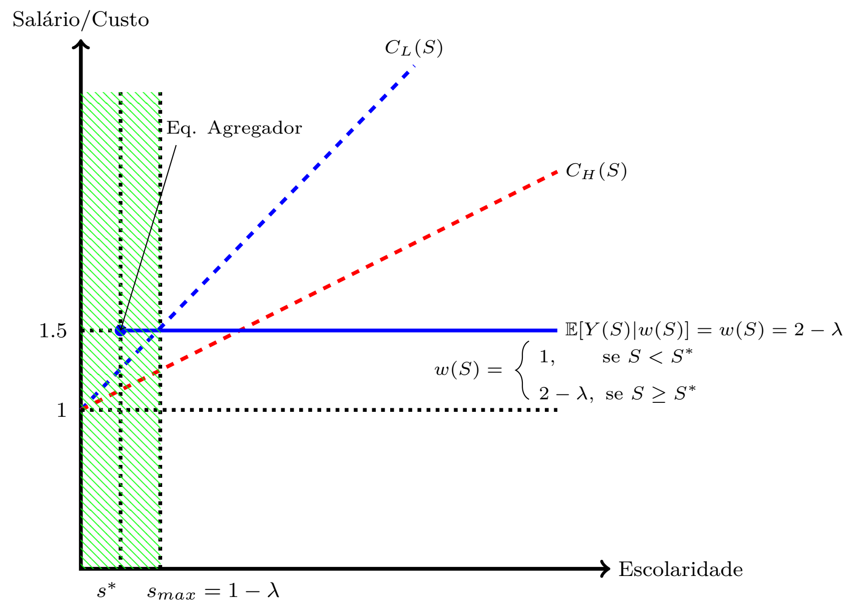
$$\mathbb{E}[Y(S \geq S^*)|w(S)] = 1 + (1 - \lambda) = 2 - \lambda = w(S > S^*). \tag{5.49}$$

Então, este é um equilíbrio agregador viável, como mostrado na Figura 5.4 abaixo.

Esse equilíbrio é um pouco estranho porque não especifica o que aconteceria se os empregadores conhecessem um grupo de trabalhadores com $S = 0$ e descobrissem que sua produtividade também era $2 - \lambda$. O modelo de Spence foi escrito quando a teoria dos jogos ainda estava em sua infância, e não faz um bom trabalho ao considerar como as crenças "fora de equilíbrio" afetam o modelo.

Observe que $S^* > 1 - \lambda$ não seria um esquema salarial de equilíbrio viável. Sob esse cronograma, trabalhadores com alta, mas não os de baixa produtividade, adquiririam educação, mas o salário pago aos trabalhadores de alta produtividade seria apenas $2 - \lambda$, enquanto sua produtividade seria 2. Os empregadores teriam um incentivo para se desviar desse cronograma salarial para aumentar os salários de trabalhadores de alta produtividade. Isso poderia teoricamente acontecer, mas criaria um equilíbrio possível completamente diferente.

Figura 5.4 – Potencial equilíbrio agregador com $\lambda = 0.5$



5.2.3.3 Equilíbrio Agregador sem Educação

É crucial lembrar que nós não especificamos com equilíbrio realmente vai acontecer aqui. Mostramos apenas exemplos de muitos equilíbrios diferentes que poderiam ocorrer. Agora considere um equilíbrio de pool diferente no qual os empregadores oferecem o plano salarial

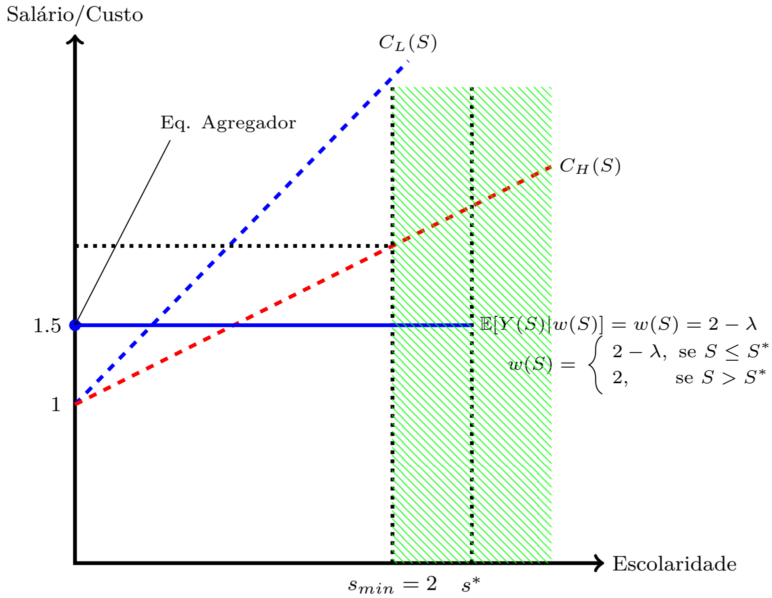
$$w(S) = (2 - \lambda) + I[S^* \geq 3]. \tag{5.50}$$

Quem obterá escolaridade neste caso? A resposta é ninguém, já que o custo de obter 3 unidades de escolaridade para H e L excede o benefício salarial de 1. Mas as crenças dos empregadores são auto-realizáveis: o conjunto de trabalhadores sem instrução tem produtividade média de $2 - \lambda$, que é igual ao salário:

$$\mathbb{E}[Y(0)|w(S)] = 2 - \lambda = w(0). \tag{5.51}$$

Então este é outro equilíbrio viável. Esses três equilíbrios (dois agregadores, um separador) podem ser classificados em termos de bem-estar total? Sim. Em um mundo em que a escolaridade é usada apenas para sinalizar a produtividade (e não gera nenhuma nova produtividade), o equilíbrio agregador com escolaridade zero é preferível a qualquer dos equilíbrios que envolvem escolaridade diferente de zero. A produtividade e os salários agregados são idênticos em todos os casos, mas qualquer equilíbrio que envolva níveis de escolaridade não nulos inclui gastos desnecessários com a escolarização. Esses investimentos em educação são pura perda de peso morto do ponto de vista social, uma vez que não aumentam a produção.

Figura 5.5 – Potencial equilíbrio agregador com $\lambda = 0.5$



5.2.4 Implicações Empíricas

O modelo de sinalização compartilha alguma implicação com o modelo de capital de humano de Gary Becker?

1. Pessoas que frequentam anos adicionais de escolaridade são mais produtivas. SIM.
2. Pessoas que frequentam anos adicionais de escolaridade recebem salários mais altos. SIM.
3. As pessoas frequentam a escola enquanto são jovens, ou seja, antes de ingressarem no mercado de trabalho. SIM.
4. A taxa de retorno à escolaridade deve ser aproximadamente igual à taxa de juros. NÃO HÁ PREVISÃO

Como as implicações empíricas dos modelos de Capital Humano e Sinalização parecem tão semelhantes, muitos economistas concluíram que esses modelos não poderiam ser distinguidos empiricamente. O artigo publicado no Quarterly Journal of Economics em 2000 por Tyler, Murnane e Willett demonstra que essa conclusão foi prematura.

Os resultados do artigo mostram grandes efeitos de sinalização para os brancos, estimados em 20% de ganho após 5 anos. Isso prova que os detentores de GED (General Educational Development) não são mais produtivos que os detentores não-GED? Não. Apenas o oposto. Para que haja um equilíbrio de sinalização, deve ser o caso de que os detentores de GED sejam, em média, mais produtivos do que desistentes da escola, que não possuem um GED.

Esses resultados provam que a educação é improdutiva? Não, eles também não têm nada a dizer sobre esta questão porque a educação/habilidade é efetivamente mantida constante por essa quase-experiência.

O que o estudo mostra sem ambiguidade é que o GED é tomado como um sinal positivo pelos empregadores. E isso só pode ser verdade se: (i) os detentores de GEDs forem, em média, mais produtivos que os não detentores de GED; (ii) o GED é, de certa forma, mais caro para alguns trabalhadores menos produtivos que os trabalhadores mais produtivos. (isso tem relação com a maturidade, o intelecto, etc.); (iii) os empregadores são incapazes de distinguir perfeitamente a produtividade diretamente e, portanto, usam o status de GED como um sinal de produtividade esperada.

5.3 Modelo de Auto-Seleção: Rothschild e Stiglitz (1976)

O mercado de seguros tem um problema de seleção adversa potencialmente enorme, problemas que podem levar à quebra do mercado. Mas, na realidade, ainda observamos o funcionamento do mercado de seguros. A questão de pesquisa aqui é: podemos projetar um modelo de equilíbrio no qual existe um equilíbrio apesar da seleção adversa? Além disso, esse equilíbrio pode ser único?

Suponha, por enquanto, que exista apenas um tipo de indivíduo cuja renda seja W quando nenhum acidente acontecer. Os indivíduos gozariam de renda $W-d$ quando um acidente acontecesse com probabilidade p. Existem muitas companhias de seguros que oferecem contratos de seguro com um prêmio de α_1 e apólice de $\widehat{\alpha}_2$. Presume-se que os indivíduos sejam avessos ao risco e as companhias de seguros são neutras ao risco. Denote como W_1 a renda dos indivíduos no caso de nenhum acidente, ou seja, $W_1 \equiv W - \alpha_1$, e denote como W_2 a renda dos indivíduos no caso de acidente, ou seja, $W_2 \equiv W - d - \alpha_1 + \widehat{\alpha}_2$.

No lado da demanda por seguro, os indivíduos estão maximizando a utilidade esperada na forma de $V(p, W_1, W_2) = (1-p)U(W_1) + pU(W_2)$, em que $U' > 0$ e $U'' < 0$. Vamos assumir que as preferências não dependem dos estados. O valor do contrato de seguro[1] (α_1, α_2), com $\alpha_2 = \widehat{\alpha}_2 - \alpha_1$, pode ser expresso como $V(p, \alpha_1, \alpha_2) = V(p, W - \alpha_1, W - d + \alpha_2)$ para os indivíduos. Portanto, o problema de otimização para o indivíduo com probabilidade de risco p é escolher (α_1, α_2) para maximizar a utilidade esperada. Aqui assumimos que a condição de racionalidade individual é válida, ou seja, $V(p, \alpha_1, \alpha_2) \geq V(p, 0, 0) \equiv V(p, W, W-d)$, ou seja, manter a apólice de seguro deve fazer indivíduos não pior do que não manter a apólice. Do lado da oferta de seguros, as seguradoras estão tentando maximizar os lucros esperados, $\pi(p, \alpha_1, \alpha_2) = (1-p)\alpha_1 - p\alpha_2$. Como vamos utilizar os rendimentos em dois casos, W_1 e W_2, como variáveis de ordenação, podemos expressar a função

[1] Os contratos de seguro reais são mais complicados porque um único contrato oferecerá cobertura contra muitas perdas potenciais. Uma generalização formal do esquema acima para cobrir este caso é simples. Suponha que um indivíduo, na ausência de seguro, tenha uma renda de W_i se o estado i ocorrer. Um contrato de seguro é simplesmente uma n-upla $(a_1, \ldots, a_n)$ cuja i-ésima coordenada descreve o pagamento líquido do indivíduo à seguradora, se o estado i ocorrer. Limitamos nossa discussão ao simples caso mencionado no texto, embora isso possa ser trivialmente estendido a esse caso mais complicado.

lucro no espaço de variáveis de ordenação. Isso é,

$$\pi = \alpha_1 - p(\alpha_1 + \alpha_2) = W - (1-p)W_1 - pW_2 - pd. \tag{5.52}$$

Por suposição, todas as companhias de seguros são neutras ao risco. A entrada gratuita nos mercados forçará o lucro esperado para a companhia de seguros a ser zero, e assim

$$(1-p)W_1 + pW_2 = W - pd, \tag{5.53}$$

será a restrição de lucro zero para as seguradoras. Além disso, notamos que a inclinação é $\frac{dW_2}{dW_1} = -\frac{(1-p)}{p}$. Portanto, sabemos que a dessa restrição é inclinada para baixo e que a magnitude da inclinação dessa reta depende do nível de risco dos indivíduos de entrar em um acidente.

Definição 5.3.1. *O equilíbrio de Nash do modelo Rothschild-Stiglitz é definido como o conjunto de contratos de seguro em que os consumidores escolhem contratos para maximizar a utilidade esperada: (i) nenhum contrato no conjunto de equilíbrio gera lucro negativo para as seguradoras; e (ii) nenhum contrato fora do conjunto de equilíbrio, se oferecido, permitirá um lucro positivo para as seguradoras.*

Esta definição pode parecer muito natural, mas acontece que é muito restritiva. Por enquanto, podemos resolver a primeira melhor solução para o modelo com informações simétricas. O preço justo correspondente à condição de lucro zero é

$$\frac{p}{1-p} = \frac{\alpha_1}{\alpha_2} \quad \text{ou} \quad p = \frac{\alpha_1}{\widehat{\alpha}_2}, \tag{5.54}$$

ou seja, o preço do seguro deve ser o mesmo que a probabilidade subjetiva de ocorrência de um acidente. O problema de otimização dos indivíduos se torna

$$\max_{\alpha_1} pU\left(W - d - \alpha_1 + \frac{\alpha_1}{p}\right) + (1-p)U(W - \alpha_1). \tag{5.55}$$

A condição de primeira ordem é

$$p\left(\frac{1}{p} - 1\right)U'(W - d + \alpha_2) - (1-p)U'(W - \alpha_1) = 0 \quad \text{ou} \tag{5.56}$$

$$U'(W - d + \alpha_2) = U'(W - \alpha_1) \quad \text{ou} \tag{5.57}$$

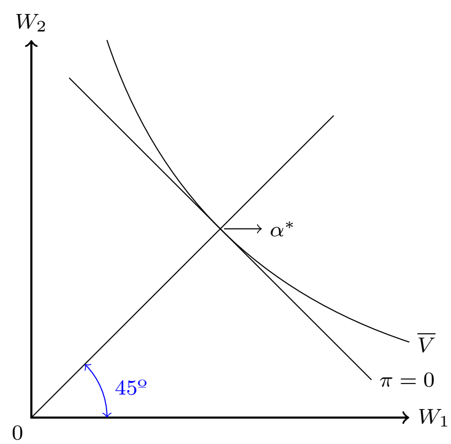
$$W_1 = W_2. \tag{5.58}$$

Alternativamente, podemos mostrar que, independentemente do seguro que detenha, a riqueza esperada dos indivíduos é

$$p\left(W - d - \alpha_1 + \frac{\alpha_1}{p}\right) + (1-p)(W - \alpha - 1) = W - pd. \tag{5.59}$$

O seguro total $\alpha_1 = pd$ pode alcançar essa riqueza esperada com certeza, e a aversão ao risco leva a um seguro total, ou seja, $W_1 = W_2$. Ou seja, o equilíbrio será o ponto de intersecção da linha de 45º (para que a renda esperada seja certa e igual entre os estados) e a restrição com a inclinação $-\dfrac{(1-p)}{p}$. Um desses equilíbrios é descrito na Figura 5.6. Observe que a mudança para o nordeste aumentaria a curva de iso-utilidade dos indivíduos e reduziria o nível de lucro das seguradoras.

Figura 5.6 – Espaço de estado da riqueza no modelo Rothschild-Stiglitz



Agora vamos introduzir assimetria no modelo. Suponha que cada indivíduo pode ser de dois tipos, alto risco (p^H) ou baixo risco (p^L), em que $p^H > p^L$. A proporção populacional do tipo de alto risco é λ. Suponha ainda que cada indivíduo reconheça seu próprio tipo, mas as companhias de seguro não. Supõe-se que a parte desinformada, as seguradoras, se mova primeiro neste modelo, fornecendo políticas de seguro para a parte informada escolher. Vamos denotar a probabilidade de risco médio ou agrupado de envolver um acidente na população como $p \equiv \lambda p^H + (1-\lambda)p^L$.

A função de utilidade esperada será

$$V_i(p_i, W_1, W_2) = (1-p^i)U(W_1) + p^i U(W_2), \tag{5.60}$$

em que $i \in \{L, H\}$.

Podemos usar o teorema da função implícita para descobrir a inclinação da curva de iso-utilidade como

$$\left(\frac{dW_2}{dW_1}\right)^i = -\frac{\partial V_i/\partial W_1}{\partial V_i/\partial W_2} = -\frac{(1-p^i)U'(W_1)}{p^i U'(W_2)}. \tag{5.61}$$

Sabemos que essas curvas de iso-utilidade têm declives negativos em todos os pontos e para ambos os tipos. Mas gostaríamos de saber qual é mais íngreme. Então descobrimos

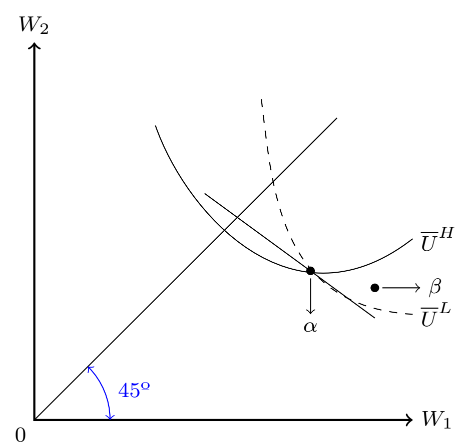
$$\left|\frac{dW_2}{dW_1}\right|^{i=H} - \left|\frac{dW_2}{dW_1}\right|^{i=L} = \frac{U'(W_1)}{U'(W_2)}\frac{p^L - p^H}{p^L p^H} < 0, \tag{5.62}$$

e concluímos que os tipos de baixo risco têm curvas de iso-utilidade mais acentuadas em cada ponto no espaço de variáveis de classificação. Ao mesmo tempo, sabemos que os tipos de baixo risco também têm curvas de iso-lucro mais acentuadas.

É intuitivo observar que, neste modelo, os tipos de alto risco estão prejudicando os tipos de baixo risco, uma vez que os últimos poderiam ter obtido contratos de seguro mais favoráveis se não houvesse tipos de alto risco. No entanto, é verdade aqui e em cada modelo de informação assimétrica que os agentes de alto risco acabam obtendo as melhores soluções na contraparte de informação completa. Por quê? Os tipos de alto risco não são melhores do que seriam na ausência de tipos de baixo risco se existe um equilíbrio separador bem-sucedido.

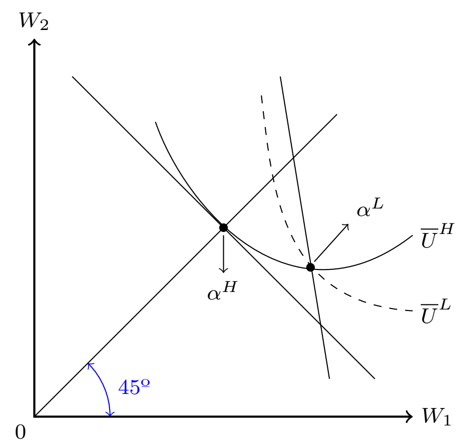
Observe ainda que um equilíbrio agregador não é possível neste modelo. Veja a Figura 5.7. Suponha que o contrário seja verdadeiro e que o equilíbrio agregador seja de fato o ponto de interseção α. Então, descobrimos que quaisquer contratos de seguro situados entre as curvas de iso-utilidade de dois tipos à direita do ponto α, por exemplo, o ponto β, não serão interessantes para os tipos de alto risco, mas mais atraentes tipos de baixo risco que estão atualmente no ponto α. Além disso, uma vez que este contrato atrai apenas os tipos de baixo risco, a companhia de seguros que fornece o contrato β claramente fará lucro positivo, violando a condição (ii) do equilíbrio de Nash do modelo Rothschild-Stiglitz. Portanto, o equilíbrio agregador α não é um equilíbrio de Nash do modelo Rothschild-Stiglitz.

Figura 5.7 – Equilíbrio agregador no modelo Rothschild-Stiglitz



Agora vamos ver apenas o equilíbrio separador, como na Figura 5.8. Novamente, sabemos que os tipos de alto risco permanecerão em α^H. Torna-se natural ver que o ponto α^L é onde os agentes do tipo de risco baixo vão permanecer. A combinação (α^L, α^H) é um equilíbrio de Nash do modelo Rothschild-Stiglitz de fato? Precisamos impor outra restrição aqui, ou seja, a restrição de compatibilidade de incentivo. O contrato α^L para os tipos de baixo risco deve satisfazer as três condições seguintes para se qualificar para um equilíbrio: (1) α^L deve estar na ou abaixo da curva de iso-utilidade do tipo de alto risco; (2) α^L deve estar no tipo de baixo risco para garantir que as companhias de seguros obtenham lucro zero, fornecendo o contrato α^L; (3) α^L deve atingir o nível máximo de utilidade possível para os tipos de baixo risco. É basicamente dizer que os tipos de baixo risco estarão dispostos a permanecer em α^L e os tipos de alto risco estarão dispostos a permanecer em α^H, e assim efetivamente revelar seus verdadeiros tipos de risco para as companhias de seguro.

Figura 5.8 – Equilíbrio separador no modelo Rothschild-Stiglitz



Agora é hora de relacionar este artigo ao artigo de Spence. Como é que o papel de Spence oferece uma quantidade infinita de equilíbrios, enquanto este artigo fornece apenas um equilíbrio, se é que existe? Acontece que é devido à diferença na configuração: embora a condição de equilíbrio do mercado de trabalho em Spence seja semelhante à condição de lucro esperado zero no modelo Rothschild-Stiglitz, existem infinitas maneiras de pagar salários para os trabalhadores em Spence, e os trabalhadores não estão limitados a escolher o salário que maximiza a utilidade dos trabalhadores.

Note que tanto no modelo de Spence quanto no modelo Rothschild-Stiglitz, o agente desinformado se move primeiro. Neste tipo de jogo, o pré-comprometimento dos desinformados é crítico. Embora exista um forte incentivo para o agente desinformado renegociar o contrato com o tipo melhor entre os agentes informados, eles não podem permitir isso. Caso contrário, o pior tipo de agentes informados não estará mais disposto a tomar sua melhor solução e efetivamente estragar o resultado de equilíbrio.

Uma lição importante e muito realista do modelo de Rothschild-Stiglitz é que a desregulamentação leva à diferenciação do produto, pois não existe um equilíbrio comum em um mercado competitivo. Aqui está um exemplo sobre o setor bancário. Antes da desregulamentação, os bancos não precisavam pagar juros sobre os depósitos, mas ofereciam gratuitamente serviços de administração de caixa, ou seja, os bancos usavam lucros de depósitos para subsidiar os serviços de administração de caixa. Depois que a desregulamentação foi implementada, esse tipo de prática de agrupamento não funcionava mais, pois muitos clientes que não precisavam de serviços de gerenciamento de caixa mudavam para bancos que ofereciam juros sobre depósitos.

5.4 Moral Hazard

Anteriormente, enfatizamos que a delegação de tarefas cria uma lacuna de informação entre o principal e seu agente quando este aprende alguma informação relevante para determinar o volume eficiente do comércio. A seleção adversa não é o único problema informacional que se pode imaginar. Os agentes também podem escolher ações que afetam o valor do comércio ou, mais geralmente, o desempenho do agente. O principal muitas vezes perde a capacidade de controlar as ações que não são mais observáveis, seja pelo principal que oferece o contrato ou pelo tribunal que o aplica. Em tais casos, diremos que há risco moral.

Os principais candidatos para essas ações de risco moral são as variáveis de esforço, que influenciam positivamente o nível de produção do agente, mas também criam uma desutilidade para o agente. Por exemplo, o rendimento de um campo depende da quantidade de tempo que o inquilino gastou selecionando as melhores culturas, ou a qualidade de sua colheita. Da mesma forma, a probabilidade de um motorista sofrer um acidente de carro depende de sua segurança, o que também afeta sua demanda por seguro. Além disso, uma empresa regulada pode ter que realizar um investimento dispendioso e não observável para reduzir seu custo de produzir um bem socialmente valioso.

Como no caso da seleção adversa, a informação assimétrica também desempenha um papel crucial no projeto do contrato de incentivo ótimo sob risco moral. No entanto, em vez de ser uma incerteza exógena para o principal, a incerteza é agora endógena. As probabilidades dos diferentes estados da natureza e, portanto, o volume esperado de comércio, agora dependem explicitamente do esforço do agente. Em outras palavras, o nível de produção realizado é apenas um sinal ruidoso da ação do agente. Essa incerteza é fundamental para entender o problema contratual sob risco moral. Se o mapeamento entre esforço e desempenho fosse completamente determinístico, o principal e o tribunal de justiça não teriam dificuldade em inferir o esforço do agente a partir do resultado observado. Mesmo que o esforço do agente não fosse observável diretamente, ele poderia ser indiretamente contraído, já que a produção seria ela mesma observável e verificável.

5.4.1 Esforço e Produção

Consideramos um agente que pode exercer um esforço dispendioso e. Dois valores possíveis podem ser tomados por e, que nós normalizamos como um nível de esforço zero e um esforço positivo de um: $e \in \{0, 1\}$. Exercer esforço e implica uma desutilidade para o agente que é igual a $\psi(e)$ com a normalização $\psi(0) = \psi_0 = 0$ e $\psi_1 = \psi$.

O agente recebe uma transferência t do principal. Assumimos que sua função de utilidade é separável entre moeda e esforço, $U = u(t) - \psi(e)$, com $u(\cdot)$ crescente e côncava ($u' > 0$ e $u'' < 0$). Às vezes, usaremos a função $h = u^{-1}$, a função inversa de $u(\cdot)$, que é crescente e convexa ($h' > 0$ e $h'' > 0$).

A produção é estocástica, e o esforço afeta o nível de produção da seguinte forma: o nível de produção estocástico $\widetilde{q}$ só pode ter dois valores $\{\underline{q}, \overline{q}\}$, com $\overline{q} - \underline{q} = \Delta q > 0$, e a influência estocástica do esforço na produção é caracterizada pelas probabilidades $\Pr(\widetilde{q} = \underline{q}|e = 0) = \pi_0$, e $\Pr(\widetilde{q} = \underline{q}|e = 1) = \pi_1$, com $\pi_1 > \pi_0$. Vamos denotar a diferença entre essas duas probabilidades por $\Delta\pi = \pi_1 - \pi_0$.

Observe que o esforço melhora a produção no sentido de dominância estocástica de primeira ordem, ou seja, $\Pr(\widetilde{q} \leqq q^*|e)$ é decrescente com e para qualquer produção q^*.

5.4.2 Contratos Viáveis de Incentivo

Como a ação do agente não é diretamente observável pelo principal, o principal só pode oferecer um contrato com base no nível de produção observável e verificável, ou seja, uma função $\{t(\widetilde{q})\}$ vinculando a compensação do agente ao produto aleatório $\widetilde{q}$. Com dois resultados possíveis $\overline{q}$ e $\underline{q}$, o contrato pode ser definido de forma equivalente por um par de transferências $\overline{t}$ e $\underline{t}$. A transferência $\overline{t}$ é o pagamento recebido pelo agente se a produção $\overline{q}$) for realizada.

A utilidade esperada do principal (neutro ao risco) é escrito como

$$V_1 = \pi_1(S(\overline{q}) - \overline{t}) + (1 - \pi_1)(S(\underline{q}) - \underline{t}), \tag{5.63}$$

se o agente faz um esforço positivo $e = 1$ e

$$V_0 = \pi_0(S(\overline{q}) - \overline{t}) + (1 - \pi_0)(S(\underline{q}) - \underline{t}), \tag{5.64}$$

se o agente faz um esforço positivo $e = 0$. Por simplicidade, denotaremos os benefícios do principal em cada estado de natureza por $S(\overline{q}) = \overline{S}$ e $S(\underline{q}) = \underline{S}$.

Cada nível de esforço que o principal deseja induzir corresponde a um conjunto de contratos que garantem a restrição de compatibilidade de incentivo por risco moral e a restrição de participação:

$$\pi_1 u(\overline{t}) + (1 - \pi_1)u(\underline{t}) - \psi \geqq \pi_0 u(\overline{t}) + (1 - \pi_0)u(\overline{t}), \tag{5.65}$$

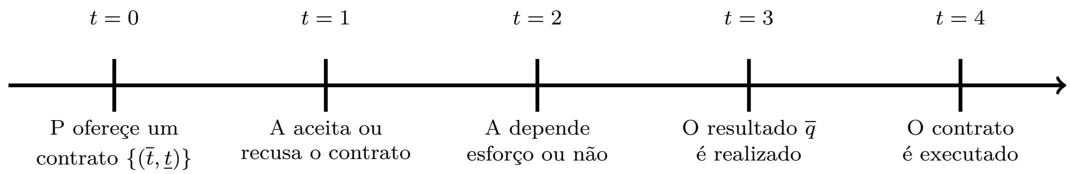
$$\pi_1 u(\overline{t}) + (1 - \pi_1)u(\underline{t}) - \psi \geqq 0. \tag{5.66}$$

Observe que a restrição de participação é assegurada no estágio ex ante, ou seja, antes da realização do choque de produção.

Definição 5.4.1. *Um contrato viável de incentivo satisfaz as restrições de compatibilidade e participação de incentivo dado pelas equações acima.*

O timing do jogo de contratação sob risco moral é resumido na figura abaixo.

Figura 5.9 – Timing de contratação sob risco moral



5.4.3 Contrato Ótimo

Como referência, vamos primeiro supor que o principal e um tribunal benevolente possam observar o esforço. Então, se ele quiser induzir esforço, o problema do principal se torna

$$\max_{\{(\bar{t}, \underline{t})\}} \; \pi_1(\overline{S} - \bar{t}) + (1 - \pi_1)(\underline{S} - \underline{t}) \tag{5.67}$$
$$\text{sujeito a } \pi_1 u(\bar{t}) + (1 - \pi_1) u(\underline{t}) - \psi \geqq 0.$$

De fato, apenas a restrição de participação de agentes é importante para o principal, porque o agente pode ser forçado a exercer um nível positivo de esforço. Se o agente não estivesse escolhendo esse nível de esforço, o agente poderia ser altamente punido e o tribunal de justiça poderia se comprometer a aplicar tal punição.

Denotando o multiplicador dessa restrição de participação por λ e otimizando em relação a $\bar{t}$ e $\underline{t}$, temos as seguintes condições de primeira ordem:

$$-\pi_1 + \lambda \pi_1 u'(\bar{t}^*) = 0, \tag{5.68}$$
$$-(1 - \pi_1) + \lambda(1 - \pi_1) u'(\underline{t}^*) = 0. \tag{5.69}$$

em que $\bar{t}^*$ e $\underline{t}^*$ são as transferências first-best.

As condições de primeira ordem implicam que $\lambda = \dfrac{1}{u'(\underline{t}^*)} = \dfrac{1}{u'(\bar{t}^*)} > 0$ e $t^* = \bar{t}^* = \underline{t}^*$.

Assim, com um esforço verificável, o agente obtém um seguro total do principal e a transferência t^* que recebe é a mesma, independentemente do estado de natureza. Como a restrição de participação é vinculativa, também obtemos o valor dessa transferência, que é suficiente apenas para cobrir a desutilidade do esforço, ou seja, $t^* = h(\psi)$. Este é também o pagamento esperado feito pelo principal ao agente, ou o custo de first-best de implementar o nível de esforço positivo, C^{FB}.

Para o principal, o esforço indutor produz um retorno esperado igual a

$$V_1 = \pi_1 \overline{S} + (1 - \pi_1)\underline{S} - h(\psi). \tag{5.70}$$

Se o principal decidisse deixar o agente não fazer nenhum esforço, e_0, ele faria um pagamento zero ao agente, qualquer que fosse a realização da produção. Nesse cenário, o principal obteria um pagamento igual a

$$V_0 = \pi_0 \overline{S} + (1 - \pi_0)\underline{S}. \tag{5.71}$$

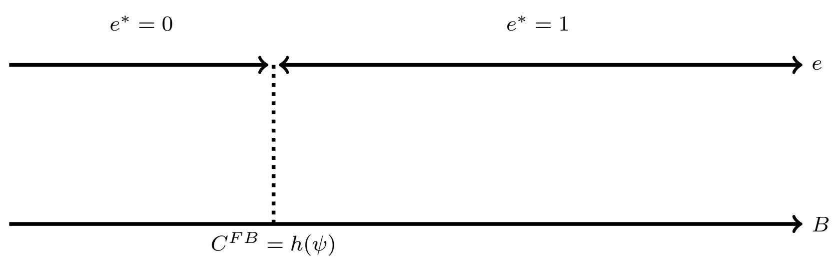
O esforço indutor é, portanto, ótimo do ponto de vista do principal quando $V_1 \geqq V_0$, isto é, $\pi_1 \overline{S} + (1 - \pi_1)\underline{S} - h(\psi) \geqq \pi_0 \overline{S} + (1 - \pi_0)\underline{S}$, ou, em outras palavras, quando o ganho de efeito esperado é maior do que o custo de induzir o custo de first-best, isto é,

$$\Delta\pi\Delta S \geqq h(\psi), \tag{5.72}$$

em que $\Delta S = \overline{S} - \underline{S} > 0$.

Denotando o benefício de induzir um nível de esforço estritamente positivo por $B = \Delta\pi\Delta S$, o primeiro melhor resultado requer $e^* = 1$ se e somente se $B > h(\psi)$, como mostrado na figura abaixo.

Figura 5.10 – Nível de esforço de fist-best



5.4.4 Neutralidade ao Risco

Se o agente é neutro ao risco, temos (até uma transformação afim) $u(t) = t$ para todo t e $h(u) = u$ para todo u. O principal que deseja induzir esforço deve, portanto, escolher o contrato que resolve o seguinte problema:

$$\max_{\{(\overline{t},\underline{t})\}} \ \pi_1(\overline{S} - \overline{t}) + (1 - \pi_1)(\underline{S} - \underline{t}) \tag{5.73}$$

$$\text{sujeito a } \pi_1 \overline{t} + (1 - \pi_1)\underline{t} - \psi \geqq \pi_0 \overline{t} + (1 - \pi_0)\underline{t} \tag{5.74}$$

$$\pi_1 \overline{t} + (1 - \pi_1)\underline{t} - \psi \geqq 0. \tag{5.75}$$

Com a neutralidade de risco, o principal pode, por exemplo, escolher transferências compatíveis de incentivo $\overline{t}$ e $\underline{t}$, que tornam a restrição de participação do agente vinculativa e não deixam nenhum espaço para o agente. Com efeito, resolvendo as restrições com igualdades, obtemos imediatamente

$$\underline{t}^* = -\frac{\pi_0}{\Delta\pi}\psi, \tag{5.76}$$

$$\overline{t}^* = \frac{1-\pi_0}{\Delta\pi}\psi = \underline{t}^* + \frac{1}{\Delta\pi}\psi. \tag{5.77}$$

O agente é recompensado se a produção for alta. Sua utilidade líquida neste estado de natureza é $\overline{U}^* = \overline{t}^* - \psi = \frac{1-\pi_1}{\Delta\pi}\psi > 0$. Por outro lado, o agente é punido se a produção é baixa. Sua utilidade líquida neste estado de natureza é $\overline{U}^* = \underline{t}^* - \psi = -\frac{\pi_1}{\Delta\pi}\psi < 0$.

O principal faz um pagamento esperado $\pi_1\overline{t}^* + (1-\pi_1)\underline{t}^* = \psi$, que é igual à desutilidade do esforço que ele incorreria se pudesse controlar perfeitamente o nível de esforço. O principal pode estruturar sem custos o pagamento do agente, para que este tenha os incentivos certos para exercer esforço. Usando as expressões para $\underline{t}^*$ e $\overline{t}^*$, seu ganho esperado de exercer esforço é, portanto, $\Delta\pi(\overline{t}^* - \underline{t}^*) = \psi$ ao aumentar seu esforço de $e = 0$ para $e = 1$.

Proposição 5.4.1. *O risco moral não é um problema com um agente neutro ao risco, apesar da não observância do esforço. O nível de esforço de first-best ainda é implementado.*

Quando a contratação ocorre ex ante, a restrição de incentivo, sob seleção adversa ou risco moral, não entra em conflito com a restrição de participação ex ante com um agente neutro ao risco, e o primeiro melhor resultado ainda é implementado.

Ineficiências na provisão de esforço devido ao risco moral surgirão quando o agente não estiver mais neutro ao risco. Existem duas maneiras alternativas de modelar esses custos de transação. Uma é manter a neutralidade do risco para níveis positivos de renda, mas impor uma restrição de responsabilidade limitada, que exige que as transferências não sejam muito negativas. A outra é deixar o agente ser estritamente avesso ao risco. A seguir, analisamos esses dois ambientes contratuais e os diferentes trade-offs que eles implicam.

5.4.5 Aversão ao Risco

Suponha que o agente é avesso ao risco. O problema do agente é escrito como:

$$\max_{\{(\overline{t},\underline{t})\}} \quad \pi_1(\overline{S} - \overline{t}) + (1-\pi_1)(\underline{S} - \underline{t}) \tag{5.78}$$

$$\text{sujeito a } \pi_1\overline{t} + (1-\pi_1)\underline{t} - \psi \geqq \pi_0\overline{t} + (1-\pi_0)\underline{t} \tag{5.79}$$

$$\pi_1\overline{t} + (1-\pi_1)\underline{t} - \psi \geqq 0. \tag{5.80}$$

Como o problema de otimização do principal pode não ser um programa côncavo para o qual as condições de Kuhn-Tucker de primeira ordem são necessárias e suficientes, fazemos a seguinte mudança de variáveis. Defina $\overline{u} = u(\overline{t})$ e $\underline{u} = u(\underline{t})$ ou, de forma equivalente, seja $\overline{t} = h(\overline{u})$ e $\underline{t} = u(\underline{h})$. Essas novas variáveis são os níveis de utilidade ex post obtidos pelo agente em ambos os estados da natureza. O conjunto de contratos viáveis de incentivo pode agora ser descrito por duas restrições lineares:

$$\pi_1\overline{u} + (1-\pi_1)\underline{u} - \psi \geqq \pi_0\overline{u} + (1-\pi_0)\underline{u}, \tag{5.81}$$
$$\pi_1\overline{u} + (1-\pi_1)\underline{u} - \psi \geqq 0. \tag{5.82}$$

Então, o problema do agente pode ser reescrito como

$$\max_{\{(\overline{u},\underline{u})\}} \ \pi_1(\overline{S} - h(\overline{u})) + (1-\pi_1)(\underline{S} - h(\underline{u})) \tag{5.83}$$
$$\text{sujeito a} \quad \pi_1\overline{u} + (1-\pi_1)\underline{u} - \psi \geqq \pi_0\overline{u} + (1-\pi_0)\underline{u} \tag{5.84}$$
$$\pi_1\overline{u} + (1-\pi_1)\underline{u} - \psi \geqq 0. \tag{5.85}$$

Observe que a função objetivo do principal agora é estritamente côncava em $(\overline{u}, \underline{u})$ porque $h(\cdot)$ é estritamente convexa. As restrições são agora lineares e o interior do conjunto restrito é obviamente não vazio.

Observe que o agente avesso ao risco não recebe mais o seguro total. De fato, com seguro total, a restrição de compatibilidade de incentivo não pode mais ser satisfeita. O esforço indutor requer que o agente tenha algum risco, a seguinte proposição fornece um resumo.

Proposição 5.4.2. *Quando o agente é estritamente avesso ao risco, o contrato ideal que induz o esforço torna a participação do agente e as restrições de incentivo vinculantes. Este contrato não oferece seguro total. Além disso, as segundas melhores transferências são dadas por*

$$\bar{t}^{SB} = h\left(\psi + (1-\pi_1)\frac{\psi}{\Delta\pi}\right) = h\left(\frac{1-\pi_0}{\Delta\pi}\psi\right), \tag{5.86}$$
$$\underline{t}^{SB} = h\left(\psi - \pi_1\frac{\psi}{\Delta\pi}\right) = h\left(-\frac{\pi_0}{\Delta\pi}\psi\right). \tag{5.87}$$

Vamos agora nos voltar para a questão da segunda melhor otimização para induzir um alto esforço, do ponto de vista do principal. O segundo melhor custo C^{SB} de induzir esforço sob risco moral é o pagamento esperado feito ao agente $C^{SB} = \pi_1\bar{t}^{SB} + (1-\pi_1)\underline{t}^{SB}$. Esse custo é reescrito como

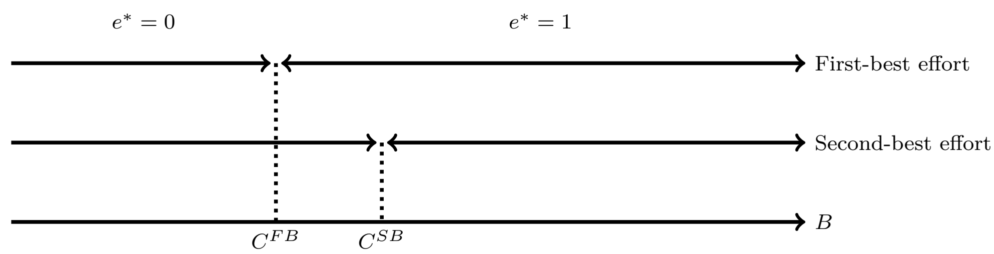
$$C^{SB} = \pi_1 h\left(\frac{1-\pi_0}{\Delta\pi}\psi\right) + (1-\pi_1)h\left(-\frac{\pi_0}{\Delta\pi}\psi\right). \tag{5.88}$$

O benefício de induzir esforço é $B = \Delta\pi\Delta S$, e um esforço positivo $e^* = 1$ é a escolha ótima do principal sempre que

$$\Delta\pi\Delta S \geqq C^{SB} = \pi_1 h\left(\frac{1-\pi_0}{\Delta\pi}\psi\right) + (1-\pi_1)h\left(-\frac{\pi_0}{\Delta\pi}\psi\right). \tag{5.89}$$

Com $h(\cdot)$ sendo estritamente convexa, a desigualdade de Jensen implica que o custo de second-best é estritamente maior do que o custo de first-best de implementação de esforço. Portanto, a indução de um esforço maior ocorre com menos frequência com risco moral do que quando o esforço é observável. A figura abaixo representa esse fenômeno graficamente. Para B pertencente ao intervalo $[C^{FB}, C^{SB}]$, o segundo melhor nível de esforço é zero e, portanto, está estritamente abaixo de seu primeiro melhor valor. Existe agora um subfornecimento de esforços por causa do risco moral e da aversão ao risco.

Figura 5.11 – Segundo melhor nível de esforço com risco moral e aversão ao risco



Proposição 5.4.3. *Com risco moral e aversão ao risco, há um trade-off entre induzir o esforço e fornecer seguro ao agente. Em um modelo com dois níveis possíveis de esforço, o principal induz um esforço positivo do agente com menos frequência do que quando o esforço é observável.*

5.4.6 Salário-Eficiência

Vamos considerar um agente neutro ao risco trabalhando para uma empresa, o principal. Este é um modelo básico estudado por Shapiro e Stiglitz (AER, 1984). Ao exercer o esforço $e \in \{0, 1\}$, o valor agregado da empresa é $\overline{V}$ (resp. $\underline{V}$) com probabilidade $\pi(e)$ (resp. $1 - \pi(e)$). O agente só pode ser recompensado por um bom desempenho e não pode ser punido por um mau resultado, uma vez que eles são protegidos por responsabilidade limitada.

Para induzir esforço, o principal deve encontrar um esquema de compensação ideal $\{(\bar{t}, \underline{t})\}$ que seja a solução para o programa abaixo:

$$\max_{\{(\bar{t},\underline{t})\}} \ \pi_1(\overline{V} - \bar{t}) + (1 - \pi_1)(\underline{V} - \underline{t}) \tag{5.90}$$

$$\text{sujeito a } \pi_1 \bar{t} + (1 - \pi_1)\underline{t} - \psi \geqq \pi_0 \bar{t} + (1 - \pi_0)\underline{t} \tag{5.91}$$

$$\pi_1 \bar{t} + (1 - \pi_1)\underline{t} - \psi \geqq 0 \tag{5.92}$$

$$t \geqq 0. \tag{5.93}$$

O problema é completamente isomórfico ao analisado anteriormente. A restrição de responsabilidade limitada é vinculativa no ótimo, e a empresa escolhe induzir um alto esforço quando $\Delta\pi\Delta V \geqq \dfrac{\pi_1 \psi}{\Delta\pi}$. O salário positivo $\bar{t}^{SB} = \dfrac{\psi}{\Delta\pi}$ é frequentemente chamado de salário de eficiência porque induz o agente a exercer um nível de esforço alto (eficiente). Para induzir a produção, o principal deve desistir de uma parte positiva do lucro da empresa para o agente.

Referências Bibliográficas

[Akerlof1970] Akerlof, G. (1970). The Market for Lemons: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 89(4):488–500.

[Bemanke and Gertler1989] Bemanke, B. and Gertler, M. (1989). Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, 79(1):14–31.

[Bolton et al.2005] Bolton, P., Dewatripont, M., et al. (2005). *Contract Theory*. MIT press.

[CDATA-Tirole1988] CDATA-Tirole, J. (1988). The Theory of Industrial Organization.

[Engers and Fernandez1987] Engers, M. and Fernandez, L. (1987). Market Equilibrium with Hidden Knowledge and Self-Selection. *Econometrica*, 55:425–439.

[Green and Kahn1983] Green, J. and Kahn, C. M. (1983). Wage-Employment Contracts. *Quarterly Journal of Economics*, 98:173–187.

[Grossman and Hart1992] Grossman, S. J. and Hart, O. D. (1992). An Analysis of the Principal-Agent Problem. *Econometrica*, 51:7–45.

[Hart1983] Hart, O. D. (1983). Optimal Labour Contracts under Asymmetric Information: An Introduction. *The Review of Economic Studies*, 50(1):3–35.

[Holmstrom et al.1979] Holmstrom, B. et al. (1979). Moral Hazard and Observability. *Bell Journal of Economics*, 10(1):74–91.

[Kreps et al.1990] Kreps, D. M. et al. (1990). *A Course in Microeconomic Theory*. Princeton University Press.

[Laffont and Martimort2009] Laffont, J.-J. and Martimort, D. (2009). *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

[Laffont and Tirole1993] Laffont, J.-J. and Tirole, J. (1993). *A Theory of Incentives in Procurement and Regulation*. MIT press.

[Maskin and Riley1984] Maskin, E. and Riley, J. (1984). Monopoly with Incomplete Information. *The RAND Journal of Economics*, 15(2):171–196.

[Mussa and Rosen1978] Mussa, M. and Rosen, S. (1978). Monopoly and Product Quality. *Journal of Economic Theory*, 18(2):301–317.

[Rothschild and Stiglitz1976] Rothschild, M. and Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets. *Quarterly Journal of Economics*, 93:541–562.

[Salanié2005] Salanié, B. (2005). *The Economics of Contracts: A Primer*. MIT press.

[Spence1973] Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics*, 87:355–374.

[Stiglitz1974] Stiglitz, J. E. (1974). Incentives and Risk Sharing in Sharecropping. *Review of Economic Studies*, 41(2):219–255.

[Stiglitz1977] Stiglitz, J. E. (1977). Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies*, 44(3):407–430.

[Stiglitz and Weiss1981] Stiglitz, J. E. and Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *American Economic Review*, 71(3):393–410.

[Wolfstetter1999] Wolfstetter, E. (1999). *Topics in Microeconomics: Industrial Organization, Auctions, and Incentives*. Cambridge University Press.

Capítulo 6

Teoria dos Jogos

Contents

6.1 Introdução

A teoria dos jogos é uma metodologia formal e um conjunto de técnicas para estudar a interação de agentes racionais em cenários estratégicos. "Racional" aqui significa a coisa padrão em economia: maximizar objetivos. "Estratégico" significa que os agentes não se preocupam apenas com suas próprias ações, mas também com as ações tomadas por outros agentes. Observe que a teoria da decisão é o estudo de como um indivíduo toma decisões em ambientes não estratégicos. Portanto, a teoria dos jogos às vezes também é chamada de teoria da decisão de várias pessoas. A terminologia comum para o campo vem de suas aplicações putativas para jogos como pôquer, xadrez, etc. No entanto, as aplicações nos quais estamos geralmente interessados pouco têm a ver diretamente com esses jogos. Em particular, esses são os jogos que chamamos de "soma zero", no sentido de que a perda de um jogador é o ganho de outro jogador; eles são jogos de puro conflito. Em aplicações econômicas, normalmente há uma mistura de conflitos e motivos de cooperação.

A teoria moderna dos jogos como um campo deve muito ao trabalho de John von Neumann. Em 1928, ele escreveu um artigo importante sobre os jogos de soma zero de duas pessoas que continham o famoso Teorema Minimax, que veremos mais adiante. Em 1944, von Neumann e Oscar Morgenstern publicaram seu livro clássico, "Theory of Games and Strategic Behavior", que estendeu o trabalho em jogos de soma zero, e também iniciou a teoria dos jogos cooperativos. No início da década de 1950, John Nash fez suas contribuições seminais para jogos de soma não-zero e começou a teoria da barganha. Em 1957, Robert Luce e Howard Raiffa publicaram seu livro, "Games and Decisions: Introduction and Critical Survey", popularizando a teoria dos jogos. Em 1967-1968, John Harsanyi formalizou métodos para estudar jogos de informações incompletas, o que foi crucial para ampliar o escopo das aplicações. Na década de 1970, houve uma explosão de trabalho teórico e aplicado na teoria dos jogos, e a metodologia foi bem longe de seu status atual como uma ferramenta proeminente não apenas na economia, mas também em outras ciências sociais.

Ao longo deste curso, vamos nos concentrar na teoria dos jogos não-cooperativos, em oposição à teoria dos jogos cooperativos. Toda a teoria dos jogos descreve as configurações estratégicas começando com o conjunto de jogadores, ou seja, os tomadores de decisão. A diferença entre teoria cooperativa e não-cooperativa é que a primeira toma as ações individuais de cada jogador como primitivas, enquanto a segunda toma ações conjuntas como primitivas. Ou seja, a teoria cooperativa dos jogos pressupõe que os acordos vinculativos podem ser feitos pelos jogadores dentro de vários grupos e os jogadores podem se comunicar livremente para fazê-lo. Vamos tomar o ponto de vista não-cooperativo de que cada jogador age como um indivíduo, e as possibilidades de acordos e comunicação devem ser explicitamente modeladas.

Note que, uma vez que as preferências de um jogador em suas ações dependem de quais ações as outras partes tomam, sua ação depende de suas crenças sobre o que os outros fazem. Claro, o que os outros fazem depende de suas crenças sobre o que cada jogador faz. Desta forma, a ação de um jogador, em princípio, depende das ações disponíveis para cada jogador, preferências de cada jogador nos resultados, crenças de cada jogador sobre quais ações estão disponíveis para cada jogador e como cada jogador classifica os resultados, além de suas crenças sobre as crenças de cada jogador, ad infinitum.

Quando os jogadores pensam no que os outros jogadores farão, levando em conta o que os outros jogadores pensam sobre eles, eles podem encontrar um jeito claro de jogar o jogo. Considere o seguinte "jogo":

		Jogador 2		
		L	m	R
	T	$(1,1)$	$(0,2)$	$(2,1)$
Jogador 1	M	$(2,2)$	$(1,1)$	$(0,0)$
	B	$(1,0)$	$(0,0)$	$(-1,1)$

Aqui, há dois jogadores, o jogador 1 e o jogador 2. O jogador 1 tem estratégias, T, M, B e o jogador 2 tem estratégias L, m, R. Eles escolhem suas estratégias simultaneamente.

Cada par de estratégias leva a um retorno (payoff) para cada jogador, uma recompensa medida por um número real. Em cada entrada, o primeiro número é a recompensa do jogador 1, e a segunda entrada é a recompensa do jogador 2. Por exemplo, se o jogador 1 jogar T e o jogador 2 jogar R, então o jogador 1 recebe um pagamento de 2 e o jogador 2 obtém um payoff de 1. Vamos supor que cada jogador saiba que estas são as estratégias e os payoffs, cada jogador sabe que cada jogador sabe disso, cada jogador sabe que cada jogador sabe que cada jogador sabe disso. . . ad infinitum.

Agora, o jogador 1 analisa seus pagamentos e percebe que, não importa o que o outro jogador jogue, é melhor que ele jogue M ao invés de B. Isto é, se o jogador 2 joga L, M dá 2 e B dá 1; se o jogador 2 joga m, M dá 1, B dá 0; e se o jogador 2 joga R, M dá 0, B dá -1. Portanto, ele percebe que não deve jogar B. Agora ele compara T e M. Ele percebe que, se o jogador 2 joga L ou m, M é melhor que T, mas se ela joga R, T definitivamente é melhor que M. O jogador 2 joga R? Para encontrar uma resposta a estas perguntas, o jogador 1 analisa o jogo do ponto de vista do jogador 2. Ele percebe que, para o jogador 2, não há estratégia que seja melhor do que qualquer outra estratégia. Por exemplo, R é a melhor estratégia se o jogador 1 jogar B, mas caso contrário é estritamente pior que m. O jogador 2 pensaria que o jogador 1 jogaria B? Bem, ele sabe que o jogador 1 está tentando maximizar o retorno esperado, dado pelas primeiras entradas, como todos sabem. Ele deve então deduzir que o jogador 1 não jogará B. Portanto, o jogador 1 conclui, ele não jogará R (como é pior que m neste caso). Excluindo a possibilidade de que o jogador 2 jogue R, o jogador 1 analisa seus pagamentos e vê que M é agora melhor que T, não importa o que aconteça. Por outro lado, o jogador 2 passa por um raciocínio semelhante e conclui que o jogador 1 deve jogar M e, portanto, joga.

Imagine que você quer se encontrar com um amigo em um dos dois lugares, sobre os quais ambos são indiferentes. Infelizmente, você não pode se comunicar entre si até se encontrar. Esta situação é formalizada no jogo seguinte, que é chamado jogo de coordenação pura:

		Jogador 2	
		Esquerda	Direita
Jogador 1	Superior	$(1,1)$	$(0,0)$
	Inferior	$(0,0)$	$(1,1)$

Aqui, o jogador 1 escolhe entre as linhas superior e inferior, enquanto o jogador 2 escolhe entre as colunas esquerda e direita. Em cada caixa, o primeiro e o segundo números indicam os payoffs dos jogadores 1 e 2, respectivamente. Note que o jogador 1 prefere de superior para inferior se ele sabe que o jogador 2 joga à esquerda; ele prefere inferior se ele sabe que o jogador 2 joga direita. Da mesma forma, o jogador 2 prefere esquerda se ele sabe que o jogador 1 joga com superior. Não há previsão clara sobre o resultado deste jogo.

Pode-se procurar os resultados estáveis (perfis estratégicos) no sentido de que nenhum jogador tem incentivo para desviar se ele souber que os outros jogadores jogam as estratégias prescritas. (Esses perfis de estratégia são chamados de equilíbrio de Nash, nomeado por John Nash.) Aqui, o superior esquerdo e o inferior direito são esses resultados. Mas o abaixo-esquerda e o acima-direita não são estáveis nesse sentido. Por exemplo, se o abaixo-esquerda é conhecido por ser jogado, cada jogador gostaria de se desviar.

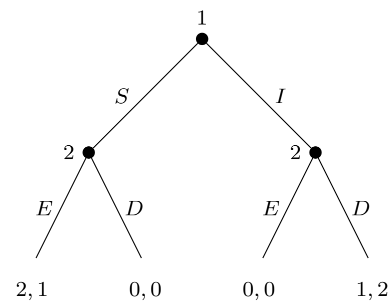
Ao contrário deste jogo, a maioria dos jogadores tem diferentes preferências nos resultados, induzindo conflitos. No jogo seguinte, conhecido como a Batalha dos Sexos, o conflito e a necessidade de coordenação estão presentes juntos.

		Jogador 2	
		Esquerda	Direita
Jogador 1	Superior	$(2,1)$	$(0,0)$
	Inferior	$(0,0)$	$(1,2)$

Aqui, mais uma vez os jogadores gostariam de coordenar na parte superior esquerda ou inferior direita, mas agora o jogador 1 prefere coordenar na parte superior esquerda, enquanto o jogador 2 prefere coordenar na parte inferior direita. Os resultados estáveis são novamente superior esquerdo e inferior direito.

A análise acima assume que os jogadores realizam suas ações simultaneamente, de modo que um jogador não observa a ação tomada pelos outros quando escolhe sua própria ação. Em geral, um jogador pode observar algumas das ações de alguns outros jogadores. Tal conhecimento pode ter um impacto dramático no resultado do jogo. Para uma ilustração, na Batalha dos Sexos, imagine que o Jogador 2 sabe o que o Jogador 1 faz quando ele toma sua ação. Isso pode ser formalizado por meio da árvore na Figura 6.1. Aqui, o Jogador 1 escolhe entre superior e inferior e o Jogador 2 escolhe entre esquerda e direita sabendo o que o Jogador 1 escolheu. Claramente, agora o Jogador 2 escolheria esquerda se o Jogador 1 jogasse superior, e escolheria direita se o Jogador 1 jogasse inferior. Sabendo disso, o jogador 1 jogaria superior. Portanto, pode-se argumentar que o único resultado razoável deste jogo é superior-esquerda. Esse tipo de raciocínio é chamado de indução retroativa.

Figura 6.1 – Batalha dos sexos com movimentos sequenciais



Outra interpretação é que o Jogador 1 pode se comunicar com o Jogador 2, que não pode se comunicar com o jogador 1. Isso permite que o Jogador 1 se comprometa com suas ações, fornecendo uma posição forte na relação.

Imagine que, antes de jogar a Batalha dos Sexos, o jogador 1 tem a opção de sair, caso em que cada jogador receberá $\frac{3}{2}$, ou joga a Batalha dos Sexos. Quando solicitado a jogar, o Jogador 2 saberá que o Jogador 1 escolheu jogar a Batalha dos Sexos. Existem dois equilíbrios "razoáveis" (ou resultados estáveis). Uma é que o Jogador 1 sai, pensando que, se ele jogar a Batalha dos Sexos, eles jogarão o equilíbrio inferior-direita da Batalha dos Sexos, rendendo apenas 1 para o Jogador 1. O segundo é que o Jogador 1 escolhe Jogar o Batalha de Sexos, e na Batalha dos Sexos eles jogam o equilíbrio superior-esquerda.

Alguns argumentariam que o primeiro resultado não é realmente razoável? Porque, quando solicitado a jogar, o Jogador 2 saberá que o Jogador 1 escolheu jogar a Batalha dos Sexos, renunciando ao pagamento de $\frac{3}{2}$. Ele deve, portanto, perceber que o jogador 1 não pode estar planejando jogar inferior, o que gera o pagamento de 1. Ou seja, quando solicitado a jogar, o Jogador 2 deve entender que o Jogador 1 está planejando jogar superior e, portanto, deve jogar esquerda. Antecipando isso, o Jogador 1 deve escolher jogar o jogo Batalha dos Sexos, no qual eles jogam supeior-esquerda. Portanto, o segundo resultado é o único razoável.

A análise anterior é um exemplo de forward induction, pois estamos requerendo que os jogadores levem em conta não somente o que eles esperam que acontecerá após as suas ações, mas também o que eles esperam que tenha ocorrido antes das suas jogadas.

Aqui estão mais alguns exemplos de jogos que serão referidos com frequência ao longo dessas notas.

1. Dilema dos Prisioneiros

		Jogador 2	
		Cooperar	Dificultar
Jogador 1	Cooperar	$(5, 5)$	$(0, 6)$
	Dificultar	$(6, 0)$	$(1, 1)$

Este é um jogo bem conhecido que a maioria de vocês conhece. Dois prisioneiros são presos por um crime para o qual não há provas firmes e estão sendo interrogados em salas separadas.

Cada prisioneiro pode cooperar com o outro e não confessar seu crime ou defeito e confessar o crime. Neste jogo, não importa o que o outro jogador faça, cada jogador gostaria de desertar, confessando seu crime. Isso produz $(1, 1)$. Se ambos cooperassem e não confessassem o crime, cada um receberia um pagamento melhor de 5.

2. Jogo do falcão e do pombo

		Jogador 2	
		Falcão	Pombo
Jogador 1	Falcão	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
	Pombo	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

Este é um importante jogo biológico, mas também é bastante semelhante a muitos jogos em Economia e Ciência Política. V é o valor de um recurso que um dos jogadores irá utilizar. Se eles compartilham o recurso, seus valores são $\dfrac{V}{2}$. Falcão significa uma estratégia "dura", em que o jogador não desiste do recurso. No entanto, se o outro jogador também estiver jogando falcão, ele acabará lutando e incorrerá no custo $\dfrac{C}{2}$. Por outro lado, um jogador de Falcão obtém todo o recurso para si mesmo ao jogar uma Pombo. Quando $V > C$, este é um jogo do dilema dos prisioneiros, produzindo uma briga.

Quando $V < C$, de modo que a luta é cara, este jogo é semelhante a outro jogo conhecido, chamado "Chicken", onde dois jogadores que se dirigem para um penhasco têm que decidir se param ou continuam. Aquele que pára primeiro perde a face, mas pode salvar sua vida. Mais geralmente, uma classe de jogos chamada "guerras de atrito" é usada para modelar esse tipo de situação. Nesse caso, um jogador gostaria de jogar Falcão se seu oponente jogar Pombo e jogar Pombo se seu oponente jogar Falcão.

3. Jogo de investimento

		Jogador 2	
		Investir	Não investir
Jogador 1	Investir	θ, θ	$\theta - c, 0$
	Não investir	$0, \theta - c$	$0, 0$

Aqui, duas partes decidem simultaneamente se investem; o investimento é mais valioso se a outra parte também investe (como no jogo de coordenação). Por exemplo, considere um potencial trabalhador e um empregador em potencial. O trabalhador em potencial decide se obtém educação (investindo em seu capital humano) e o potencial empregador decide investir em uma tecnologia que exija capital humano. Pense sobre quais são os resultados razoáveis para vários valores de θ e c. Como você analisaria essas situações se os jogadores não conhecerem os valores reais desses parâmetros, mas tiverem alguma informação privada sobre o que esses valores poderiam ser?

6.2 Representação de Jogos

Agora estamos prontos para apresentar formalmente jogos e alguns conceitos fundamentais, como uma estratégia. Para analisar situações estratégicas, é necessário conhecer

- quem são os jogadores
- quais ações estão disponíveis para eles
- quanto cada jogador valoriza cada resultado
- o que cada jogador conhece

Um jogo é apenas uma representação formal da informação acima. Isso geralmente é feito de uma das duas maneiras a seguir:

1. A representação na forma extensiva, em que as informações acima são explicitamente descritas usando árvores de jogos e conjuntos de informações.
2. A representação na forma normal (ou forma estratégica), na qual as informações acima são resumidas por meio de estratégias.

Ambas as formas de representação são úteis a sua própria maneira, e vou usar ambas as representações extensivamente ao longo do curso. É importante enfatizar que, ao descrever o que um jogador sabe, é necessário especificar não apenas o que ele sabe sobre parâmetros externos, como os payoffs, mas também o que ele sabe sobre os conhecimentos e crenças dos outros jogadores sobre esses parâmetros, o que ele sabe sobre o conhecimento dos outros jogadores sobre suas próprias crenças e assim por diante. Em ambas as representações, essas informações são codificadas de maneira econômica. Na primeira metade deste curso, nos concentraremos em questões não-informacionais, limitando-nos aos jogos de informação completa, nos quais tudo o que é conhecido por um jogador é conhecido por todos. Na segunda parte, nos concentraremos em questões informacionais, permitindo que os jogadores tenham informações assimétricas, para que se possa conhecer uma informação que não é conhecida por outra.

6.2.1 Representação na Forma Extensiva

A representação em forma extensiva de um jogo contém todas as informações sobre o jogo explicitamente, definindo quem se move quando, o que cada jogador sabe quando se movimenta, quais movimentos estão disponíveis para ele e para onde cada movimento leva, etc. Isso é feito pelo uso de uma árvore de jogos e conjuntos de informações, bem como informações mais básicas, como jogadores e os payoffs.

6.2.1.1 Árvore de um Jogo

Definição 6.2.1. *Uma árvore é um conjunto de nós e arestas direcionadas que conectam esses nós, tais que:*

1. *existe um nó inicial, para o qual não há borda de entrada;*
2. *para todos os outros nós, existe exatamente uma borda de entrada;*
3. *para quaisquer dois nós, existe um caminho único que conecta esses dois nós.*

Para uma ajuda visual, imagine os ramos de uma árvore que surgem do tronco. Por exemplo, o gráfico na Figura 6.2 é uma árvore. Existe um nó inicial único e ramifica-se a partir daí sem formar um loop. Parece uma árvore. Por outro lado, o gráfico da Figura 6.3 não é uma árvore. No gráfico há dois caminhos alternativos para o mesmo nó.

Figura 6.2 – Uma árvore

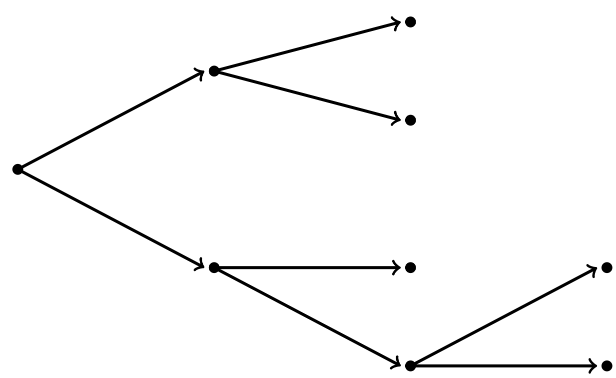
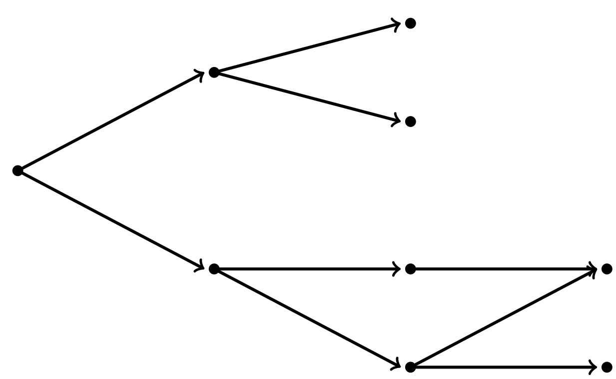


Figura 6.3 – Um gráfico que não é uma árvore

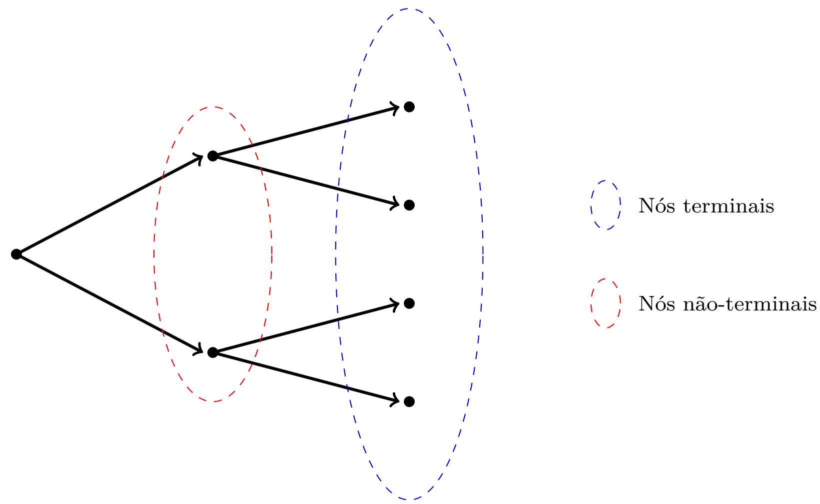


Observe que as arestas (ou setas) podem vir com rótulos, que podem ser iguais para duas setas diferentes. Em uma árvore de jogos, existem dois tipos de nós, nós terminais, nos quais o jogo termina, e nós não terminais, nos quais um jogador precisaria tomar uma decisão adicional. Isto é formalmente declarado como segue.

Definição 6.2.2. *Os nós que não são seguidos por outro nó são chamados de terminal. Os outros nós são chamados não-terminais.*

Por exemplo, os nós terminais e não-terminais da árvore de um jogo é representado na Figura 6.4. Não há seta de saída em nenhum nó terminal, indicação de que o jogo terminou. Um nó terminal também pode ser referido como um resultado no jogo. Nesse nó, precisamos especificar os pagamentos dos jogadores para descrever suas preferências entre os resultados. Por outro lado, existem algumas setas de saída em qualquer nó não terminal, indicando que algumas outras decisões devem ser tomadas. Nesse caso, precisa descrever quem toma uma decisão e o que sabe no momento da decisão.

Figura 6.4 – Nós terminais e nós não-terminais



6.2.1.2 Jogos na Forma Extensiva

Definição 6.2.3. *Um jogo consiste de*

- *um conjunto de jogadores*
- *uma árvore*
- *uma alocação de nós não-terminais da árvore para os jogadores*
- *uma partição informativa dos nós não-terminais*
- *payoffs para cada jogador em cada nó terminal.*

Definição 6.2.4 (Jogadores)**.** *O conjunto de jogadores consiste em tomadores de decisão ou atores que tomam alguma decisão durante o decorrer do jogo. Alguns jogos também podem conter um jogador especial, natureza, que representa a incerteza que os jogadores enfrentam. O conjunto de jogos é frequentemente denotado por*

$$N = \{1, 2, \ldots, n\}, \tag{6.1}$$

em que $i, j \in N$ *são designados como jogadores.*

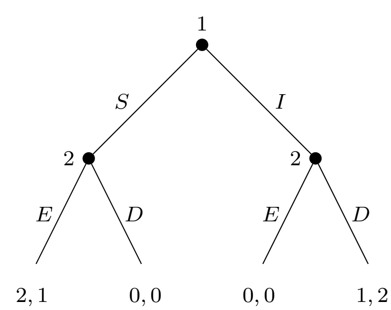
Definição 6.2.5 (Resultados e payoff)**.** *O conjunto de nós terminais geralmente denotado por* Z*. Em um nó terminal, o jogo terminou, levando a algum resultado. Nesse ponto, especifica-se um* *payoff, que é um número real, para cada jogador i. O mapeamento*

$$u_i \colon Z \to \mathbb{R}, \tag{6.2}$$

que mapeia cada nó terminal para o retorno do jogador i naquele nó é a função de utilidade Von-Neumann e Morgenstern do jogador i. Lembre-se que isso significa que o jogador i tenta maximizar o valor esperado de u. Ou seja, considerando quaisquer duas loterias p e q em Z ele prefere p a q se e somente se p levar a um valor esperado maior para a função u_i do que q, ou seja, $\sum_{z\in Z} u_i(z)p(z) \geq \sum_{z\in Z} u_i(z)q(z)$. Lembre-se também que essas preferências não mudam se multiplicarmos todos os payoffs por um número positivo fixo ou se adicionarmos um número fixo a todos os payoffs. As preferências mudam sob qualquer outra transformação.

Definição 6.2.6 (Nós de decisão)**.** *Em um nó não terminal, uma nova decisão deve ser tomada. Assim, na definição de um jogo, um jogador é atribuído a cada nó não terminal. Este é o jogador que tomará a decisão nesse momento. Para descrever o problema de decisão do jogador no momento, define-se as escolhas disponíveis para o jogador no momento. Estas são as setas de saída no nó, cada uma delas levando a um nó diferente. Cada uma dessas escolhas também é chamada de movimento ou ação (de forma intercambiável). Observe que os movimentos vêm com seus rótulos e duas setas diferentes podem ter o mesmo rótulo. Nesse caso, eles são o mesmo movimento.*

Figura 6.5 – Jogo das moedas com informação perfeita



Exemplo 6.2.1. *Considere o jogo na Figura 6.5. A árvore consiste de 7 nós. O primeiro é alocado para o jogador 1 e os dois seguintes para o jogador 2. Os quatro nós finais têm payoffs anexados a eles. Como existem dois jogadores, os vetores de payoffs têm dois elementos. O primeiro número é o payoff do jogador 1 e o segundo é o payoff do jogador 2. Estes payoffs são os rótulos da utilidade de von Neumann Morgenstern. Isto é, cada jogador tenta maximizar o valor esperado de seus próprios ganhos, considerando suas crenças sobre como os outros jogadores irão jogar o jogo.*

6.2.1.3 Conjunto Informacional

Definição 6.2.7. *Um conjunto de informações é uma coleção de nós, tal que*

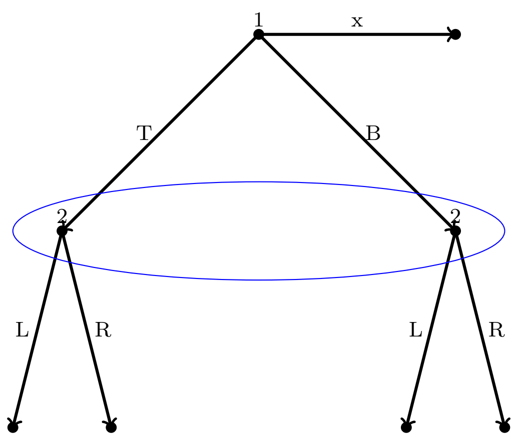
1. *o mesmo jogador i deve se mover em cada um desses nós;*

2. *os mesmos movimentos estão disponíveis em cada um desses nós.*

Definição 6.2.8. *Uma partição de informação é uma alocação de cada nó não terminal da árvore para um conjunto de informações; o nó inicial deve estar "sozinho".*

O significado de um conjunto de informações é que, quando o indivíduo está nesse conjunto de informações, ele sabe que um dos nós no conjunto de informações foi atingido, mas não pode descartar nenhum dos nós no conjunto de informações. Além disso, em um jogo, o conjunto de informações pertence ao jogador que deve se mover no conjunto de informações fornecido, representando sua incerteza. Ou seja, o jogador que deve mover-se no conjunto de informações não consegue distinguir entre os pontos no conjunto de informações, mas é capaz de distinguir entre os pontos fora do conjunto de informações daqueles contidos nele. Portanto, a definição acima seria sem sentido sem a condição 1, enquanto a condição 2 requer que o jogador conheça suas escolhas disponíveis. Esta última condição pode ser tomada como uma suposição simplificadora. Eu também me refiro a conjuntos de informações como histórico e escrevo h_i para um histórico genérico no qual o jogador i se move.

Figura 6.6 – Representação de um jogo



Por exemplo, considere o jogo na Figura 6.6. Aqui, o jogador 2 sabe que o jogador 1 tomou a ação T ou B e não a ação x; mas o jogador 2 não pode saber ao certo se o jogador 1 tomou a ação T ou B[1].

Diz-se que um jogo tem informação perfeita se cada conjunto de informações tiver apenas um elemento. Lembre-se de que, em uma árvore, cada nó é alcançado por meio de um caminho exclusivo. Assim, em um jogo de informação perfeita, um jogador pode construir o jogo anterior perfeitamente. Por exemplo, na Figura 6.5, o jogador 2 sabe se o jogador 1 escolheu car ou coroa. E o jogador 1 sabe que quando ele joga cara ou coroa, o jogador 2 saberá o que o jogador 1 jogou.

6.2.1.4 Natureza como um Jogador e a Representação da Incerteza

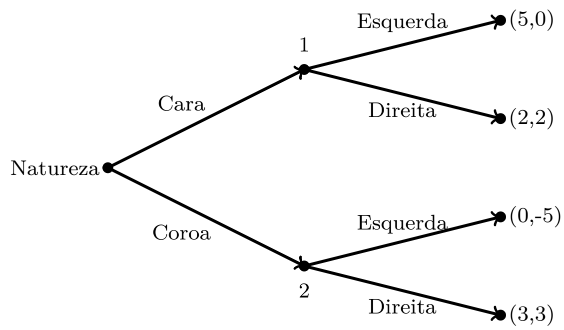
O conjunto de jogadores inclui os decisores que participam no jogo. No entanto, em muitos jogos existe espaço para o acaso. Mais amplamente, os jogadores frequentemente enfrentam incer-

[1] Ao longo dessas notas, os conjuntos de informações são representados por círculos (como em conjuntos) ou por curvas tracejadas conectando os nós nos conjuntos de informações, dependendo da conveniência.

tezas sobre alguns fatos relevantes, incluindo o que os outros jogadores sabem. Nesse caso, mais uma vez o acaso desempenha um papel (como representação). Para representar essas possibilidades, introduzimos um jogador fictício: a natureza. Não há recompensa para a natureza nos nós finais e toda vez que um nó é alocado para a natureza, uma distribuição de probabilidade sobre os ramos que seguem precisa ser especificada, por exemplo, cara com probabilidade de $\frac{1}{2}$ e coroa com probabilidade de $\frac{1}{2}$. Note que isto é o mesmo que adicionar loterias ao jogo.

Por exemplo, considere o jogo na Figura 6.7. Neste jogo, uma moeda justa é lançada, onde a probabilidade de cara é $\frac{1}{2}$. Se der cara, o jogador 1 escolhe entre esquerda e direita; se der coroa, o jogador 2 escolhe entre esquerda e direita. Os payoffs também dependem do sorteio.

Figura 6.7 – Representação de um jogo com sorteio



6.2.1.5 Conhecimento Comum

A estrutura de um jogo é assumida como sendo conhecida por todos os jogadores, e é assumido que todos os jogadores conhecem a estrutura e assim por diante. Ou seja, em uma linguagem mais formal, a estrutura do jogo é de conhecimento comum. Por exemplo, no jogo da Figura 6.6, o jogador 1 sabe que, se ele escolher T ou B, o jogador 2 saberá que o jogador 1 escolheu uma das ações acima, sem poder descartar nenhuma delas. Além disso, o jogador 2 sabe que o jogador 1 possui o conhecimento acima, e o jogador 1 sabe que o jogador 2 sabe disso, e assim por diante. Usando conjuntos de informações e árvores de jogo mais ricas, pode-se modelar estruturas de informação arbitrárias como essa. Por exemplo, pode-se também modelar uma situação na qual o jogador 1 não sabe se o jogador 2 poderia distinguir as ações T e B. No sentido de modelar a incerteza do jogador 1, introduzir-se-ia ainda um movimento de aleatoriedade, cujo resultado conduza aos dois primeiros conjuntos de informação (caso observável) ou ao último caso de informação (caso não observável).

6.2.2 Estratégias

Definição 6.2.9. *Uma estratégia de um jogador é um plano contingente completo que determina qual ação ele tomará em cada conjunto de informações que ele deve se mover (incluindo os conjuntos de informações que não serão atingidos de acordo com essa estratégia). Matematicamente, uma estratégia de um jogador i é uma função s_i que mapeia todo conjunto de informações h_i do jogador* *i para uma ação que está disponível em h_i.*

É importante observar as três sutilezas a seguir na definição.

1. Deve-se atribuir um movimento a todos os conjuntos de informações do jogador. Se omitirmos atribuir um movimento a um conjunto de informações, não saberíamos o que o jogador teria feito quando esse conjunto de informações fosse atingido.

2. A jogada atribuída deve estar disponível no conjunto de informações. Se a movimentação atribuída não estiver disponível em um conjunto de informações, o plano não será viável, pois não poderá ser executado quando esse conjunto de informações for atingido.

3. Em todos os nós em um determinado conjunto de informações, o jogador executa o mesmo movimento. Afinal, o jogador não consegue distinguir esses nós uns dos outros.

Exemplo 6.2.2. *Na Figura 6.5, o jogador* 1 *possui apenas um conjunto de informações. Assim, o conjunto de estratégias para o jogador* 1 *é* $\{$*cara, coroa*$\}$*. Por outro lado, o jogador* 2 *possui dois conjuntos de informações. Por isso, uma estratégia do jogador* 2 *determina o que fazer em cada conjunto de informação, isto é, dependendo do que o jogador* 1 *faz.*

6.3 Estratégias

6.3.1 Dominância

As seções anteriores se concentraram em como descrever formalmente uma situação estratégica. Agora, começamos a analisar situações estratégicas para descobrir quais resultados são mais razoáveis e prováveis de se realizar. Para fazer isso, consideramos certos conjuntos de suposições sobre as crenças dos jogadores e descobrimos suas implicações no que eles jogariam. Tais análises levarão a conceitos de solução, que geram um conjunto de perfis estratégicos. Estes são os perfis estratégicos considerados possíveis pelo conceito de solução. Esta seção é dedicada a dois conceitos de solução: equilíbrio de estratégia dominante e racionalização. Esses conceitos de solução são baseados na ideia de que um jogador racional não desempenha uma estratégia que é dominada por outra estratégia.

Diz-se que um jogador é racional se e somente se ele maximiza o valor esperado de seus resultados (dadas suas crenças sobre as estratégias dos outros jogadores). Por exemplo, considere o seguinte jogo.

		Jogador 2	
		L	R
	T	$2, 0$	$-1, 1$
Jogador 1	M	$0, 10$	$0, 0$
	B	$-1, -6$	$2, 0$

Considere o jogador 1. Ele está pensando se deve jogar T, M ou B. Uma rápida inspeção de seus resultados revela que seu melhor jogo depende do que ele acha que o outro jogador faz. Vamos então escrever p para a probabilidade que ele atribui a L (como o jogador 2 joga). Então, seus retornos esperados de jogar T, M e B são

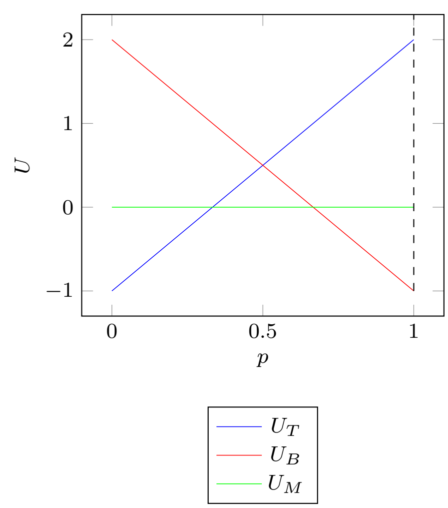
$$U_T = 2p - (1 - p) = 3p - 1, \tag{6.3}$$

$$U_M = 0, \tag{6.4}$$

$$U_B = -p + 2(1 - p) = 2 - 3p, \tag{6.5}$$

respectivamente. Esses valores em função de p são plotados na Figura 6.8. Como fica claro no gráfico, U_T é maior quando $p > \dfrac{1}{2}$ e U_B é maior quando $p < \dfrac{1}{2}$ Em $p = \dfrac{1}{2}$, $U_T = U_B > 0$. Portanto, se o jogador 1 for racional, ele jogará B quando $p < \dfrac{1}{2}$, T quando $p > \dfrac{1}{2}$e B ou T quando $p = \dfrac{1}{2}$. Observe que, se o jogador 1 for racional, ele nunca jogará M – não importa o que ele acredite sobre a estratégia do jogador 2. Portanto, se assumirmos que o jogador 1 é racional (e que o jogo é como descrito acima), então podemos concluir que o jogador 1 não joga M. Isso porque M é uma estratégia estritamente dominada, um conceito que definimos agora.

Figura 6.8 – Payoff esperado como função da probabilidade de L



Para descrever essa ideia de maneira mais geral e formal, vamos usar a notação s_{-i} para denotar a lista de estratégias s_j desempenhadas por todos os jogadores j que não o i, isto é,

$$s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n). \tag{6.6}$$

Definição 6.3.1. *Uma estratégia s_i^* domina estritamente s_i se e somente se*

$$u_i(s_i^*, s_{-i}) > u_i(s_i, s_{-i}), \quad \forall s_{-i} \in S_{-i}. \tag{6.7}$$

Ou seja, não importa o que os outros jogadores joguem, jogar s_i^* é estritamente melhor do que jogar s_i para o jogador i. Nesse caso, se i é racional, ele nunca jogaria a estratégia estritamente dominada s_i. Isto é, não há crença sob a qual ele jogaria s_i, pois s_i^* sempre renderia uma recompensa maior do que s_i não importando o que o jogador i acredite sobre os outros jogadores.

Uma estratégia mista σ_i domina uma estratégia s_i de maneira semelhante: σ_i domina estritamente s_i se e somente se

$$\sigma_i(s_{i1})u_i(s_{i1}, s_{-i}) + \sigma(s_{i2})u_i(s_{i2}, s_{-i}) + \ldots + (s_{ik})u_i(s_{ik}, s_{-i}) > u_i(s_i, s_{-i}), \quad \forall s_{-i} \in S_{-i}. \tag{6.8}$$

Observe que nenhuma das estratégias puras T, M e B domina qualquer estratégia. No entanto, M é dominado pela estratégia mista que σ_1 que atribui probabilidade de $\frac{1}{2}$ em T e B. Para cada p, a recompensa de jogar σ_1 é

$$U_{\sigma_1} = \frac{1}{2}(3p - 1) + \frac{1}{2}(2 - 3p) = \frac{1}{2}, \tag{6.9}$$

que é maior que 0, a recompensa de M.

Este é realmente um resultado geral. Para indicar o resultado, introduzo alguns conceitos básicos. Escreva

$$S_{-i} = \prod_{j \neq i} S_j, \tag{6.10}$$

para o conjunto de estratégias de outros jogadores e defina uma crença do jogador i como uma distribuição de probabilidadeβ_{-i} sobre S_{-i}.

Definição 6.3.2. *Para qualquer jogador i, uma estratégia s_i é a melhor resposta a s_{-i} se e somente se*

$$u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}), \quad \forall s_i' \in S_i. \tag{6.11}$$

Uma estratégia s_i é dita ser a melhor resposta a crença β_{-i} se e somente se s_i gera o maior payoff sob a crença β_{-i}, isto é,

$$\sum_{s_{-i} \in S_{-i}} u_i(s_i, s_{-i})\beta_{-i}(s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} u_i(s_i', s_{-i})\beta_{-i}(s_{-i}), \quad \forall s_i' \in S_i. \tag{6.12}$$

O conceito de melhor resposta é um dos principais conceitos da teoria dos jogos. É importante entender bem a definição e ser capaz de calcular a melhor resposta em jogos relativamente simples. Um jogador racional pode jogar uma estratégia sob uma crença apenas se for uma melhor resposta a essa crença.

Teorema 6.3.1. *Uma estratégia s_i é a melhor resposta para alguma crença se e somente se s_i não for dominada. Portanto, jogar a estratégia s_i nunca é racional se s_i for dominada por uma estratégia (mista ou pura).*

Se assumirmos que os jogadores são racionais (e que o jogo é como descrito), então podemos concluir que nenhum jogador joga uma estratégia estritamente dominada (por alguma estratégia mista ou pura), e isso é tudo o que se pode concluir.

6.3.2 Equilíbrio em Estratégia Dominante

Esta seção apresenta dois conceitos de dominância, um é mais forte que o outro. Usaremos oo conceito de dominância fraca para definir o equilíbrio em estratégia dominante.

Definição 6.3.3. *Uma estratégia s_i^* é uma estratégia estritamente dominante para o jogador i se e somente se s_i^* domina estritamente todas as outras estratégias do jogador i.*

Por exemplo, no jogo do dilema dos prisioneiros, não cooperar domina estritamente a única outra estratégia que é cooperar. Por isso, não cooperar é uma estratégia estritamente dominante. Se i é racional e tem uma estratégia estritamente dominante s_i^*, então ele não jogará nenhuma outra estratégia. Nesse caso, é razoável esperar que ele jogue s_i^*.

O problema é que existem poucas situações estratégicas interessantes em que os jogadores têm estratégias estritamente dominantes. Tais situações podem ser analisadas como problemas individuais de decisão. Uma forma ligeiramente mais fraca de dominância é mais comum, especialmente em jogos dinâmicos (que analisaremos no futuro) e em situações que surgem em ambientes estruturados, como sob mecanismos de negociação adequadamente projetados como em leilões. Esta forma mais fraca é chamada de dominância fraca:

Definição 6.3.4. *Uma estratégia s_i^* domina fracamente s_i se e somente se*

$$u_i(s_i^*, s_{-i}) \geq u_i(s_i, s_{-i}), \quad \forall s_{-i} \in S_{-i}, \tag{6.13}$$

e

$$u_i(s_i^*, s_{-i}) > u_i(s_i, s_{-i}). \tag{6.14}$$

para algum $s_{-i} \in S_{-i}$.

Ou seja, não importa o que os outros jogadores joguem, jogar s_i^* é pelo menos tão bom quanto jogar s_i, e existem algumas contingências nas quais jogar s_i^* é estritamente melhor que s_i. Nesse caso, se for racional, i "jogaria" somente se ele acreditar que essas contingências nunca ocorrerão. Se ele é cauteloso no sentido de atribuir uma probabilidade positiva para cada contingência, então ele não jogará s_i. Esta dominância fraca é usada na definição de uma estratégia dominante.

Definição 6.3.5. *Uma estratégia s_i^* de um jogador i é uma estratégia (fracamente) dominante se e somente se s_i^* domina fracamente todas as outras estratégias do jagador i.*

Quando há uma estratégia fracamente dominante, se o jogador é racional e cauteloso, então ele vai jogar a estratégia dominante.

Exemplo 6.3.1. *Neste jogo, o jogador* 1 *(firma) tem uma estratégia estritamente dominante: "contratar". O jogador* 2 *tem apenas uma estratégia fracamente dominada. Se os jogadores são racionais e, além disso, o jogador* 2 *é cauteloso, então o jogador* 1 *contrata e o jogador* 2 *faz um baixo esforço.*

		Jogador 2	
		Alto esforço	Baixo esforço
Jogador 1	Contrata	$2,2$	$1,3$
	Não contrata	$0,0$	$0,0$

6.3.3 Racionalização

Um jogador é considerado racional se maximizar o valor esperado de sua função de utilidade, conforme descrito no jogo. A seção anterior explorou as implicações da racionalidade. Isso foi capturado pela dominância. Em ambientes estratégicos naturais, isso geralmente gera previsões fracas. Além disso, os jogos em que a dominância por si só leva a uma previsão nítida (por exemplo, os jogos com um equilíbrio de estratégia dominante) não são interessantes para a teoria dos jogos porque em tal jogo a decisão de cada jogador pode ser analisada separadamente sem exigir uma análise teórica do jogo.

No entanto, na definição de um jogo, assume-se muito mais que a racionalidade dos jogadores. Supõe-se ainda que é de conhecimento comum que os jogadores são racionais. Ou seja, todo mundo é racional; todo mundo sabe que todo mundo é racional; todo mundo sabe que todo mundo sabe que todo mundo é racional; ad infinitum. Se algumas dessas suposições falharem, então seria necessário considerar um jogo diferente, o jogo que reflete o fracasso dessas suposições. Esta seção explora as implicações do conhecimento comum da racionalidade. Essas implicações são precisamente capturadas por um conceito de solução chamado racionalização, que é equivalente à eliminação iterativa de estratégias estritamente dominadas. Dessa forma, a racionalidade captura precisamente as implicações dos pressupostos embutidos na definição do jogo.

Vimos anteriormente que a estratégia M é estritamente dominada (por uma mistura de T e B) e, portanto, não pode ser a melhor resposta para qualquer crença. Portanto, a racionalidade do jogador 1 implica que o jogador 1 não joga M. Nenhuma outra estratégia é estritamente dominada. Por exemplo, para o jogador 2, ambas as estratégias podem ser a melhor resposta. Se ele acha que o jogador 1 provavelmente não jogará M, então ele deve jogar R, e se ele acha que é muito provável que o jogador 1 jogue M, então ele deve jogar L. Portanto, a racionalidade do jogador 2 não coloca nenhuma restrição em seu comportamento. Mas, e se ele achar que é muito provável que o jogador 1 seja racional? Nesse caso, uma vez que um jogador racional 1 não joga M, ele deve atribuir probabilidade muito pequena para o jogador 1 jogar M. De fato, se ele sabe que o jogador 1 é racional, então ele deve ter certeza de que ele não irá jogar M. Nesse caso, sendo racional, o jogador 2 deve jogar R. Em resumo, se o jogador 2 é racional e ele sabe que o jogador 1 é racional, então ele deve jogar R.

Observe que primeiro eliminamos todas as estratégias que são estritamente dominadas (ou seja, M) e, em seguida, tomando o jogo resultante, eliminamos novamente todas as estratégias que são estritamente dominadas (ou seja, L). Isso é chamado de eliminação iterada de estratégias estritamente dominadas. Em geral, se um jogador é racional e sabe que os outros jogadores também são racionais (e os ganhos são como são dados), então ele deve jogar uma estratégia que sobrevive duas vezes ao iterar a eliminação de estratégias estritamente dominadas.

Sob outras hipóteses de racionalidade, pode-se eliminar de maneira iterativa estratégias estritamente dominadas (se houver alguma). No exemplo anterior, lembre-se que a racionalidade do jogador 1 exige que ele jogue T ou B, e o conhecimento do fato de que o jogador 2 também é racional não coloca qualquer restrição em seu comportamento - já que a racionalidade em si não restringe o comportamento do jogador 2. Agora, suponha que o jogador 1 também saiba que o jogador 2 é racional e que o jogador 2 sabe que o jogador 1 é racional. Então, como mostra a análise acima, o jogador 1 deve saber que o jogador 2 irá jogar R. Nesse caso, sendo racional, ele deve jogar B.

Essa análise produz um procedimento mecânico para analisar jogos a partir de eliminação iterada de estratégias estritamente dominadas k vezes: eliminar todas as estratégias estritamente dominadas e iterar essas k-vezes. Neste procedimento, elimina-se todas as estratégias estritamente dominadas e itera-se k-vezes.

Dois pontos são cruciais para o procedimento de eliminação:

1. É preciso eliminar apenas as estratégias estritamente dominadas. Não se pode eliminar uma estratégia se ela é fracamente dominada, mas não estritamente dominada.

2. É preciso eliminar as estratégias estritamente dominadas por estratégias mistas (mas não necessariamente por estratégias puras).

Quando existem apenas algumas estratégias finitas, este processo de eliminação deve parar em algum k. Ou seja, em algum momento, não haverá estratégia dominada para eliminar. Nesse caso, iterar a eliminação não teria qualquer efeito.

O processo de eliminação que continua a eliminar iterativamente todas as estratégias estritamente dominadas até que não exista uma estratégia estritamente dominada é chamado de Eliminação Iterada de Estratégias Estritamente Dominadas. Diz-se que uma estratégia é racionalizável se e somente se sobreviver à eliminação iterada de estratégias estritamente dominadas.

Teorema 6.3.2. *Se é de conhecimento comum que todo jogador é racional (e o jogo é como descrito), então todo jogador deve jogar uma estratégia racionalizável. Além disso, qualquer estratégia racionalizável é consistente com o conhecimento comum da racionalidade.*

Um problema geral com a racionalização é que geralmente existem muitas estratégias racionalizáveis; o processo de eliminação geralmente pára cedo demais. Nesse caso, não se pode fazer muita previsão com base em tal análise. Por exemplo, no jogo abaixo

		Jogador 2	
		Cara	Coroa
Jogador 1	Cara	$-1, 1$	$1, -1$
	Coroa	$1, -1$	$-1, 1$

toda estratégia é racionalizável e não podemos dizer o que os jogadores farão.

6.4 Equilíbrio de Nash

6.4.1 Equilíbrio de Nash

Tanto o equilíbrio de estratégia dominante quanto a racionalização são conceitos de solução bem fundamentados. Se os jogadores são racionais e são cautelosos no sentido de atribuir uma probabilidade positiva a cada uma das estratégias dos outros jogadores, então esperamos que os jogadores joguem de acordo com o equilíbrio da estratégia dominante sempre que tal equilíbrio existir. Por outro lado, a racionalidade descreve exatamente o que está implícito na definição do jogo (também conhecido como conhecimento comum da racionalidade). Se é de conhecimento comum que os jogadores são racionais (ou seja, eles maximizam o valor esperado de sua função de utilidade), então cada jogador deve estar jogando uma estratégia racionalizável. Além disso, toda estratégia racionalizável pode ser racionalizável no sentido de que um jogador pode jogar essa estratégia e ainda acreditar que é de conhecimento comum que os jogadores são racionais.

Infelizmente, esses conceitos de solução não são úteis na maioria das situações em economia. Com exceção dos jogos projetados especificamente, como no leilão de segundo preço, muitas vezes não há equilíbrio de estratégia dominante. O conjunto de estratégias racionalizáveis tende a ser grande em jogos analisados em economia. Nesse caso, pode-se fazer apenas previsões fracas sobre o resultado usando racionalização.

Esta seção apresenta um novo conceito de solução: o Equilíbrio de Nash. Ele pressupõe que os jogadores adivinham corretamente as estratégias dos outros jogadores. Esta suposição pode ser razoável quando há uma interação prévia longa que leva os jogadores a formar opinião sobre como os outros jogadores jogam. Também pode ser razoável quando existe uma convenção social, aderida pelos outros jogadores. Para definir o equilíbrio de Nash, considere o jogo Batalha dos Sexos.

		João	
		Ópera	Futebol
Maria	Ópera	$4, 1$	$0, 0$
	Futebol	$0, 0$	$1, 4$

Neste jogo, não há estratégia dominante e tudo é racionalizável. Suponha que Maria jogue ópera. Então, a melhor coisa que João pode fazer é jogar ópera também. Assim, a ópera é a melhor resposta para João se Maria jogar ópera. Assim, (ópera, ópera) é um Equilíbrio de Nash.

Vamos formalizar esta ideia para um jogo qualquer lembrando que, para um jogador i, uma estratégia s_i^{BR} é a melhor resposta a s_{-i} se e somente se

$$u_i(s_i^{BR}, s_{-i}) \geq u_i(s_i, s_{-i}), \quad \forall s_i \in S_i. \tag{6.15}$$

Lembre-se também que a definição de melhor resposta difere da de estratégia dominante, exigindo a desigualdade acima apenas para uma estratégia específica s_{-i}, em vez de exigi-la para todos os $s_{-i} \in S_i$. Se a desigualdade fosse verdadeira para todo s_{-i}, então s_i^{BR} também seria uma estratégia dominante, que é um requisito mais forte do que ser uma melhor resposta contra alguma estratégia s_{-i}.

Definição 6.4.1. *Um perfil de estratégia $s^* = (s_1^*, \ldots, s_n^*)$ é um Equilíbrio de Nash se e somente se s_i^* é uma melhor resposta a $s_i^* = (s_1^*, \ldots, s_{i-1}^*, s_{i+1}^*, \ldots, s_n^*)$ para cada i. Isto é, para todo i,*

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \quad \forall s_i \in S_i. \tag{6.16}$$

Em outras palavras, nenhum jogador teria um incentivo para desviar, se ele adivinhar corretamente as estratégias dos outros jogadores. Se alguém vê um perfil estratégico como uma convenção social, então, estar em equilíbrio de Nash está ligado a ser self-enforcing, isto é, ninguém quer se desviar quando pensa que os outros seguirão a convenção.

Por exemplo, na batalha do jogo dos sexos, (ópera, ópera) é um equilíbrio de Nash porque

$$u_{Maria}(\text{ópera, ópera}) = 4 > 0 = u_{Maria}(\text{futebol, ópera}),$$

e

$$u_{Joao}(\text{ópera, ópera}) = 1 > 0 = u_{Joao}(\text{ópera, futebol}).$$

Da mesma forma, (futebol, futebol) também é um equilíbrio de Nash.

Todo equilíbrio de estratégia dominante é também um equilíbrio de Nash, mas o inverso não é verdadeiro.

Teorema 6.4.1. *Se s^* é um equilíbrio de estratégia dominante, então s^* é um equilíbrio de Nash.*

Para ver que o inverso não é verdadeiro, considere a Batalha dos Sexos. Neste jogo, ambos (ópera, ópera) e (futebol, futebol) são os equilíbrios de Nash, mas nenhum deles é um equilíbrio de estratégia dominante. Além disso, pode haver no máximo um equilíbrio de estratégia dominante, mas como mostra a Batalha dos Sexos, o equilíbrio de Nash não é único em geral.

Também pode haver outros equilíbrios de Nash quando há um equilíbrio de estratégia dominante. Por exemplo, considere o jogo

		1	
		a	b
2	a	1, 1	0, 0
	b	0, 0	0, 0

Neste jogo, (a, a) é um equilíbrio de estratégia dominante, mas (b, b) também é um equilíbrio de Nash.

Este exemplo também ilustra que um equilíbrio de Nash pode estar em estratégias fracamente dominadas. Nesse caso, pode-se descartar alguns equilíbrios de Nash, eliminando estratégias fracamente dominadas. Embora possa achar esse equilíbrio irracional e estar disposto a descartar tais equilíbrios, todos os equilíbrios de Nash podem precisar estar em estratégias dominadas em alguns jogos.

Se uma estratégia é jogada em um equilíbrio de Nash, então é racionalizável, mas pode haver estratégias racionalizáveis que não são jogadas em qualquer equilíbrio de Nash.

Teorema 6.4.2. *Se s^* é um equilíbrio de Nash, então s_i^* é racionalizável para todo jogador i.*

O inverso não é verdadeiro. Isto é, pode haver uma estratégia racionalizável que não seja jogada em qualquer equilíbrio de Nash, como o próximo exemplo ilustra.

		Jogador 2		
		a	b	c
	a	$(1, -2)$	$(2, -1)$	$(0, 0)$
Jogador 1	b	$(-1, 2)$	$(1, -2)$	$(0, 0)$
	c	$(0, 0)$	$(0, 0)$	$(0, 0)$

Note que (c, c) é o único equilíbrio de Nash. Em contraste, nenhuma estratégia é estritamente dominada e, portanto, todas as estratégias são racionalizáveis.

6.4.2 Equilíbrio de Nash em Estratégia Mista

A definição acima abrange apenas as estratégias puras. Podemos definir o equilíbrio de Nash para estratégias mistas, alterando as estratégias puras com as estratégias mistas. Novamente, dada a estratégia mista dos outros, cada agente maximiza sua recompensa esperada sobre suas próprias estratégias (mistas).

Definição 6.4.2. *Um perfil de estratégia mista $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*)$ é um equilíbrio de Nash se e somente se para todo jogador i, σ_i^* é a melhor resposta a σ_{-i}^*.*

Considere o jogo Batalha dos Sexos novamente.

		João	
		Ópera	Futebol
Maria	Ópera	$4,1$	$0,0$
	Futebol	$0,0$	$1,4$

Já identificamos dois equilíbrios em estratégia pura. Além disso, há um equilíbrio em estratégia mista. Para calcular o equilíbrio, seja p a probabilidade de Maria ir à ópera; com probabilidade $1-p$ ela vai ao jogo de futebol. Escreva também q para a probabilidade de João ir à ópera. Para Maria, o retorno esperado da ópera é

$$U_M(\text{ópera}, q) = qu_M(\text{ópera, ópera}) + (1-q)u_M(\text{ópera, futebol}) = 4q. \tag{6.17}$$

e o retorno esperado do futebol é

$$U_M(\text{futebol}, q) = qu_M(\text{futebol, ópera}) + (1-q)u_M(\text{futebol, futebol}) = 1-q. \tag{6.18}$$

Sua recompensa esperada da estratégia mista é

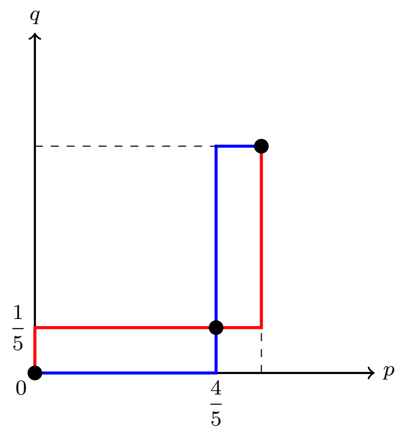
$$U_M(p;q) = pU_M(\text{ópera}, q) + (1-p)U_M(\text{futebol}, q) = 4pq + (1-p)(1-q). \tag{6.19}$$

A função payoff $U_M(p;q)$ é estritamente crescente em p quando $U_M(\text{ópera}, q) > U_M(\text{futebol}, q)$. Este é o caso quando $4q > 1-q$ ou equivalentemente quando $q > \dfrac{1}{5}$. Nesse caso, a única melhor resposta para Maria é $p = 1$, e ela vai à ópera com certeza. Da mesma forma, quando $q < \dfrac{1}{5}$, $U_M(\text{ópera}, q) > U_M(\text{futebol}, q)$ e seu resultado esperado $U_M(p;q)$ é decrescente em p. Nesse caso, a melhor resposta de Maria é $q = 0$, ou seja, ir ao jogo de futebol com certeza. Finalmente, quando $q = \dfrac{1}{5}$, o resultado esperado $U_M(p;q)$ não depende de p, e qualquer $p \in [0,1]$ é a melhor resposta. Em outras palavras, Maria escolheria ópera se sua utilidade esperada da ópera fosse maior, futebol se a utilidade esperada do futebol fosse maior, e escolheria ópera ou futebol ou qualquer randomização entre eles se ela fosse indiferente entre os dois.

Da mesma forma, pode-se calcular que $q = 1$ é a melhor resposta se $p > \dfrac{4}{5}$; $q = 0$ é a melhor resposta se $p < \dfrac{4}{5}$ e qualquer q pode ser melhor resposta se $p = \dfrac{4}{5}$.

As melhores respostas são apresentadas na Figura 6.9. O equilíbrio de Nash é onde essas melhores respostas se cruzam. Há um em $(0,0)$, quando ambos vão para o futebol, um em $(1,1)$, quando ambos vão para ópera, e há um em $\left(\dfrac{4}{5}, \dfrac{1}{5}\right)$, quando Maria vai para ópera com probabilidade de $\dfrac{4}{5}$, e João vai para ópera com probabilidade de $\dfrac{1}{5}$.

Figura 6.9 – Melhor resposta na batalha dos sexos



6.4.3 Equilíbrio de Nash em um Ambiente Dinâmico

Considere o seguinte jogo:

		Jogador 2	
		Falcão	Pombo
Jogador 1	Falcão	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
	Pombo	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

Suponha que $V < C$, de modo que os pagamentos sejam negativos quando dois falcões se encontrarem. Pode-se verificar facilmente que existem dois equilíbrios de Nash em estratégias puras: (falcão, pomba) e (pomba, falcão). Há também um equilíbrio de estratégia mista onde ambas as estratégias são jogadas com probabilidade positiva. Seja h a probabilidade de o jogador 2 jogar falcão e $d = 1 - h$ a probabilidade de ele jogar pomba. Como o jogador 1 joga ambas as estratégias com probabilidade positiva, ele deve ser indiferente entre elas:

$$\frac{V-C}{2} \cdot h + V \cdot d = \frac{V}{2} \cdot d, \tag{6.20}$$

onde o lado esquerdo é o resultado esperado de jogar falcão e o lado direito é o resultado esperado de jogar pomba. A solução para esta equação é

$$h = \frac{V}{C}. \tag{6.21}$$

Similarmente, para que o jogador 2 jogue tanto falcão quanto pomba com probabilidades positivas (que são jogadas com probabilidades positivas $\frac{V}{C}$ e $1 - \frac{V}{C}$, respectivamente), o jogador 1 deve jogue falcão com probabilidade $\frac{V}{C}$. Portanto, no equilíbrio de Nash de estratégia mista, cada jogador joga falcão com probabilidade $\frac{V}{C}$ e pomba com probabilidade $1 - \frac{V}{C}$.

Agora imagine uma ilha onde falcões e pombos vivem juntos. Seja H_0 falcões e D_0 pombas no começo, onde ambos H_0 e D_0 são muito grandes. Suponha que em cada estação, as aves sejam casadas aleatoriamente e o número de descendentes de uma ave seja dado pela matriz de payoffs acima. Ou seja, se uma pomba é correspondida a uma pomba como vizinha, então ela terá $\frac{V}{2}$ descendentes e, na próxima geração, teremos $1 + \frac{V}{2}$ pombos em sua família. Se uma pomba é combinada com um falcão, então ela terá descendentes zero e sua família terá apenas um membro, na próxima temporada. Se dois falcões são combinados, então cada um terá $\frac{V-C}{2}$ descendentes, o que é negativo, refletindo a situação de que o número de falcões de tais jogos diminuirá quando formos para a próxima temporada. Finalmente, se um falcão encontra pomba, ele terá descendentes, e haverá $1+V$ falcões em sua família na próxima temporada. Queremos saber a proporção de falcões e pombos nesta ilha milhões de temporadas mais tarde.

Seja H_t e D_to número de falcões e pombas, respectivamente, na estação t. Defina

$$h_t = \frac{H_t}{H_t + D_t} \quad \text{e} \quad \frac{D_t}{H_t + D_t}, \tag{6.22}$$

como as proporções de falcões e pombas em t. De acordo com a lei forte dos grandes números, suponha que o número de falcões que correspondem aos falcões seja $H_t h_t$, e o número de falcões que são correspondidos a pombos seja $H_t d_t$. Cada falcão no primeiro grupo multiplica para $1 + \frac{V-C}{2}$ e cada falcão no segundo grupo multiplica para $1 + \frac{V}{2}$. O número de falcões na próxima temporada será então

$$\begin{aligned} H_{t+1} &= \left(1 + \frac{V-C}{2}\right) H_t h_t + (1+V) H_t d_t \\ &= \left(1 + \frac{(V-C)h_t}{2} + V d_t\right) H_t. \end{aligned} \tag{6.23}$$

O número de pombas que são combinados a falcões é $D_t h_t$, e o número de pombas que são combinadas a pombos é $D_t d_t$. Cada pomba no primeiro e no segundo grupo se multiplica para 1 e $1 + \frac{V}{2}$, respectivamente. Assim, o número de pombos na próxima temporada será então

$$\begin{aligned} D_{t+1} &= (1+0) D_t h_t + \left(1 + \frac{V}{2}\right) D_t d_t \\ &= \left(1 + \frac{V d_t}{2}\right) D_t. \end{aligned} \tag{6.24}$$

É fácil encontrar os estados estacionários da razão h_t e d_t, definida por

$$h_{t+1} = h_t \quad \text{e} \quad d_{t+1} = d_t. \tag{6.25}$$

Encontramos que $h_t = 0$ e $d_t = 1$ é um estado estacionário, que pode ser alcançado se começarmos com todos os pombos. Nesse caso, teremos somente pombos ao final. Da mesma forma, outro estado estável é $h_t = 1$ e $d_t = 0$, que pode ser alcançado se começarmos com todos os falcões. Uma vez que começamos com falcões e pombos, D_t e D_{t+1} são positivos. Assim, podemos calcular os estados estacionários

$$\frac{H_t}{D_t} = \frac{H_{t+1}}{D_{t+1}} = \frac{H_t}{D_t} \frac{1 + \frac{(V-C)h_t}{2} + Vd_t}{1 + \frac{Vd_t}{2}}. \tag{6.26}$$

A igualdade é válida se e somente se

$$\frac{(V-C)h_t}{2} + Vd_t = \frac{Vd_t}{2} \quad \implies \quad h_t = \frac{V}{C}. \tag{6.27}$$

Este é o único estado estacionário alcançado a partir de uma distribuição com falcões e pombas. Observe que é a estratégia mista do equilíbrio de Nash do jogo subjacente. Este é um fato geral: se uma dinâmica populacional é como descrita nesta seção, então os estados estacionáveis alcançáveis a partir de uma distribuição completamente mista são os equilíbrios de Nash simétricos.

6.4.4 Aplicação: Competição Imperfeita

Algumas das primeiras aplicações da teoria dos jogos são as análises da concorrência imperfeita de Cournot (1838) e Bertrand (1883), um século antes de Nash (1950). Esta seção aplica os conceitos de solução por racionalidade e equilíbrio de Nash a esses modelos de concorrência imperfeita.

6.4.4.1 Competição em Cournot

Considere que haja n firmas. Cada firma i produz $q_i \geq 0$ unidades de um bem ao custo marginal $c \geq 0$ e vende ao preço

$$P = \max\{1 - Q, 0\}, \tag{6.28}$$

em que

$$Q = q_1 + \ldots + q_n, \tag{6.29}$$

é a oferta total. Cada firma maximiza o lucro esperado. Portanto, o payoff da firma i é

$$\pi_i = q_i(P - C). \tag{6.30}$$

Assumindo que as informações acima são de conhecimento comum, podemos escrever o jogo na forma normal, em que:

- $N = \{1, 2, \ldots, n\}$ é o conjunto de jogadores.
- $S_=[0, \infty)$ é o espaço de estratégias do jogador i, em que uma estratégia típica é a quantidade q_i a ser produzida pela firma i.
- $\pi_i = S_1 \times S_2 \times \ldots \times S_n \to \mathbb{R}$ é a função payoff.

É útil conhecer a melhor resposta de uma empresa aos níveis de produção das outras empresas. Seja

$$Q_{-i} = \sum_{j \neq i} q_j, \tag{6.31}$$

a oferta total das empresas que não a empresa i. Se $Q_{-i} > 1$, então o preço é $P = 0$ e a melhor resposta da empresa i é produzir zero e obter lucro zero. Agora assuma $Q_{-i} < 1$. Para qualquer $q_i \in (0, 1 - Q_{-i})$, a função lucro da firma i é

$$q_i^B(Q_{-i}) = \frac{1 - Q_{-i} - c}{2}. \tag{6.32}$$

A função lucro é plotada na Figura 6.10. A função melhor resposta é plotada na Figura 6.11.

Figura 6.10 – Função lucro

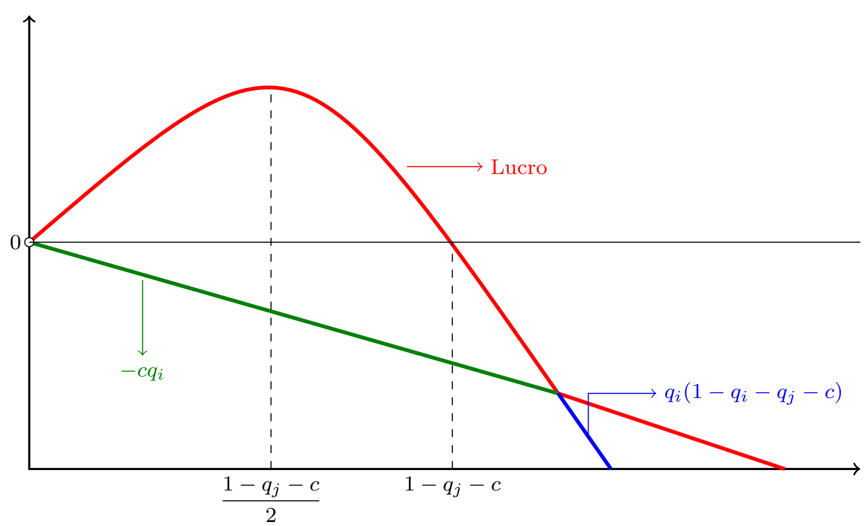
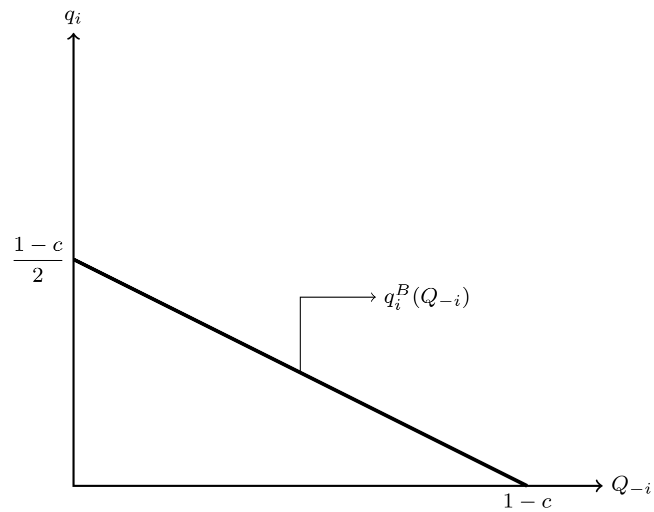


Figura 6.11 – Função melhor resposta



Agora considere o caso de duas firmas, duopólio de Cournot. Nesse caso, $Q_{-i} = q_j$ para $i \neq j$.

Um equilíbrio de Nash deve (q_1, q_2) deve satisfazer

$$q_1 = q_1^B(q_2) \equiv \frac{1 - q_2 - c}{2}, \tag{6.33}$$

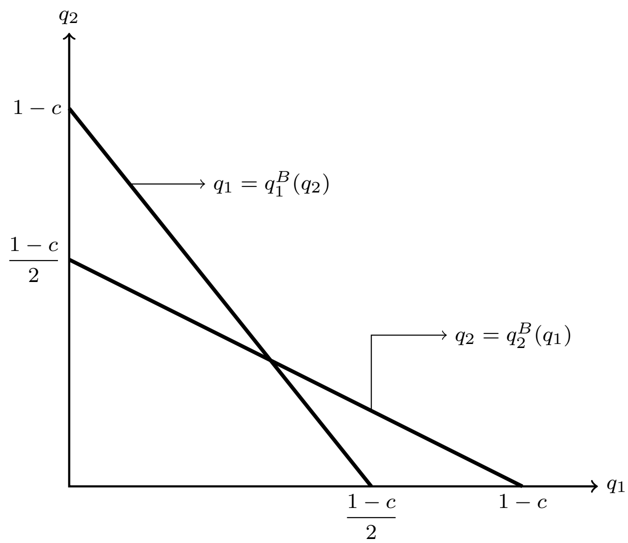
$$q_2 = q_2^B(q_1) \equiv \frac{1 - q_1 - c}{2}. \tag{6.34}$$

Resolvendo essas duas equações simultaneamente, podemos obter

$$q_1^* = q_2^* = \frac{1-c}{3}, \tag{6.35}$$

como o único equilíbrio de Nash. Graficamente, Figura 6.12, podemos plotar a função melhor resposta de ambas as firmas e a intersecção corresponde ao equilíbrio de Nash.

Figura 6.12 – Função melhor resposta em duopólio de Cournot



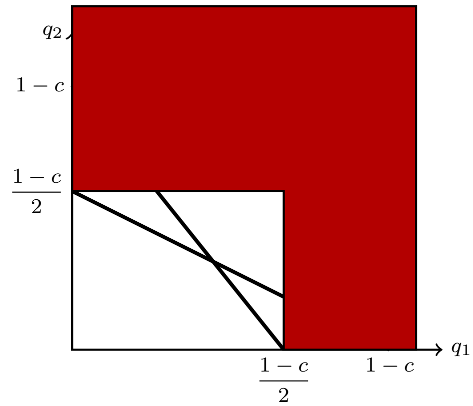
O jogo do duopólio de Cournot (linear) considerado aqui é de domínio solucionável. Isto é, há uma estratégia racionalizável única. Vamos primeiro considerar as primeiras rodadas de eliminação para ver isso intuitivamente. Então mostrarei matematicamente que esse é realmente o caso.

- Round 1

  Observe que uma estratégia $\widehat{q}_i > \dfrac{1-c}{2}$ é estritamente dominada por $\dfrac{1-c}{2}$. Para ver isso, considere qualquerq_j. Como na Figura 6.10, $\pi_i(q_i, q_j)$ é estritamente crescente até $q_i = \dfrac{1-c-q_j}{2}$ e decrescente a partir desse ponto. Em particular,

$$\pi_i\left(\frac{1-c-q_j}{2}, q_j\right) \geq \pi_i\left(\frac{1-c}{2}, q_j\right) > \pi_i(\widehat{q}_i, q_j). \tag{6.36}$$

  mostrando que $\widehat{q}_i$ é estritamente dominada por $\dfrac{1-c}{2}$. Portanto, eliminamos todos $\widehat{q}_i > \dfrac{1-c}{2}$ para todo jogador i. As estratégias resultantes são as seguintes, em que a área pintada foi eliminada.

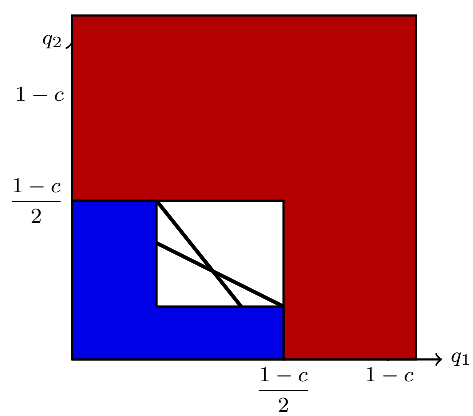


- Round 2

  No restante do jogo, $q_j \leq \frac{1-c}{2}$. Consequentemente, qualquer estratégia $\widehat{q}_i < \frac{1-c}{4}$ é estritamente dominada por $\frac{1-c}{4}$. Para ver isso, considere qualquer $q_j \leq \frac{1-c}{2}$. Como na Figura 6.10, $\pi_i(q_i, q_j)$ é estritamente crescente até $q_i = \frac{1-c-q_j}{2}$, o que é maior ou igual a $\frac{1-c}{4}$. Em particular,

$$\pi_i(\widehat{q}_i, q_j) < \pi_i\left(\frac{1-c}{4}, q_j\right) \leq \pi_i\left(\frac{1-c-q_j}{2}, q_j\right), \tag{6.37}$$

  mostrando que $\widehat{q}_i$ é estritamente dominada por $\frac{1-c}{4}$. Portanto, eliminamos todos $\widehat{q}_i < \frac{1-c}{4}$ para todo jogador i. As estratégias resultantes são as seguintes, em que a área pintada foi eliminada.



  Observe que o jogo restante é uma réplica menor do jogo original. Aplicando o mesmo procedimento repetidamente, pode-se eliminar todas as estratégias, exceto o equilíbrio de Nash. (Após cada duas rodadas, uma réplica menor é obtida.) Portanto, a única estratégia racionalizável é a estratégia única do equilíbrio de Nash.

Lema 6.4.1. *Dado que* $q_j \leq \overline{q}$*, toda estratégia* $\widehat{q}_i$ *com* $\widehat{q}_i < q_i^B(\overline{q})$ *é estritamente dominada por* $q_i^B(\overline{q}) = \dfrac{1-\overline{q}-c}{2}$*. Dado que* $q_j \geq \overline{q}$*, toda estratégia* $\widehat{q}_i$ *com* $\widehat{q}_i > q_i^B(\overline{q})$ *é estritamente dominada por* $q_i^B(\overline{q}) = \dfrac{1-\overline{q}-c}{2}$

Demonstração. Para provar a primeira parte do lema, tome um q_j qualquer tal que $q_j \leq \overline{q}$. Note que $\pi_i(q_i; q_j)$ é estritamente crescente em q_i para qualquer $q_i < q_i^B(q_j)$. Dado que $\widehat{q}_i < q_i^B(\overline{q}) < q_i^B(q_j)$, uma vez que q_i^B é decrescente, isso implica que

$$\pi(\widehat{q}_i, q_j) < \pi_i(q_i^B(\overline{q}), q_j). \tag{6.38}$$

Isto é, $\widehat{q}_i$ é estritamente dominado por $q_i^B(\overline{q})$.

Para provar a segunda parte do lema, tome um q_j qualquer tal que $q_j \leq \overline{q}$. Note que $\pi_i(q_i; q_j)$ é estritamente decrescente em q_i para qualquer $q_i > q_i^B(q_j)$. Dado que $q_i^B(q_j) < q_i^B(\overline{q}) < \widehat{q}_i$, isso implica que

$$\pi(\widehat{q}_i, q_j) < \pi_i(q_i^B(\overline{q}), q_j). \tag{6.39}$$

Isto é, $\widehat{q}_i$ é estritamente dominado por $q_i^B(\overline{q})$. ∎

Considere a sequência $q^0, q^1, q^2, \ldots$, com $q^0 = 0$ e

$$q^m = q_i^B(q^{m-1}) \equiv \frac{1-q^{m-1}-c}{2} = \frac{1-c}{2} - \frac{q^{m-1}}{2}, \quad \forall m > 0. \tag{6.40}$$

Isto é,

$$\begin{aligned}
q^0 &= 0 && (6.41)\\
q^1 &= \frac{1-c}{2} && (6.42)\\
q^2 &= \frac{1-c}{2} - \frac{1-c}{4} && (6.43)\\
q^2 &= \frac{1-c}{2} - \frac{1-c}{4} + \frac{1-c}{8} && (6.44)\\
&\vdots\\
q^m &= \frac{1-c}{2} - \frac{1-c}{4} + \frac{1-c}{8} - \ldots - (-1)^m \frac{1-c}{2^m} && (6.45)\\
&\vdots
\end{aligned}$$

Teorema 6.4.3. *O conjunto de estratégias restantes após qualquer rodada ímpar* m*,* $m = (1, 3, \ldots)$*, é* $[q^{m-1}, q^m]$*. O conjunto de estratégias restantes após qualquer rodada par* m*,* $m = (2, 4, \ldots)$*, é*

$[q^m, q^{m-1}]$. *O conjunto de estratégias racionalizáveis é* $\left\{\dfrac{1-c}{3}\right\}$.

Vamos agora considerar o caso de três ou mais empresas. Quando existem três ou mais empresas, a racionalização não ajuda: não se pode eliminar qualquer estratégia menos do que a produção monopolista $q = \dfrac{1-c}{2}$.

Na primeira rodada, pode-se eliminar qualquer estratégia $q_i > \dfrac{1-c}{2}$ usando o mesmo argumento no caso do duopólio. Mas na segunda rodada, a oferta total máxima possível pelas outras empresas é

$$(n-1)\frac{1-c}{2} \geq 1-c, \tag{6.46}$$

em que n é o número de firmas. A melhor resposta para este nível de oferta agregada é 0. Portanto, não é possível eliminar qualquer estratégia na rodada 2. O processo de eliminação é interrompido, gerando $\left[0, \dfrac{1-c}{2}\right]$ como o conjunto de estratégias racionalizáveis. Como o conjunto de estratégias racionalizáveis é grande, a racionalização tem um poder preditivo fraco neste jogo.

Embora a racionalização tenha um fraco poder preditivo, na medida em que o conjunto de estratégias racionalizáveis é grande, o equilíbrio de Nash continua a ter um forte poder preditivo. Existe um equilíbrio único de Nash. Lembre-se que $q^* = (q_1^*, q_2^*, \ldots, q_n^*)$ é um equilíbrio de Nash se e somente se

$$q_i^* = q_i^B\left(\sum_{j\neq i} q_j^*\right) = \frac{1 - \sum_{j\neq i} q_j^* - c}{2}, \quad \forall i \tag{6.47}$$

Reescrevendo este sistema mais explicitamente:

$$2q_1^* + q_2^* + \ldots + q_n^* = 1-c \tag{6.48}$$
$$q_1^* + 2q_2^* + \ldots + q_n^* = 1-c \tag{6.49}$$
$$\vdots \tag{6.50}$$
$$q_1^* + q_2^* + \ldots + 2q_n^* = 1-c \tag{6.51}$$

Para qualquer i e j, subtraindo a j-ésima equação da i-ésima equação, obtemos

$$q_i^* - q_j^* = 0. \tag{6.52}$$

Portanto,

$$q_1^* = q_2^* = \ldots = q_n^*. \tag{6.53}$$

Substituindo dentro da primeira equação, encontramos:

$$(n+1)q_1^* = 1 - c, \tag{6.54}$$

ou seja,

$$q_1^* = q_2^* = \ldots = q_n^* = \frac{1-c}{n+1}. \tag{6.55}$$

Portanto, existe um único equilíbrio de Nash, no qual cada firma produz $\frac{1-c}{n+1}$.
No único equilíbrio, a oferta total é

$$Q = \frac{n}{n+1}(1-c), \tag{6.56}$$

e o preço é dado por

$$P = c + \frac{1-c}{n+1}. \tag{6.57}$$

O lucro por firma é

$$\pi = \left(\frac{1-c}{n+1}\right)^2. \tag{6.58}$$

À medida que n vai para o infinito, a oferta total Q converge para $1-c$ e o preço P converge para c. Estes são os valores nos quais a demanda ($P = \max\{1-Q, 0\}$) é igual a oferta ($P = c$), que é chamado de equilíbrio perfeitamente competitivo. Quando há poucas empresas, no entanto, o preço é significativamente maior do que o preço competitivo c, e a oferta total é significativamente menor do que a oferta competitiva $1-c$.

6.4.4.2 Competição em Bertrand

Considere duas empresas. Simultaneamente, cada empresa define um preço p_i. A empresa i com o preço mais baixo $p_i < p_j$ vende $1-p_i$ unidades e a outra empresa não vende nenhuma. Se as firmas definem o mesmo preço, a demanda é dividida igualmente entre elas. Ou seja, a quantidade de vendas para a empresa i é

$$Q_i(p_1, p_2) = \begin{cases} 1 - p_i & \text{se } L\ p_i < p_j \\ \dfrac{1 - p_i}{2} & \text{se } p_i = p_j \\ 0 & \text{caso contrário} \end{cases} \tag{6.59}$$

Suponha que não custa nada produzir o bem (ou seja, $c = 0$). Portanto, o lucro de uma empresa i é

$$\pi_i = p_i Q_i(p_1, p_2) = \begin{cases} (1 - p_i)p_i & \text{se } L\ p_i < p_j \\ \dfrac{(1 - p_i)p_i}{2} & \text{se } p_i = p_j \\ 0 & \text{caso contrário} \end{cases} \tag{6.60}$$

Assumindo que as informações acima são de conhecimento comum, podemos escrever o jogo na forma normal, em que:

- $N = \{1, 2\}$ é o conjunto de jogadores.
- $S_= [0, \infty)$ é o espaço de estratégias do jogador i, em que uma estratégia típica é o preço p_i a ser cobrado pela firma i.
- π_i é a função payoff.

Observe que quando $p_j = 0$, $\pi_i = (p_1, p_2) = 0$ para todo p_i e, portanto, todo p_i é a melhor resposta para $p_j = 0$. Isso tem duas implicações importantes:

1. Toda estratégia é racionalizável (não se pode eliminar nenhuma estratégia, porque cada uma delas é a melhor resposta a zero).
2. $p_1^* = p_2^* = 0$ é um equilíbrio de Nash.

Este é de fato o único equilíbrio de Nash. Em outras palavras, mesmo com duas firmas, quando as firmas competem definindo preços, o equilíbrio competitivo vai emergir. Se modificarmos o jogo ligeiramente, discretizando o conjunto de preços permitidos e colocando um preço mínimo, então o jogo se torna solucionável por dominância, ou seja, apenas uma estratégia permanece racionalizável. No jogo modificado, o preço mínimo é a única estratégia racionalizável, como no equilíbrio competitivo.

Teorema 6.4.4. *No modelo de Bertrand, o único equilíbrio de Nash é* $p^* = (0, 0)$.

Demonstração. Já vimos que $p^* = (0, 0)$ é um equilíbrio de Nash. Eu mostrarei aqui que se (p_1, p_2) é um equilíbrio de Nash, então $p_1 = p_2 = 0$. Para fazer isso, tome qualquer equilíbrio de Nash (p_1, p_2). Primeiro mostro que (p_1, p_2). Por contradição, suponha que $p_i > p_j$. Se $p_j = 0$, então $\pi_i(p_i, p_j) = 0$ enquanto $\pi(p_i, p_j) = \dfrac{(1 - p_i)p_i}{2}$, ou seja, escolher p_i é um desvio lucrativo para a empresa j, mostrando que $p_i > p_j = 0$ não é um equilíbrio de Nash. Portanto, para $p_i > p_j$ ser um equilíbrio, deve ser que $p_j > 0$. Mas, então, a empresa i tem um desvio lucrativo: $\pi_i(p_i, p_j) = 0$ enquanto $\pi(p_i, p_j) = \dfrac{(1-p_i)p_i}{2}$. Em suma, isso mostra que não se pode ter$p_i > p_j$ em equilíbrio. Portanto, $p_1 = p_2$. Mas se $p_1 = p_2$ é um equilíbrio de Nash, então deve ser que $p_1 = p_2 = 0$. Isto ocorre porque se $p_1 > p_2 = 0$, então a firma 1 teria um desvio lucrativo e esse desvio tenderia a voltar ao equilíbrio. ∎

6.5 Jogos Dinâmicos com Informações Completas

6.5.1 Indução Retroativa

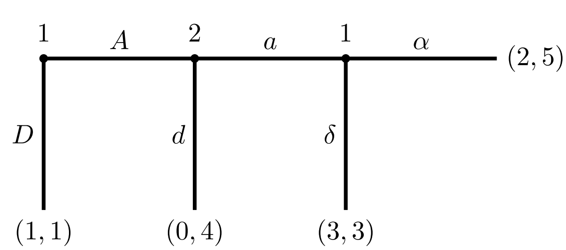
Agora começamos a analisar os jogos dinâmicos com informações completas. Esta seção concentra-se nos jogos de informação perfeita, em que cada conjunto de informações é único e se aplica a noção de indução retroativa a esses jogos. Vamos supor que o jogo tem "horizonte finito", ou seja, só pode haver movimentos finitos em qualquer histórico de movimentos.

O conceito de indução retroativa corresponde à suposição de que é de conhecimento comum que cada ator irá agir racionalmente em cada nó futuro para onde ele se move − mesmo que sua racionalidade implique que tal nó não será alcançado. A suposição de que o jogador se move racionalmente em cada conjunto de informações que ele move é chamada de racionalidade sequencial.

Mecanicamente, a indução inversa corresponde ao seguinte procedimento. Considere qualquer nó que vem logo antes dos nós terminais, isto é, após cada movimento que se origina deste nó, o jogo termina. Tais nós são chamados de nós terminais. Se o jogador que se move nesse nó age racionalmente, ele escolhe o melhor movimento para ele nesse nó. Por isso, selecione um dos movimentos que dão maior retorno ao jogador. Atribuindo o vetor de payoff associado a essa movimentação para o nó em questão, exclua todos os movimentos decorrentes desse nó para que tenhamos um jogo mais curto, em que o nó acima é um nó terminal. Repita este procedimento até que a origem seja o único nó restante. O resultado é os movimentos que são escolhidos no caminho. Como um movimento é escolhido em cada conjunto de informações, o resultado é um perfil de estratégia.

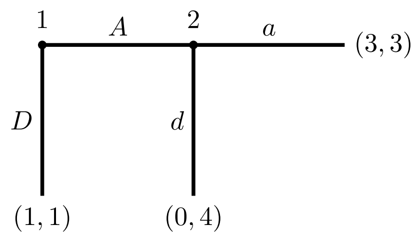
Para uma ilustração do procedimento, considere o jogo na figura a seguir. Este jogo descreve uma situação em que é mutuamente benéfico para todos os jogadores permanecerem em um relacionamento, enquanto um jogador gostaria de sair do relacionamento, se ela souber que o outro jogador sairá no dia seguinte.

Figura 6.13 – Jogo do relacionamento



No terceiro dia, o jogador 1 se move, escolhendo entre ficar (α) ou sair (δ). Se ele ficar, ele ganharia 2; se ele sair, ele obterá a maior recompensa de 3. Assim, de acordo com o procedimento, ele sai. Selecionando o movimento δ para o nó em questão, reduz-se o jogo da seguinte forma:

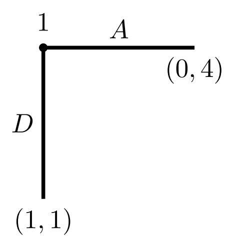
Figura 6.14 – Jogo do relacionamento



Aqui, a parte do jogo que começa no último nó de decisão é substituída pelo vetor de payoff associado ao movimento selecionado, δ, do jogador neste nó de decisão.

No segundo dia, o jogador 2 se move, escolhendo entre ficar (a) ou sair (d). Se ela ficar, ela ganha 3; se ela sair, ela recebe o maior pagamento de 4. Assim, de acordo com o procedimento, ela sai. Selecionando o movimento d para o nó em questão, reduz-se o jogo da seguinte forma:

Figura 6.15 – Jogo do relacionamento



Mais uma vez, a parte do jogo que começa com o nó em questão é substituída pelo vetor de retorno associado ao movimento selecionado, d. Agora, o jogador 1 recebe se ele ficar, (A), e recebe 1 se ele sair (D). Portanto, ele sair. O procedimento resulta no seguinte perfil de estratégia:

Ou seja, em cada nó, o jogador que deve se mover na direção de sair do relacionamento.

Vamos analisar as suposições que fizemos na construção desse perfil de estratégia. Assumimos que o jogador 1 irá agir racionalmente na última data, quando consideramos que ele sai. Quando calculamos que o jogador 2 sai no segundo dia, assumimos que o jogador 2 assume que o jogador 1 irá agir racionalmente no terceiro dia, e também assumiu que ela é racional. No primeiro dia, o jogador 1 antecipa tudo isso. Isto é, presume-se que ele saiba que o jogador 2 é racional e que continuará acreditando que o jogador 1 irá agir racionalmente no terceiro dia.

Este exemplo também ilustra outra noção associada à indução retroativa - comprometimento (ou falta de comprometimento). Observe que os resultados no terceiro dia (ou seja, $(3, 3)$ e $(2, 5)$) são ambos estritamente melhores do que o resultado de equilíbrio $(1, 0)$. Mas eles não podem alcançar estes resultados, porque o jogador 2 não pode comprometer-se a ficar, e antecipando que o jogador 2 irá sair, o jogador 1 fica no relacionamento no primeiro dia. Há também um outro problema de compromisso neste exemplo. Se o jogador 1 conseguir se comprometer a ficar no terceiro dia, então o jogador 2 definitivamente entrará em ação no segundo dia. Nesse caso, o jogador 1 passaria no primeiro. É claro que o jogador 1 não pode se comprometer a ficar no terceiro dia, e o jogo termina no primeiro dia, gerando os baixos ganhos $(1, 0)$.

6.5.2 Indução Retroativa e Equilíbrio de Nash

Leitores cuidadosos devem ter notado que o perfil de estratégia resultante da indução retroativa acima é um equilíbrio de Nash. Isto não é uma coincidência.

Proposição 6.5.1. *Em um jogo com finitos nós, a indução retroativa sempre resulta em um equilíbrio de Nash.*

É tentador concluir que a indução retroativa resulta no equilíbrio de Nash, porque um deles reproduz uma melhor resposta em cada nó. Como se tomam seus movimentos futuros dados e seleciona-se apenas um movimento para o nó em questão, fazer os melhores movimentos nos nós dados não leva necessariamente a uma melhor resposta entre todos os planos contingentes em geral.

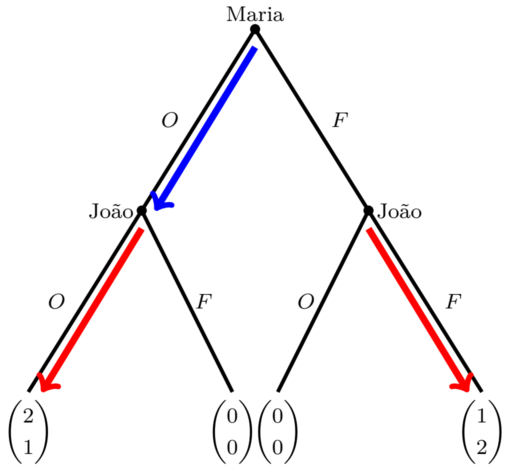
Considere um único jogador, que escolhe entre o bem e o mal todos os dias para sempre. Se ele escolhe bem todos os dias, ele recebe 1, e ele recebe 0 caso contrário. Claramente, o plano ideal é jogar bem todos os dias, produzindo 1. Agora considere a estratégia de acordo com a qual ele joga mal todos os dias em todos os nós. Isso dá a ele 0. Mas a estratégia satisfaz a condição de indução retroativa (embora a indução retroativa não possa ser aplicada a este jogo sem nenhum nó final). Em qualquer nó, de acordo com os movimentos selecionados no futuro, ele recebe zero, independentemente do que ele faz no nó atual.

O caso patológico acima é um contraexemplo para a ideia de que, se alguém está jogando o melhor movimento em cada nó, seu plano é a melhor resposta. A última ideia é um princípio importante de otimização dinâmica, chamado Princípio do Desvio Único[2]. Aplica-se na maioria dos casos, exceto para os casos patológicos como acima. O Princípio do Desvio Único será a principal ferramenta na análise dos jogos do horizonte infinito nas próximas seções.

Mas nem todos os equilíbrios de Nash podem ser obtidos por indução retroativa. Considere o seguinte jogo da Batalha dos Sexos com informação perfeita, onde Maria se move primeiro.

[2] O princípio do desvio único é fundamental para a teoria dos jogos extensivos. Foi originalmente formulado por David Blackwell (1965) no contexto de programação dinâmica. Como a estratégia de outros jogadores induz um problema normal de maximização para qualquer jogador, podemos formular o princípio no contexto de uma árvore de decisão de uma única pessoa.

Figura 6.16 – Jogo batalha dos sexos



Neste jogo, a indução retroativa leva ao perfil de estratégia identificado na figura, de acordo com o qual João vai para onde Maria vai e Maria vai para a Ópera. Há outro equilíbrio de Nash: Maria vai para o jogo de futebol e João vai para o jogo de futebol em ambos os seus nós de decisão. Vamos ver porque este é um equilíbrio de Nash. Maria tem uma melhor resposta à estratégia de João: se ela vai para o futebol, ela ganha 1, e se ela for para a ópera, ela recebe 0 (como eles não se encontram). A estratégia de João também é a melhor resposta à estratégia de Maria: sob essa estratégia, ele recebe 2, que é o maior que ele pode obter neste jogo.

Pode-se, no entanto, desacreditar o último equilíbrio de Nash porque ele se baseia em um movimento sequencialmente irracional no nó após Maria ir à Ópera. Esse nó não acontece de acordo com a estratégia de Maria e, portanto, é ignorado no equilíbrio de Nash. No entanto, se Maria for à Ópera, ir ao jogo de futebol seria irracional para João, e ele também iria racionalmente à Ópera. E Maria deveria prever isso e ir à Ópera. Às vezes, dizemos que esse equilíbrio é baseado em "uma ameaça não crível", com a interpretação óbvia.

Este exemplo ilustra uma deficiência da condição de racionalidade usual, que exige que se tenha uma melhor resposta (como um plano contingente completo) no início do jogo. Enquanto isto requer que o jogador jogue uma melhor resposta nos nós que ele atribui probabilidade positiva, deixa o jogador livre para escolher qualquer movimento nos nós que ele coloca probabilidade zero - porque todos os payoffs após esses nós são multiplicados por zero no cálculo de utilidade esperada. Como as probabilidades dos nós são determinadas como parte da solução, isso pode levar a soluções um tanto errôneas nas quais um nó não é alcançado porque um jogador joga irracionalmente no nó, antecipando que o nó não será alcançado em equilíbrio. Naturalmente, isso é errôneo, pois quando esse nó é alcançado, o jogador não pode fingir que o nó não será alcançado, pois ele saberá que é atingido pela definição do conjunto de informações. Então, ele deve jogar uma melhor resposta, uma vez que o nó é alcançado.

6.5.3 Comprometimento

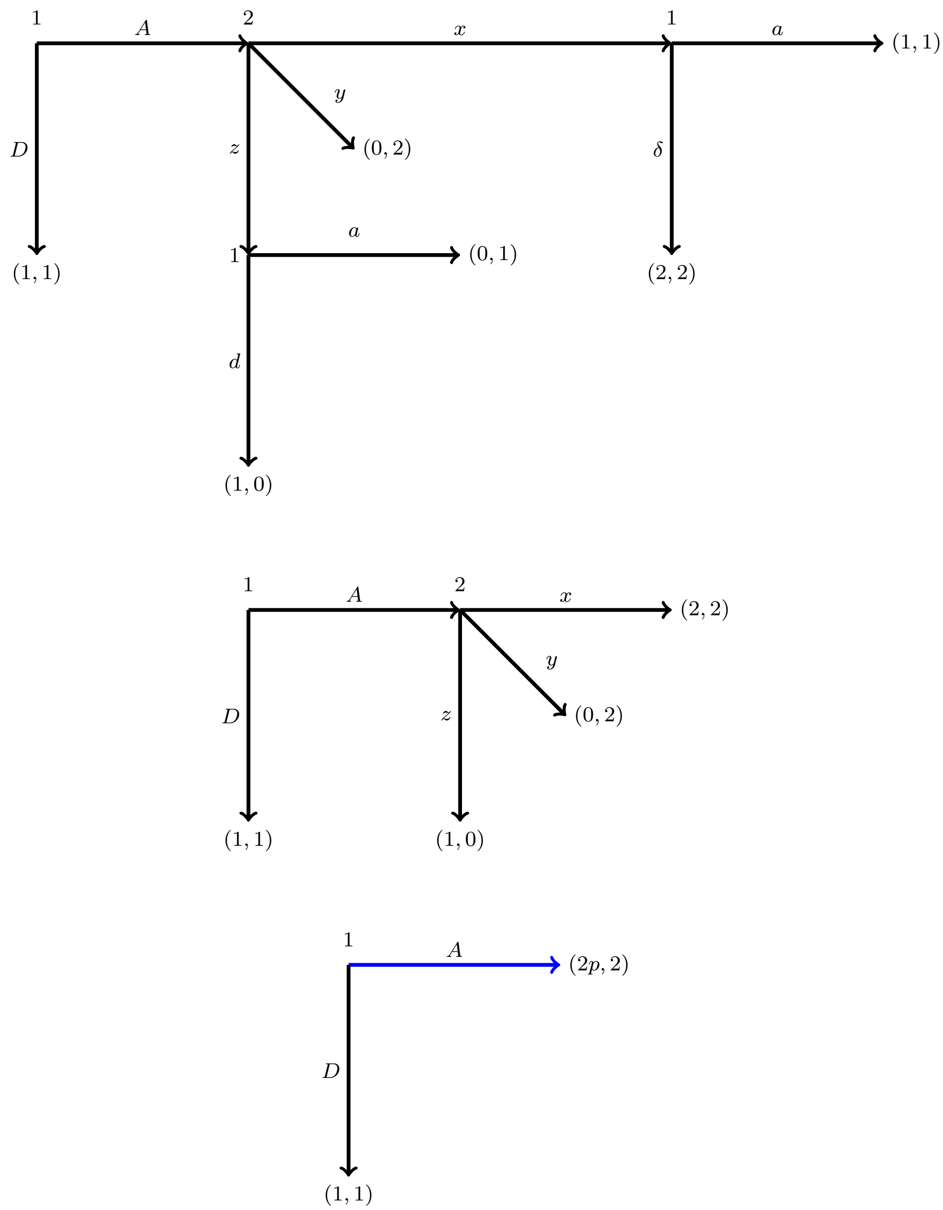
Neste jogo, Maria pode se comprometer a ir a um lugar, mas João não pode. Se confiarmos no resultado da indução retroativa, esse compromisso ajudará Maria e machucará João. (Embora o jogo seja simétrico, Maria recebe uma recompensa maior.) É tentador concluir que a capacidade de comprometimento é sempre boa. Embora isso seja verdade em muitos jogos, em alguns jogos não é o caso.

6.5.4 Múltiplas Soluções

A indução retroativa pode levar a múltiplos equilíbrios. Aqui, em um equilíbrio, o jogador 1 escolhe cara, em outro o jogador 1 escolhe coroa, e ainda em outro equilíbrio de estratégia mista, ele mistura as duas estratégias. Cada probabilidade de mistura corresponde a um equilíbrio diferente. Em todos esses equilíbrios, os payoffs são os mesmos. Em geral, no entanto, a indução retroativa pode levar a múltiplos equilíbrios com resultados bem diferentes.

Considere o jogo na Figura 6.17. De acordo com a indução retroativa, nos seus nós à direita e na parte inferior, o jogador 1 desce, escolhendo δ e a, respectivamente. Isso leva ao jogo reduzido. Claramente, no jogo reduzido, tanto x como y rendem 2 para o jogador 2, enquanto z rende 1. Portanto, ela deve escolher x ou y ou qualquer randomização entre os dois. Em outras palavras, para qualquer $p \in [0, 1]$, a estratégia mista que coloca p em x, $1 - p$ em y e 0 em z pode ser selecionada pela indução retroativa. Selecione essa estratégia. Então, o vetor de payoff associado à decisão do jogador 2 é $(2p, 2)$.

Figura 6.17 – Um jogo com múltiplos equilíbrios de indução retroativa



A estratégia selecionada para o jogador 1 depende da escolha de p. Se algum $p > \frac{1}{2}$ for selecionado para o jogador 2, o jogador 1 deve escolher A. Isso resulta no equilíbrio no qual o jogador 1 joga Add e o jogador 2 joga x com probabilidade p e y com probabilidade $1 - p$. Se $p < \frac{1}{2}$, o jogador 1 deve escolher D. No equilíbrio resultante, o jJogador 1 joga $Dd\delta$ e o jogador 2 joga x com probabilidade p e y com probabilidade $1 - p$. Finalmente, se $p = \frac{1}{2}$ for selecionado, então o jogador 1 é indiferente, e podemos selecionar qualquer randomização entre A e D, cada um resultando em um equilíbrio diferente.

6.5.5 Duopólio de Stackelberger

No duopólio de Cournot, assumimos que as empresas definem as quantidades simultaneamente. Isso reflete a suposição de que nenhuma empresa pode se comprometer com um nível de quantidade. Às vezes, uma empresa pode se comprometer com um nível de quantidade. Por exemplo, uma empresa pode já estar no mercado e construir sua fábrica e armazéns, e seu nível de produção é fixo. A outra firma entra no mercado sabendo mais tarde o nível de produção da primeira firma. Vamos considerar tal situação, que é chamada de duopólio de Stackelberg. Existem duas empresas. A primeira firma é chamada de líder, e a segunda firma é chamada de seguidora. Como antes, tomamos o custo marginal constante.

- A firma líder escolhe seu nível de produção q_1.
- Então, conhecendo q_2, a seguidora escolhe seu nível de produção q_2.
- Cada firma i escolhe sua quantidade q_i aos preços de mercado,

$$P(q_1 + q_2) = \max\{1 - (q_1 + q_2), 0\}, \tag{6.61}$$

resultando em um lucro

$$\pi_i(q_1, q_2) = q_i(P(q_1 + q_2) - c). \tag{6.62}$$

- No nó inicial, a firma 1 escolhe uma ação q_1, em que o conjunto de ações disponíveis é $[0, \infty)$.
- Depois da ação da firma 1, a firma 2 se move e escolhe uma ação q_2, em que o conjunto de ações disponíveis é $[0, \infty)$.
- Cada uma das ações é levada ao nó terminal, no qual o vetor de payoffs é $(\pi_1(q_1, q_2), \pi_2(q_1, q_2))$.

Observe que a estratégia da empresa 1 é um número real q_1 definido no intervalo $[0, \infty)$ e, mais importante, a estratégia da empresa 2 é uma função definido no intervalo $[0, \infty) \times [0, \infty)$, que atribui um nível de produção $q_2(q_1)$ para cada q_1. Essas estratégias com a função utilidade $u_i(q_1, q_2) = \pi_i(q_1, q_2(q_1))$ nos dão a forma normal do jogo.

Vamos resolver por indução retroativa. Dado que $q_1 \leq 1 - c$, o melhor nível de produção para a seguidora é

$$q_2^*(q_1) = \frac{1 - q_1 - c}{2}, \tag{6.63}$$

resultando no seguinte vetor de payoffs[3]:

[3] Note que $q_1\left(1 - q_1 - \dfrac{1 - q_1 - c}{2} - c\right) = \dfrac{1}{2}q_1(1 - q_1 - c)$.

$$\begin{pmatrix} \pi_1(q_1, q_2^*(q_1)) \\ \pi_2(q_1, q_2^*(q_1)) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}q_1(1-q_1-c) \\ \frac{1}{4}(1-q_1-c)^2 \end{pmatrix}. \tag{6.64}$$

Substituindo os movimentos da empresa 2 com os payoffs associados, obtemos um jogo no qual a empresa escolhe um nível de produção q_1 que leva ao vetor de payoffs acima. Neste jogo, a empresa 1 maximiza seu lucro ao escolher

$$q_1^* = \frac{1-c}{2}. \tag{6.65}$$

Também devemos verificar se há um equilíbrio de Nash desse jogo no qual o seguidor produz a quantidade de Cournot independentemente do que o líder produz, e o líder produz a quantidade de Cournot. É claro que isso não é consistente com a indução retroativa: quando o seguidor sabe que o líder produziu a quantidade de Stackelberg, ele mudará de ideia e produzirá uma quantidade menor, a quantidade computada durante a indução retroativa.

6.6 Equilíbrio de Nash Perfeito de Subjogos

A indução retroativa é um poderoso conceito de solução com algum apelo intuitivo. Infelizmente, ele pode ser aplicado apenas para jogos de informação perfeitos com um horizonte finito. Sua intuição, no entanto, pode ser estendida para além desses jogos através da perfeição do subjogo. Esta seção define o conceito de equilíbrio perfeito para subjogos e ilustra como se pode verificar se um perfil estratégico é um equilíbrio perfeito para subjogos.

6.6.1 Definição e Exemplos

Um jogo na forma extensiva pode conter uma parte que poderia ser considerada um jogo menor em si; um jogo tão pequeno que é incorporado em um jogo maior é chamado de sub-jogo. Uma propriedade principal da indução retroativa é que, quando restrito a um sub-jogo, o equilíbrio computado usando indução retroativa permanece como um equilíbrio (computado novamente via indução retroativa) do sub-jogo. A perfeição do sub-jogo generaliza essa noção para jogos dinâmicos gerais:

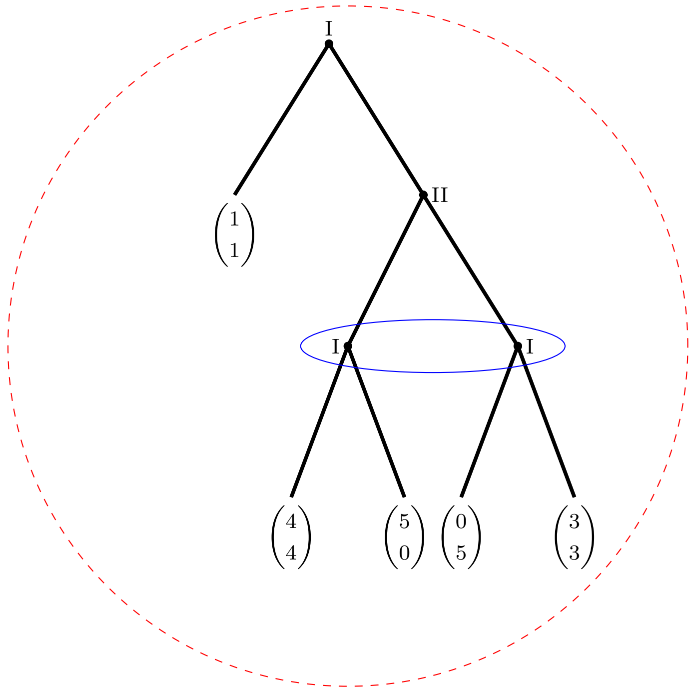
Definição 6.6.1. *Um equilíbrio de Nash é dito ser um subjogo perfeito se e somente se é um equilíbrio de Nash em todos os sub-jogos do jogo.*

Um sub-jogo deve ser um jogo bem definido quando é considerado separadamente. Isso é,

1. deve conter um nó inicial,

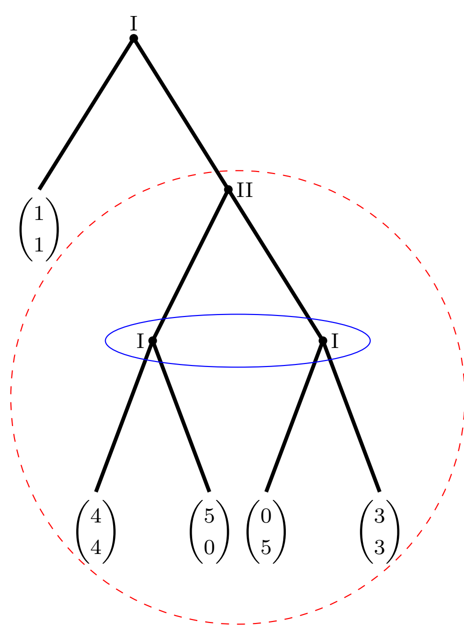
2. todos os movimentos e conjuntos de informações desse nó devem permanecer no sub-jogo.

Figura 6.18 – Exemplo de um jogo que é um subjogo



O jogo todo é um subjogo

Figura 6.19 – Exemplo de um subjogo



Este é outro subjogo

Figura 6.20 – Exemplo que não é um subjogo

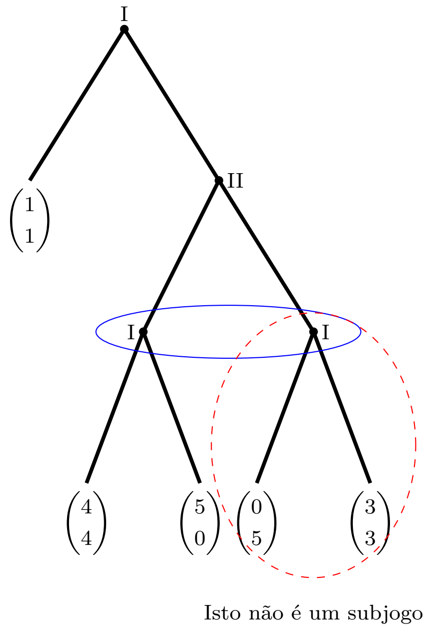
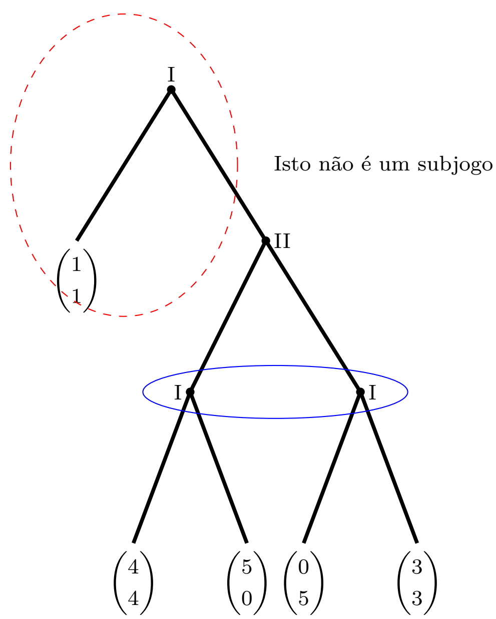


Figura 6.21 – Exemplo que não é um subjogo



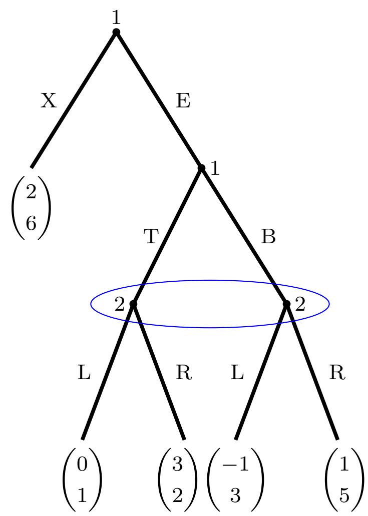
Vamos ver uma técnica para calcular os equilíbrios perfeitos do subjogo em jogos finitos:

1. Escolha um sub-jogo que não contenha nenhum outro sub-jogo.
2. Calcule um equilíbrio de Nash deste jogo. Atribua o vetor de payoff associado a este equilíbrio ao nó inicial e elimine o sub-jogo.

3. Itere este procedimento até que um movimento seja atribuído a cada contingência, quando não houver nenhum sub-jogo para eliminar.

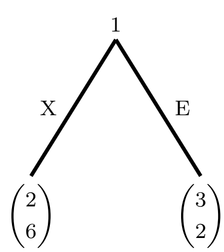
Vejamos o seguinte exemplo.

Figura 6.22 – Jogo de informação imperfeita



Não se pode aplicar indução reversa neste jogo porque não é um jogo de informação perfeita. Pode-se, porém, calcular o equilíbrio do sub-jogo. Este jogo tem dois subjogos: um começa depois que o jogador 1 joga E; o segundo é o jogo em si. Os equilíbrios perfeitos do subjogo são calculados da seguinte forma. Primeiro calcule um equilíbrio de Nash do sub-jogo, e então fixe as ações de equilíbrio como estão (neste sub-jogo), e considerando os payoffs de equilíbrio neste sub-jogo como os payoffs para entrar no sub-jogo, calcule um equilíbrio de Nash no jogo remanescente.

O subjogo tem apenas um equilíbrio de Nash, como T domina B e R domina L. No único equilíbrio de Nash, o jogador 1 joga T e o jogador 2 joga R, produzindo o vetor payoff $(3, 2)$, como ilustrado na Figura 11.3. Dado isto, o jogo reduz para

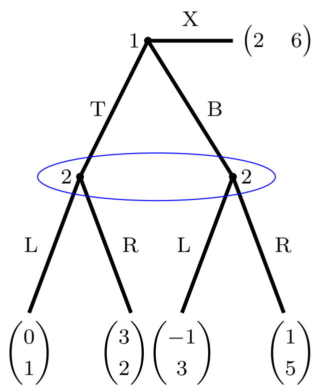


O jogador 1 escolhe E neste jogo reduzido. Portanto, no equilíbrio perfeito do subjogo, o jogador 1 escolhe E e depois joga (T, R) simultaneamente.

Como na indução retroativa, quando há múltiplos equilíbrios no subjogo escolhido, pode-se escolher qualquer um dos equilíbrios de Nash, incluindo um em uma estratégia mista. Cada escolha de equilíbrio leva a um diferente equilíbrio perfeito de Nash no subjogo no jogo original. Variando o equilíbrio de Nash para os subjogos à mão, pode-se calcular todos os equilíbrios de Nash perfeitos do subjogo.

Um equilíbrio de Nash perfeito no subjogo é um equilíbrio de Nash, porque o jogo inteiro também é um sub-jogo. O inverso não é verdadeiro. Pode haver um equilíbrio de Nash que não seja perfeito para subjogos. Por exemplo, seja o jogo abaixo.

Figura 6.23 – Jogo de informação imperfeita



Às vezes, o equilíbrio perfeito do subjogo pode ser altamente sensível à forma como modelamos a situação. Por exemplo, considere o jogo na Figura acima. Este é essencialmente o mesmo jogo que acima. A única diferença é que o jogador 1 faz suas escolhas aqui de uma só vez. Alguém poderia pensar que tal escolha de modelagem não deveria fazer diferença na solução do jogo. No entanto, faz uma enorme diferença para o equilíbrio perfeito de Nash no subjogo. No novo jogo, o único sub-jogo deste jogo é ele mesmo, portanto, qualquer equilíbrio de Nash é um equilíbrio perfeito de Nash do subjogo. No novo jogo, ele é formalmente escrito como o perfil de estratégia (X, L). Claramente, poder-se-ia ter usado a ideia de racionalidade sequencial para resolver este jogo. Isto é, pela racionalidade sequencial do jogador 2 em seu conjunto de informações, ele deve escolher R. Sabendo disso, o jogador 1 deve escolher T. Portanto, o equilíbrio perfeito do subjogo não formaliza totalmente a idéia de racionalidade sequencial. Produz soluções razoáveis em muitos jogos e é amplamente utilizado na teoria dos jogos.

6.6.2 Princípio do Desvio Único

Pode ser difícil verificar se um perfil de estratégia é um equilíbrio perfeito de subjogo em jogos de horizonte infinito, onde alguns caminhos no jogo podem durar para sempre sem terminar o jogo. No entanto, existe uma técnica simples que pode ser usada para verificar se um perfil de estratégia é um equilíbrio perfeito para subjogos na maioria dos jogos. A técnica é chamada de princípio do desvio único. Em um jogo pode haver histórias em que todas as ações anteriores são conhecidas, mas os jogadores podem se mover simultaneamente. Tais histórias são chamadas de etapas. Por exemplo, suponha que todos os dias os jogadores joguem a batalha dos sexos, sabendo o que cada jogador jogou em cada dia anterior. Nesse caso, a cada dia, após qualquer histórico de jogo nos dias anteriores, temos um estágio no qual os jogadores se movem simultaneamente e um novo sub-jogo é iniciado.

Esses jogos são chamados de jogos multi-estágio. Em um jogo de múltiplos estágios, se duas estratégias prescrevem o mesmo comportamento em todos os estágios, elas são estratégias idênticas e geram o mesmo vetor de payoffs. Suponha que duas estratégias sejam diferentes, mas elas prescrevem o mesmo comportamento para estágios sucessivos muito, muito longos, por exemplo. Então, esperamos que as duas estratégias gerem retornos muito semelhantes. Se este é realmente o caso, então chamamos esses jogos "contínuos no infinito". O princípio de desvio único se aplica a jogos multiestágios que são contínuos no infinito.

Considere um perfil de estratégia s^*. Escolha qualquer estágio (após qualquer histórico de movimentos). Suponha que estamos nesse estágio. Escolha também um jogador que se mova naquele estágio. Conserte todos os movimentos dos outros jogadores conforme prescrito pelo perfil de estratégia s^* no estágio atual, assim como no jogo a seguir. Corrija também os movimentos do jogador i em todas as datas futuras, mas deixe seus movimentos no estágio atual variarem. Podemos encontrar um movimento no estágio atual que dê uma recompensa maior do que s^*, considerando todas as jogadas que consertamos? Se a resposta for sim, então s^* falha no teste de desvio único nessa fase para o jogador i.

Claramente, se s^* falhar no teste de desvio único em qualquer estágio para qualquer jogador i, então s^* não pode ser um equilíbrio perfeito para o subjogo. Isto porque s^* não leva a um equilíbrio de Nash no sub-jogo que começa naquele estágio, pois o jogador i tem um incentivo para desviar para a estratégia de acordo com a qual i joga melhor no estágio atual, mas segue s_i^* no restante do sub-jogo. Acontece que em um jogo multi-estágio que é contínuo no infinito, o inverso também é verdadeiro. Se s^* passa no princípio do desvio único em cada estágio (após cada histórico de movimentos anteriores) para cada jogador, então é um equilíbrio perfeito para o subjogo.

Teorema 6.6.1. *Em um jogo multi-estágio que é contínuo no infinito, um perfil de estratégia é um equilíbrio de Nash perfeito para o subjogo se e somente se ele passar no teste de desvio único em cada estágio para cada jogador.*

6.6.3 Barganha Sequencial

Imagine que dois jogadores possuem um real, que eles podem usar somente depois de decidirem como dividi-lo. Cada jogador é neutro ao risco e desconta o futuro exponencialmente. Ou seja, se um jogador recebe x reais no dia t, seu pagamento é $\delta^t x$ para algum $\delta \in (0, 1)$. O conjunto de todas as divisões viáveis é $D = \{(x, y) \in [0, 1] \times [0, 1] | x + y \leq 1\}$. Os jogadores estão negociando a divisão do real fazendo ofertas e contra-ofensivas, como ficará claro. Queremos aplicar indução retroativa a este jogo para entender quando as partes chegarão a um acordo e quais serão os termos do acordo.

Primeiro, considere o seguinte modelo simplificado com apenas duas rodadas de negociações. No primeiro dia, o jogador 1 faz uma oferta $(x_1, y_1) \in D$ Então, sabendo o que foi oferecido, o jogador 2 aceita ou rejeita a oferta. Se ele aceitar a oferta, a oferta é implementada, gerando payoffs (x_1, y_1) Se ele rejeitar a oferta, esperará até o dia seguinte, quando o jogador 2 fizer uma oferta $(x_2, y_2) \in D$. Agora, sabendo o que o jogador 2 ofereceu, o jogador 1 aceita ou rejeita a oferta. Se o jogador 1 aceitar a oferta, a oferta é implementada, produzindo recompensas $(\delta x_2, \delta y_2)$. Se o jogador 2 rejeitar a oferta, então o jogo termina, quando eles perdem o real e recebem payoffs $(0, 0)$.

A análise de indução retroativa deste modelo simplificado é a seguinte. No segundo dia, se o jogador 1 rejeitar a oferta, ele recebe 0. Assim, ele aceita qualquer oferta que lhe dê mais que 0, e ele é indiferente entre aceitar e rejeitar qualquer oferta que lhe dê 0. Como vimos no seção anterior, ele aceita a oferta $(0, 1)$ em equilíbrio. Então, no segundo dia, o jogador 2 ofereceria $(0, 1)$, que é o melhor que o jogador 2 pode receber. Portanto, se eles não concordarem no primeiro dia, então o jogador 2 recebe o real inteiro no segundo dia, deixando o jogador 1 sem nada. O valor de tomar o real no dia seguinte para o jogador 2 é δ. Assim, no primeiro dia, o jogador 2 aceita qualquer oferta que lhe dê mais do que δ, rejeita qualquer oferta que lhe dê menos de δ e é indiferente entre aceitar e rejeitar qualquer oferta que lhe dê δ. Suponha que o jogador 2 aceita a oferta $(1 - \delta, \delta)$. Nesse caso, o jogador 1 oferece $(1 - \delta, \delta)$ que é aceito. O jogador 1 poderia receber mais de $1 - \delta$? Se ele oferecesse algo melhor do que $1 - \delta$ para si mesmo, sua oferta seria necessariamente menor que δ para o jogador 2, e o jogador 2 rejeitaria a oferta. Nesse caso, as negociações continuariam até o dia seguinte e ele receberia 0, o que é claramente pior do que $1 - \delta$.

Agora, considere o jogo em que o jogo acima é repetido n vezes. Ou seja, se eles ainda não chegaram a um acordo até o final do segundo dia, no terceiro dia, o jogador 1 faz uma oferta $(x_3, y_2) \in D$. Então, sabendo o que foi oferecido, o jogador 2 aceita ou rejeita a oferta. Se ele aceitar a oferta, a oferta será implementada, gerando payoffs $(\delta^2 x_3, \delta^2 y_3)$ Se ele rejeitar a oferta, esperará até o dia seguinte, quando o jogador 2 fizer uma oferta $(x_4, y_4) \in D$. Agora, sabendo o que o jogador 2 ofereceu, o jogador 1 aceita ou rejeita a oferta. Se o jogador 1 aceitar a oferta, a oferta é implementada, oferecendo retornos $(\delta^3 x_4, \delta^3 y_4)$ Se o jogador 1 rejeitar a oferta, então eles vão para o 5º dia e isso continua assim até o final do dia $2n$. Se eles ainda não concordaram no final do dia, o jogo termina, eles perdem o dólar e recebem payoffs $(0, 0)$.

A aplicação da indução retroativa a este jogo resulta no perfil de estratégia a seguir. Em qualquer $t = 2n - 2k$ (k é um inteiro não-negativo), o jogador 1 aceita qualquer oferta (x, y) com

$$x \geq \frac{\delta(1 - \delta^{2k})}{1 + \delta}, \tag{6.66}$$

e rejeita qualquer oferta com

$$x < \frac{\delta(1 - \delta^{2k})}{1 + \delta}. \tag{6.67}$$

O jogador 2 oferece

$$(x_t, y_t) = \left(\frac{\delta(1 - \delta^{2k})}{1 + \delta}, 1 - \frac{\delta(1 - \delta^{2k})}{1 + \delta} \right) \equiv \left(\frac{\delta(1 - \delta^{2k})}{1 + \delta}, \frac{1 + \delta^{2k+1}}{1 + \delta} \right). \tag{6.68}$$

Em $t - 1 = 2n - 2k - 1$, o jogador 2 aceita uma oferta se e somente se

$$y \geq \frac{\delta(1+\delta^{2k+1})}{1+\delta}, \tag{6.69}$$

e o jogador 1 oferece

$$(x_{t-1}, y_{t-1}) = \left(1 - \frac{\delta(1+\delta^{2k+1})}{1+\delta}, \frac{\delta(1+\delta^{2k+1})}{1+\delta}\right) \equiv \left(\frac{1-\delta^{2k+2}}{1+\delta}, \frac{\delta(1+\delta^{2k+1})}{1+\delta}\right). \tag{6.70}$$

Vamos ilustrar como aplicar o princípio do desvio-único no jogo de barganha com horizonte infinito com ofertas alternadas. O jogo é o mesmo que o analisado acima, exceto que não há data final. Ou seja, se uma oferta for rejeitada, sempre prosseguiremos para a próxima data em que o outro jogador fizer uma oferta. Note que o jogo é contínuo no infinito, pois se duas estratégias descrevem o mesmo comportamento nos primeiros t períodos, a diferença de pagamento sob as duas estratégias não pode exceder δ^t, que tende a zero quando t tende a ∞.

Quando $n \to \infty$, temos o seguinte comportamento:

- s_i^*: em cada histórico onde i faz uma oferta, ofereça para receber $\frac{1}{1+\delta}$ e deixar $\frac{\delta}{1+\delta}$ para o outro jogador, e em cada histórico onde i responder a uma oferta, aceite a oferta se e somente se a oferta der para i pelo menos $\frac{\delta}{1+\delta}$.

Agora vamos usar o princípio do desvio simples para verificar se s^* é um equilíbrio perfeito para o subjogo. Existem dois tipos de etapas: (i) um jogador faz uma oferta, (ii) um jogador responde a uma oferta.

Primeiro, considere um estágio como em (ii) para algum t, em que a oferta atual dá $x_j \geq \frac{\delta}{1+\delta}$ ao jogador j. Corrija a estratégia do jogador i deste estágio como em s_i^*, isto é, em $t+1$ o jogador i aceita uma oferta se sua parte for pelo menos igual a $\frac{\delta}{1-\delta}$, e ele oferecer $\frac{\delta}{1+\delta}$ para o outro jogador sempre que ele for fazer uma oferta. Da mesma forma, corrija a estratégia do jogador j em $t+1$ como em s_j^*, de modo que em $t+1$ e depois j oferece $\frac{\delta}{1+\delta}$ para i aceitar uma oferta se e somente se j obtiver pelo menos $\frac{\delta}{1+\delta}$. De acordo com o comportamento fixo, em $t+1$, j oferece tomar $\frac{1}{1+\delta}$ para si mesmo, deixando $\frac{\delta}{1+\delta}$ para I e a oferta é aceita; o retorno de j associado a esse resultado é

$$\delta^{t+1} \cdot \frac{1}{1-\delta} = \frac{\delta^{t+1}}{1-\delta}. \tag{6.71}$$

Agora, de acordo com s_j^*, no estágio atual, j deve aceitar a oferta. Isso dá a recompensa de

$$\delta^t x_j \geq \frac{\delta^{t+1}}{1-\delta}. \tag{6.72}$$

Se j se desvia e rejeita a oferta, então, de acordo com o comportamento fixo, ele recebe apenas $\dfrac{\delta^{t+1}}{1-\delta}$, e ele não tem incentivo para se desviar. Por isso, s^* passa no teste de desvio único nesta fase para o jogador j.

6.7 Jogos Repetidos

Na vida real, a maioria dos jogos é jogada dentro de um contexto maior, e as ações em uma determinada situação afetam não apenas a situação atual, mas também as situações futuras que podem surgir. Quando um jogador atua em uma determinada situação, ele leva em conta não apenas as implicações de suas ações para a situação atual, mas também suas implicações para o futuro. Se os jogadores são pacientes e as ações atuais têm implicações significativas para o futuro, então as considerações sobre o futuro podem assumir o controle. Isso pode levar a um rico conjunto de comportamentos que podem parecer irracionais quando se considera a situação atual. Tais idéias são capturadas nos jogos repetidos, nos quais um jogo estágio é jogado repetidamente. O jogo estágio é repetido independentemente do que foi jogado nos jogos anteriores.

6.7.1 Jogos Repetidos Finitamente

Seja $T = \{0, 1, \dots, n\}$ o conjunto de todas as datas possíveis. Considere um jogo em que em cada um dos jogadores jogue um jogo estágio G, sabendo o que cada jogador jogou no passado. Suponha que o payoff de cada jogador neste jogo maior seja a soma dos payoffs que ele obtém nos jogos estágio. Denote o jogo maior por G^T.

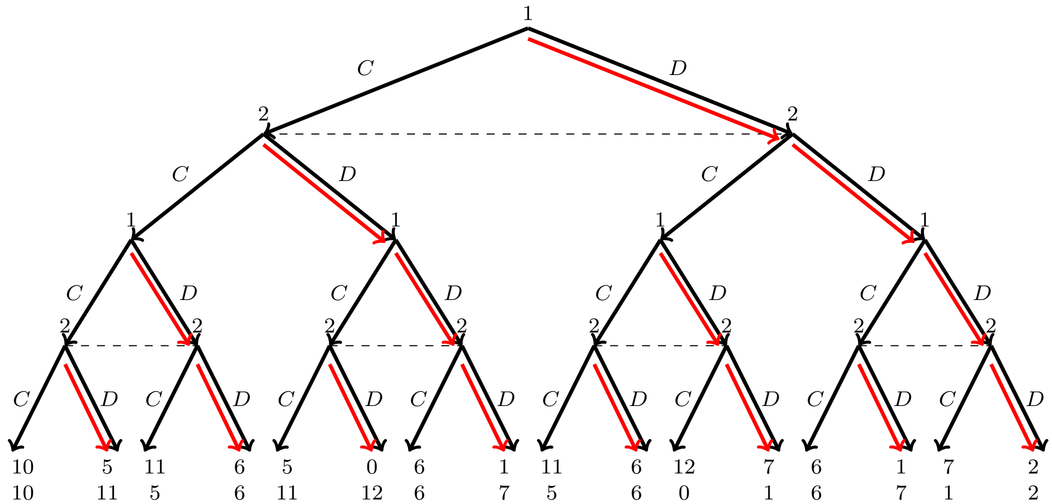
Note que um jogador simplesmente se preocupa com a soma de seus pagamentos nos jogos estágio. Mais importante, no início de cada repetição, cada jogador lembra o que cada jogador jogou em cada jogada anterior. Uma estratégia então prescreve o que o jogador joga em cada t como uma função das jogadas nas datas $0, 1, \dots, t-1$. Mais precisamente, vamos chamar os resultados dos jogos da fase anterior de uma história, que será uma sequência $(a_0, \dots, a_{t-1})$ em cada período t.

Por exemplo, considere uma situação em que dois jogadores jogam o jogo Dilema dos Prisioneiros,

		Jogador 2	
		C	D
Jogador 1	C	$5, 5$	$0, 6$
	D	$6, 0$	$1, 1$

duas vezes. Nesse caso, $T = \{0, 1\}$ e G é o jogo do Dilema dos Prisioneiros. O jogo repetido, G^T, pode ser representado na forma extensiva como

Figura 6.24 – Dilema dos prisioneiros repetido duas vezes



Agora na data $t = 1$, uma história é um perfil de estratégia do jogo Dilema dos Prisioneiros, indicando o que foi jogado em $t = 0$. Existem quatro histórias em $t = 1$: $(C, C), (C, D), (D, C), (D, D)$. Uma estratégia é descrever o que o jogador joga em $t = 0$ e o que ele joga em cada uma dessas quatro histórias. Isso é bastante claro no jogo na foma extensiva.

Vamos calcular o equilíbrio subjogo perfeito de G^T. O jogo tem quatro sub-jogos apropriados, cada um correspondendo ao jogo da última rodada após uma história de jogadas na rodada inicial. Por exemplo, depois de (C, C) na rodada inicial, temos o subjogo onde adicionamos 5 aos payoffs de cada jogador, correspondendo ao pagamento que ele recebe ao jogar (C, C) na primeira rodada. Lembre-se que adicionar uma constante ao payoff de um jogador não altera as preferências em um jogo e, portanto, o conjunto de equilíbrios neste jogo é o mesmo que o jogo original do Dilema dos Prisioneiros, que possui o equilíbrio único de Nash (D, D). Este equilíbrio é representado na figura. Da mesma forma, em cada subjogo adequado, adicionamos alguma constante ao payoff dos jogadores e, portanto, temos (D, D) como o único equilíbrio de Nash em cada um desses subjogos.

Portanto, as ações na última rodada são independentes do que é jogado na rodada inicial. Assim, os jogadores irão ignorar o futuro e jogar o jogo como se não houvesse jogo futuro, cada um jogando D. De fato, dado o comportamento na última rodada, o jogo na rodada inicial se reduz para

		Jogador 2	
		C	D
Jogador 1	C	$6, 6$	$1, 7$
	D	$7, 1$	$2, 2$

em que adicionamos 1 ao payoff de cada jogador, respondendo por sua recompensa na última rodada. O único equilíbrio deste jogo reduzido é (D, D). Isto leva a um equilíbrio perfeito em subjogo: em cada história, cada jogador joga D.

O que aconteceria com um n arbitrário? A resposta continua a mesma. No último dia, n, independente do que foi jogado nas rodadas anteriores, existe um equilíbrio único de Nash para o subjogo resultante: cada jogador joga D. Assim, as ações no dia $n-1$ não têm nenhum efeito no que será jogado no dia seguinte. Então, podemos considerar o sub-jogo como um jogo separado do Dilema dos Prisioneiros. De fato, o jogo reduzido para qualquer subjogo começando em $n-1$ é

		Jogador 2	
		C	D
Jogador 1	C	$5+1+\pi_1, 5+1+\pi_2$	$0+1+\pi_1, 6+1+\pi_2$
	D	$6+1+\pi_1, 0+1+\pi_2$	$1+1+\pi_1, 1+1+\pi_2$

em que π_1 é a soma dos payoffs de i das jogadas anteriores nas datas $0, 1, \ldots, n-2$. Aqui adicionamos π_1 para esses payoffs e 1 para o payoff da última rodada, todos os quais são independentes do que acontece na data$n-1$. Esta é outra versão do Dilema dos Prisioneiros, que tem como único equilíbrio de Nash (D, D). Procedendo deste modo todo o caminho até à data 0, descobrimos que existe um equilíbrio único no sub-jogo perfeito: em cada t e para cada histórico de jogadas anteriores, cada jogador joga D.

Ou seja, embora existam muitas repetições no jogo e as apostas no futuro possam ser altas, qualquer plano de ação que não seja jogar D de forma míope por toda parte se desfaz, já que os jogadores não podem se comprometer com qualquer plano de ação na última rodada. Este é realmente um resultado geral.

Teorema 6.7.1. *Seja n finito e assuma que G tem um único equilíbrio perfeito em subjogo, s^*. Então, G^T tem um único equilíbrio perfeito em subjogo, e de acordo com este equilíbrio s^* é jogado em cada data independente da história das jogadas anteriores.*

6.7.2 Jogos Repetidos Infinitamente com Ações Observáveis

Agora, considere os jogos repetidos infinitamente em que todos os movimentos anteriores são de conhecimento comum no início de cada estágio. Isto é, tome $T = \{0, 1, 2, \ldots\}$ como o conjunto de números naturais em vez de $T = \{0, 1, 2, \ldots, n\}$. O jogo continua indefinidamente, independentemente do que os jogadores joguem ao longo do caminho.

6.7.2.1 Valor Esperado

Em um jogo infinitamente repetido, não se pode simplesmente adicionar os payoffs de cada estágio, pois a soma se torna infinita. Para esses jogos, suponha que os jogadores maximizem a soma descontada dos pagamentos dos jogos. O valor presente de qualquer fluxo de payoff $\pi = (\pi_0, \pi_1, \ldots, \pi_t, \ldots)$ é calculado por

$$VP(\pi; \delta) = \sum_{t=0}^{\infty} \delta^t \pi_t = \pi_0 + \delta\pi_1 + \ldots + \delta^t \pi_t + \ldots, \tag{6.73}$$

em que $\delta \in (0,1)$ é o fator de desconto. O valor médio é

$$(1-\delta)VP(\pi;\delta) \equiv (1-\delta)\sum_{t=0}^{\infty}\delta^t\pi_t. \tag{6.74}$$

Os valores presentes e médios podem ser calculados em relação à data atual. Isto é, dado qualquer t, o valor presente em t é

$$VP_t(\pi;\delta) = \sum_{s=t}^{\infty}\delta^{s-t}\pi_s = \pi_t + \delta\pi t + 1 + \ldots + \delta^k\pi_{t+k} + \ldots, \tag{6.75}$$

e o valor médio em t é $(1-\delta)VP_t(\pi;\delta)$.

6.7.2.2 História e Estratégia

Mais uma vez, num jogo repetido, uma história no início de uma determinada data t é a sequência dos resultados da jogada nas datas $0, 1, \ldots, t-1$. Por exemplo, no jogo Entrar-Resistir, os possíveis resultados do jogo de teatro são X, EO = (Entrar,Observar) e EL = (Entrar,Lutar), e as possíveis histórias são t-uplas desses três resultados para cada t. Exemplos de histórias são

$$XXEOXEL$$
$$EOXELXX$$

Uma história no início da data t é denotada por $h = (a_0, \ldots, a_{t-1})$, em que $a_{t'}$ é o resultado do jogo de fase na rodada t'; h é vazio quando $t = 0$. Por exemplo, no jogo dilema dos prisioneiros, $((C,C),(C,D))$ é uma história para $t = 2$.

Uma estratégia em um jogo repetido, mais uma vez, determina uma estratégia no jogo de estágio para cada história e para cada t. O ponto importante é que a estratégia no jogo em uma determinada data pode variar de acordo com as histórias. Aqui estão algumas estratégias possíveis no jogo repetido do Dilema do Prisioneiros:

- Estratégia grim (resistência, intransigência): jogar C em $t = 0$; em seguida, jogar C se os jogadores sempre jogaram (C,C) no passado, jogar D caso contrário (ou seja, se alguém já jogou D no passado).
- Cooperar: sempre jogar C (não importa o que aconteceu no passado).
- Tit-for-Tat: jogar C em $t = 0$, e em cada $t > 0$, jogar o contrário do que o outro jogou em $t-1$.

Note que as estratégias (Grim,Grim), (Cooperar,Cooperar) e (Tit-for-Tat,Tit-for-Tat) levam a mesma trajetória:

$$((C, C), (C, C), \ldots). \tag{6.76}$$

No entanto, eles são perfis estratégicos bastante distintos. De fato, (Cooperar,Cooperar) não é nem um equilíbrio de Nash, enquanto (Grim,Grim) é um equilíbrio de Nash perfeito para subjogos para grandes valores de δ. Por outro lado, enquanto (Tit-for-Tat,Tit-for-Tat) é um equilíbrio de Nash para valores grandes de δ, não é perfeito para subjogos.

6.7.2.3 Princípio do Desvio ùnico

Em um jogo infinitamente repetido, usa-se o princípio de desvio único para verificar se um perfil de estratégia é um equilíbrio de Nash perfeito para subjogos. Em tal jogo, o princípio assume uma forma simples e é aplicado através de jogos-estágio aumentados. Aqui, aumentado refere-se ao fato de que simplesmente aumentamos os payoffs adicionando o valor presente dos payoffs futuros sob o equilíbrio pretendido. Pode-se também usar o termo jogo reduzido em vez do jogo de aumentado, alternadamente.

Formalmente considere um perfil de estratégia $s^* = (s_1^*, s_2^*, \ldots, s_n^*)$ no jogo repetido. Considere qualquer período t e qualquer história $h = (a_0, \ldots, a_{t-1})$, em que $a_{t'}$ é o resultado da jogada em t'. O jogo aumentado para s^* e h é o mesmo jogo que

$$U_i(z|s^*, h) = u_i(z) + \delta V P_{i,t+1}(h, z, s^*), \tag{6.77}$$

em que $u_i(z)$ é é o payoff do jogador i em z no jogo original e $V P_{i,t+1}(h, z, s^*)$ é o valor presente de i em $t + 1$ do fluco de payoffs que resulta quando todos os jogadores seguem s^* iniciando com a história $(h, z) = (a_, 0, \ldots, a_{t-1}, z)$, que é a história no começo no período $t + 1$.

O Princípio do Desvio Único agora afirma que um perfil de estratégia no jogo repetido é subjogo-perfeito se ele sempre produzir um equilíbrio de Nash perfeito no subjogo do jogo estágio aumentado:

Teorema 6.7.2. *O perfil de estratégia s^* é um equilíbrio de Nash perfeito para o subjogo do jogo repetido se e somente se $(s_1^*(h), \ldots, s_n^*(h))$ é um equilíbrio de Nash perfeito no subjogo do jogo de estágio aumentado para s^* e h para cada período t e toda história no início de t.*

Vamos ver o exemplo do Dilema dos Prisioneiros repetido infinitamente.

Quando o jogo estágio é um jogo de ação simultânea, não há distinção entre o equilíbrio de Nash perfeito no subjogo e o equilíbrio de Nash. Assim, para o teste de desvio único, basta verificar se $s^*(h)$ é um equilíbrio de Nash do jogo de estágio aumentado para h para cada história h. Isso simplifica a análise substancialmente porque só é preciso calcular os payoffs sem desvio e com desvios unilaterais para verificar se o perfil da estratégia é um equilíbrio de Nash.

Como exemplo, considere o perfil da estratégia (Grim,Grim). Existem dois tipos de histórias que precisamos considerar separadamente para esse perfil de estratégia.

1. Cooperação: histórias em que D nunca foi jogado por nenhum jogador.

2. Deserção: histórias em que D foi jogado por alguém em alguma data.

Primeiro considere uma história de cooperação para alguns t. Agora, se ambos os jogadores jogarem C, então, de acordo com (Grim,Grim), a partir de $t+1$ cada jogador irá jogar C para sempre. Isso produz o valor presente

$$V_C = 5 + 5\delta + 5\delta^2 + \ldots = \frac{5}{1-\delta}, \tag{6.78}$$

em $t+1$.

Se algum jogador jogar D, então de $t+1$ em diante, todas as histórias serão de Deserção e cada um jogará D para sempre. Isso produz o valor presente de

$$V_D = 1 + 1\delta + 1\delta^2 + \ldots = \frac{1}{1-\delta}, \tag{6.79}$$

em $t+1$.

Agora, em t, se ambos jogarem C, o pagamento de cada jogador será $5+\delta V_C$. Se o jogador 1 jogar D enquanto o jogador 2 estiver jogando C, então o jogador 1 recebe $6+\delta V_D$ e o jogador 2 recebe $0+\delta V_D$. Assim, o jogo aumentado é:

		Jogador 2: C	Jogador 2: D
Jogador 1	C	$5+\delta V_C, 5+\delta V_C$	$0+\delta V_D, 6+\delta V_D$
	D	$6+\delta V_D, 0+\delta V_D$	$1+\delta V_D, 1+\delta V_D$

Para passar no teste, (C,C) deve ser um equilíbrio de Nash desse jogo. Este é o caso se e somente se

$$\begin{aligned} 5+\delta V_C &\geq 6+\delta V_D \\ 5+\delta\left(\frac{5}{1-\delta}\right) &\geq 6+\delta\left(\frac{1}{1-\delta}\right) \\ \delta &\geq \frac{1}{5}. \end{aligned} \tag{6.80}$$

Também precisamos considerar histórias de deserção. Considere uma história de cooperação para alguns t. Agora, independente do que é jogado em t, de acordo com a estratégia (Grim,Grim), de $t+1$ em diante teremos histórias de deserção e cada jogador jogará D para sempre. O valor presente dos payoffs de $t+1$ em diante será V_D. Então, estágio jogo aumentado nesta história é

		Jogador 2	
		C	D
Jogador 1	C	$5+\delta V_D, 5+\delta V_D$	$0+\delta V_D, 6+\delta V_D$
	D	$6+\delta V_D, 0+\delta V_D$	$1+\delta V_D, 1+\delta V_D$

O teste de desvio único para (Grim,Grim) requer que (D, D) seja um equilíbrio de Nash desse jogo, e de fato (D, D) é o único equilíbrio de Nash. Como (Grim,Grim) passa no teste de desvio único em cada história, é um equilíbrio de Nash perfeito em subjogogo quando $\delta \geq \frac{1}{5}$ (não é um equilíbrio de Nash perfeito em subjogogo quando $\delta < \frac{1}{5}$).

Agora usaremos a mesma técnica para mostrar que (Tit-for-tat, Tit-por-tat) não é um equilíbrio de Nash perfeito no subjogo (exceto no caso degenerado $\delta = \frac{1}{5}$). As estratégias tit-for-tat em $t+1$ dependem apenas do que é jogado em t e não de qualquer jogada anterior. Se (C, C) é jogado em t, então começando em $t+1$ teremos (C, C) para sempre e, portanto, o vetor de valores presentes em $t+1$ será

$$\left(\frac{5}{1-\delta}, \frac{5}{1-\delta}\right) = (5,5) + \delta(5,5) + \delta^2(5,5) + \dots. \tag{6.81}$$

Se (C, D) for jogada em t, então de acordo com a estratégia tit-for-tat a sequência de jogadas iniciadas em $t+1$ serão

$$(D, C), (C, D), (D, C), (C, D), \dots,$$

e, portanto, o vetor de valores presentes em $t+1$ será

$$\left(\frac{6}{1-\delta^2}, \frac{6\delta}{1-\delta^2}\right) = (6,0) + \delta(0,6) + \delta^2(6,0) + \dots. \tag{6.82}$$

Similarmente, se (D, C) é jogado em t, então o valor presente em $t+1$ será

$$\left(\frac{6\delta}{1-\delta^2}, \frac{6}{1-\delta^2}\right). \tag{6.83}$$

Finalmente, se jogarem (D, D) em t, teremos (D, D) para sempre, resultando em $t+1$ o seguinte valor presente: $\left(\frac{1}{1-\delta}, \frac{1}{1-\delta}\right)$.

Para mostrar que (Tit-for-tat, Tit-por-tat) não é um equilíbrio de Nash perfeito para subjogos, consideraremos duas histórias. Para mostrar que um perfil de estratégia não é um equilíbrio perfeito perfeito para o subjogo, basta encontrar um caso em que ele falha no princípio doe desvio único. O jogo reduzido em qualquer t para qualquer história anterior é

		Jogador 2	
		C	D
Jogador 1	C	$5+\delta\frac{5}{1-\delta}, 5+\delta\frac{5}{1-\delta}$	$0+\delta\frac{6}{1-\delta^2}, 6+\delta\frac{6\delta}{1-\delta^2}$
	D	$6+\delta\frac{6\delta}{1-\delta^2}, 0+\delta\frac{6}{1-\delta^2}$	$1+\frac{\delta}{1-\delta}, 1+\frac{\delta}{1-\delta}$

1. Considere $t=0$, quando a estratégia (Tit-for-tat, Tit-por-tat) prescreve (C, C). O princípio do desvio único então requer que (C, C) seja um equilíbrio de Nash do jogo reduzido acima. Ou seja, devemos ter

$$5+\delta\frac{5}{1-\delta} \geq 6+\delta\frac{6\delta}{1-\delta^2}. \tag{6.84}$$

2. Considere uma história em que (C, D) é jogado em $t-1$. Agora, de acordo com a estratégia (Tit-for-tat, Tit-por-tat), devemos ter (D, C) em t. O princípio do desvio-único agora requer que (D, C) seja um equilíbrio de Nash do jogo acima. Ou seja, devemos ter

$$6+\delta\frac{6\delta}{1-\delta^2} \geq 5+\delta\frac{5}{1-\delta}, \tag{6.85}$$

o que é uma contradição com o suposto acima.

Portanto, a estratégia (Tit-for-tat, Tit-por-tat) não é um equilíbrio de Nash perfeito em subjogos, a menos que $6+\delta\frac{6\delta}{1-\delta^2} = 5+\delta\frac{5}{1-\delta}$, ou, equivalentemente, $\delta = \frac{1}{5}$.

6.7.3 Teorema Folk

Um objetivo principal do estudo de jogos repetidos é explorar a relação entre os incentivos a curto prazo (dentro de um único período) e os incentivos de longo prazo (dentro do jogo repetido mais amplo). A sabedoria convencional na teoria dos jogos sugere que quando os jogadores são pacientes, seus incentivos de longo prazo assumem o controle, e um grande conjunto de comportamentos pode resultar em equilíbrio. De fato, para qualquer vetor de payoff viável e individualmente racional e para valores suficientemente grandes de δ, existe algum equilíbrio perfeito no subjogo que produz o vetor payoff como o valor médio do fluxo de payoff. Este fato é chamado de Teorema Popular (Teorema Folk). Esta seção é dedicada a apresentar uma versão básica do teorema popular e ilustrar sua prova.

Ao longo desta seção, assume-se que o jogo é um jogo de ação simultânea (N, A, u) em que $N = \{1, 2, \ldots, n\}$ é o conjunto de jogadores, $A = A_1 \times A_2 \times \ldots A_n$ é o conjunto finito de perfil de estratégias, e $u_i \colon A \to \mathbb{R}$ é a função payoff.

6.7.3.1 Payoffs Viáveis

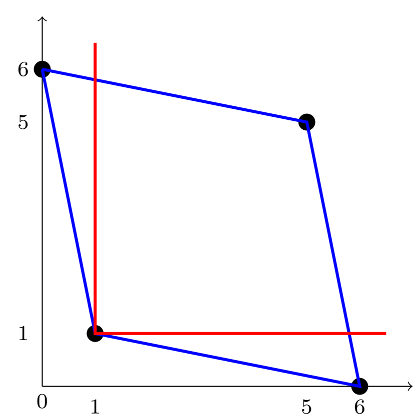
Imagine que os jogadores aleatoriamente coletam os perfis de estratégia do jogo $a \in A$. Quais vetores de payoff poderiam obter se pudessem escolher qualquer distribuição de probabilidade $p\colon A \to [0,1]$ sobre A? (Lembre-se que $\sum_{a\in A} p(a) = 1$). A respota é o conjunto V de vetores de payoffs $v = (v_1, \ldots, v_n)$ tal que:

$$v = \sum_{a\in A} p(a)(u_1(a), \ldots, u_n(a)), \tag{6.86}$$

para alguma distribuição de probabilidade $p\colon A \to [0,1]$ sobre A. Note que V é o menor conjunto convexo que contém todos os vetores de payoff $(u_1(a), \ldots, u_n(a))$ do perfil de estratégias puras no estágio jogo. Diz-se que um vetor de payoff é viável se $v \in V$. Ao longo desta seção, V é assumido como sendo n-dimensional.

Para uma ilustração visual, considere o jogo Dilema dos Prisioneiros. O conjunto V é plotado na Figura 6.25. Como existem dois jogadores, V contém os pares $V = (v_1, v_2)$. Os vetores de payoff de estratégias puras são $(1,1)$, $(5,5)$, $(6,0)$ e $(0,6)$. O conjunto V é a área em forma de diamante que fica entre as linhas que conectam esses quatro pontos.

Figura 6.25 – Payoffs viáveis no dilema dos prisioneiros



Note que para cada perfil de estratégia s no jogo repetido, o vetor de payoff médio de s está em V. Isso também implica que o mesmo é verdadeiro para perfis de estratégia mista no jogo repetido. Inversamente, se os jogadores puderem aleatoriamente coletar em perfis estratégicos nos jogos repetidos, todos os vetores $v \in V$ poderiam ser obtidos como vetores de payoff médios.

6.7.3.2 Racionalidade Individual

Há um limite inferior em quanto um jogador fica em equilíbrio. Por exemplo, no dilema dos prisioneiros (repetido), se continuarmos a jogar deserção todos os dias, não importa o que aconteça, ele recebe pelo menos 1 por dia, ganhando uma recompensa média de 1 ou mais. Então, ele deve obter pelo menos 1 em qualquer equilíbrio de Nash, porque poderia desviar-se lucrativamente da estratégia acima.

Para encontrar um limite inferior nos payoffs dos equilíbrios de Nash de estratégia pura, para cada jogador eu defino o payoff minmax de estratégia pura como

$$m_i = \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a_i, a_{-i}). \tag{6.87}$$

Aqui, os outros jogadores tentam minimizar a recompensa do jogador i escolhendo uma estratégia pura s_{-i} para eles mesmos, sabendo que o jogador i jogará uma melhor resposta para a_{-i}. Então, o castigo mais severo que eles podem infligir sobre i é m_i. Por exemplo, no jogo do dilema dos prisioneiros, $m_i = 1$ porque i obtém o máximo de 6 se o outro jogador jogar C e recebe no máximo 1 se o outro jogador jogar D.

Observe que em qualquer equilíbrio de Nash de estratégia pura s^* do jogo repetido, a recompensa média do jogador i é pelo menos m_i. Para ver isso, suponha que o payoff médio de i seja menor que m_i em s^*. Agora considere a estratégia $\widehat{s}_i$, tal que para cada história h, $\widehat{s}_i(h)$ é a melhor resposta do estágio jogo a $\widehat{s}_{-i}(h)$, ou seja,

$$u_i\left(\widehat{s}_i(h), s^*_{-i}(h)\right) = \max_{a_i \in A_i} u_i\left(a_i, s^*_{-i}(h)\right). \tag{6.88}$$

Dado que

$$\max_{a_i \in A_i} u_i\left(a_i, s^*_{-i}(h)\right) \geq m_i \tag{6.89}$$

para todo h, isto implica que o payoff médio de $(\widehat{s}_i, s^*_{-i})$ é no mínimo 1, dando ao jogador i um incentivo para se desviar.

Um limite inferior para o payoff médio de um equilíbrio de Nash em estratégia mista é dado pelo payoff minmax, definido como

$$\mu_i = \min_{\alpha_j, j \neq i} \max_{a_i \in A_i} \sum_{a_{-i} \in A_{-i}} \prod_{j \neq i} \alpha_j(a_j) u_i(a_i, a_{-i}), \tag{6.90}$$

em que α_j é uma estratégia mista do jogador j no estágio jogo. Similarmente às estratégias puras, pode-se mostrar que o payoff médio do jogador i é pelo menos μ_i em qualquer equilíbrio de Nash (misto ou puro). Note que, por definição, $\mu_i \leq m_i$. A igualdade pode ser estrita. Por exemplo, no jogo da moeda,

		Jogador 2	
		Cara	Coroa
Jogador 1	Cara	$-1, 1$	$1, -1$
	Coroa	$1, -1$	$-1, 1$

o payoff minmax da estratégia pura m_i é 1, enquanto o payoff minmax da estratégia mista μ_i é 0. Isto é obtido quando $\alpha_j(cara) = \alpha_j(coroa) = \frac{1}{2}$. Por uma questão de exposição, assume-se que $(\mu_1, \ldots, \mu_n) \in V$.

Um vetor de payoffs v é considerado individualmente racional se e somente se $v_i \geq \mu_i$ para todo $i \in N$.

Teorema 6.7.3. *Seja sv $\in V$ tal que $v_i > \mu_i$ para cada jogador I. Então, existe $\overline{\delta} \in (0, 1)$ tal que para todo $\delta > \overline{\delta}$ existe um equilíbrio perfeito no subjogo do jogo repetido sob o qual o valor médio de cada jogador i é v_i. Além disso, se $v_i > m_i$ para todo i, então o equilíbrio perfeito para o subjogo é em estratégias puras.*

O teorema afirma que qualquer vetor de payoff estritamente individualmente racional e viável pode ser apoiado no equilíbrio perfeito de Nash no subjogo quando os jogadores são suficientemente pacientes. Como todos os vetores de payoff equilíbrio precisam ser individualmente racionais e viáveis, o teorema fornece uma caracterização aproximada dos vetores de payoff de equilíbrio quando os jogadores são pacientes: o conjunto de todos os vetores de payoffs viáveis e individualmente racionais.

6.7.4 Aplicação: Cournot Repetido Infinitamente

Usarei o modelo de oligopólio linear de Cournot infinitamente repetido como o principal status de um cartel. Existem empresas, cada uma com custo marginal $c \in (0, 1)$. No estágio jogo cada empresa i produz simultaneamente q_i unidades de um bem e vende-o a preço

$$P = \max\{1 - Q, 0\}, \tag{6.91}$$

em que $Q = q_1 + \ldots + q_n$ é a oferta total. No jogo repetido, todos os níveis de produção passados de todas as empresas são publicamente observáveis, e a função de utilidade de cada empresa é a soma descontada de seus lucros em cada estágio, onde o fator de desconto é δ:

$$u_i = \sum_{t=0}^{\infty} \delta^t q_{i,t}(P(q_{1,t} + \ldots + q_{n,t}) - c), \tag{6.92}$$

em que $q_{j,t}$ é o nível de produção da da firma j no período t. Às vezes, será mais conveniente usar o valor médio descontado, que é $(1 - \delta)u_i$.

Para um q qualquer escreva,

$$f(q) = q(P(qn) - c) = q(\max\{1 - nq, 0\} - c) \tag{6.93}$$

para o lucro de uma firma quando cada firma produz q e

$$g(q) = \max_{q'} q(P(q' + (n-1)q) - c) = \begin{cases} \dfrac{(1-(n-1)q-c)^2}{4} & \text{se } L \ (n-1)q \leq 1 \\ 0 & \text{se } L \text{ caso contrário} \end{cases} \tag{6.94}$$

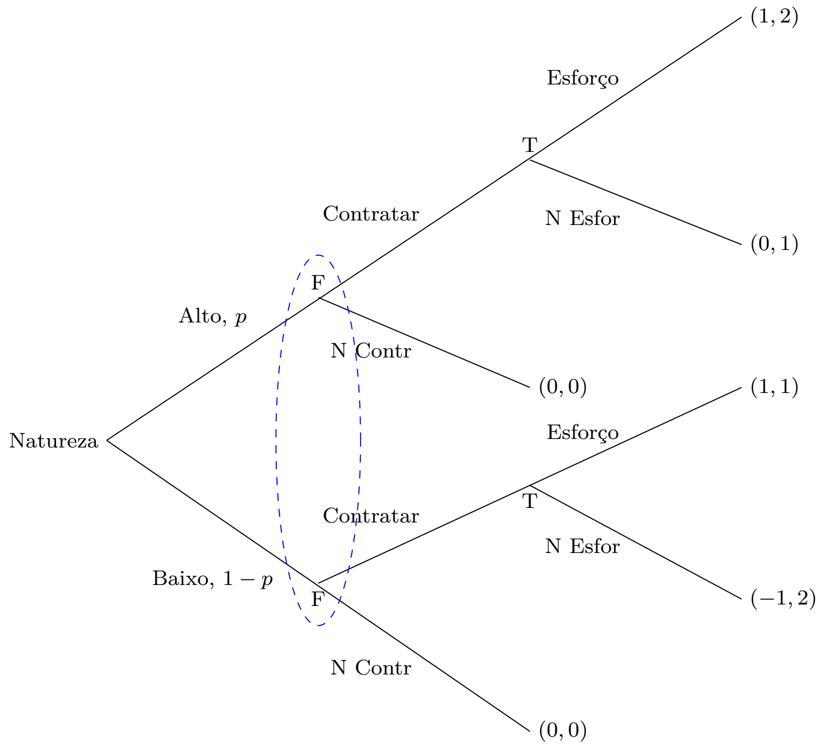
para o lucro máximo de uma firma a partir da melhor resposta quando todas as outras firmas produzem q.

6.8 Jogos Estáticos com Informação Incompleta

Até agora nos concentramos em jogos em que qualquer informação que é conhecida por qualquer jogador é conhecida por todos os jogadores (e, na verdade, pelo conhecimento comum). Tais jogos são chamados os jogos de informação completa. As preocupações informacionais não desempenham nenhum papel em tais jogos. Na vida real, os jogadores sempre têm alguma informação privada que não é conhecida por outras partes. Por exemplo, dificilmente podemos conhecer as preferências e crenças de outros jogadores tão bem quanto eles. As preocupações informacionais desempenham um papel central na tomada de decisão dos jogadores nesses ambientes estratégicos. A partir de agora nos concentraremos em tais questões informacionais. Vamos considerar casos em que uma parte pode ter alguma informação que não é conhecida por outra parte. Tais jogos são chamados jogos de informação incompleta ou informação assimétrica. As assimetrias informacionais são modeladas pelos movimentos da Natureza. Alguns jogadores podem distinguir certos movimentos da natureza enquanto outros não. Considere o seguinte exemplo simples, em que uma empresa está contemplando a contratação de um trabalhador, sem saber qual a capacidade do trabalhador.

Considere o jogo na Figura abaixo. Há uma empresa e um trabalhador.O trabalhador pode ser de alta habilidade, caso em que ele gostaria de trabalhar quando ele é contratado, ou de baixa habilidade, caso em que ele preferiria fazer o mínimo de esforço. A empresa gostaria de contratar o trabalhador que vai trabalhar, mas não o trabalhador que vai fazer o mínimo de esforço. O trabalhador sabe o seu nível de habilidade. A empresa não sabe se o trabalhador é de alta capacidade ou baixa capacidade. A empresa acredita que o trabalhador é de alta capacidade com probabilidade p e de baixa habilidade com probabilidade $1-p$. Mais importante, a empresa sabe que o trabalhador conhece seu próprio nível de habilidade. Para modelar essa situação, deixamos a Natureza escolher entre Alta e Baixa, com probabilidades p e $1-p$, respectivamente. Então deixamos o trabalhador observar a escolha da Natureza, mas não permitimos que a empresa observe a escolha da natureza.

Figura 6.26 – Decisão de empregabilidade com informação incompleta



As informações privadas de um jogador são chamadas de "tipo". Por exemplo, no exemplo acima, o trabalhador (T) possui dois tipos: alto e baixo. Como a firma (F) não possui informações privadas, a firma possui apenas um tipo. Como no exemplo acima, informações incompletas são modeladas por meio de jogos de informações imperfeitos, nos quais a natureza escolhe o tipo de cada jogador e informa a ele em particular. Estes jogos são chamados de jogo de informação incompleta ou jogo Bayesiano.

6.8.1 Jogos Bayesianos

Formalmente, um jogo estático de informação incompleta é como segue. Primeiro, a natureza escolhe um tipo t, $t = (t_1, t_2, \ldots, t_n) \in T$, em que cada $t \in T$ é selecionado com probabilidade $p(t)$. Aqui $t_i \in T_i$ é o tipo do jogador $i \in N = \{1, 2, \ldots, n\}$. Então, cada jogador sabe seu próprio tipo, mas não o dos outros. Finalmente, os jogadores escolhem simultaneamente suas ações. Denote por $a = (a_1, a_2, \ldots, a_n) \in A$ a lista das ações tomadas por todos os agentes, em que $a_\in A_i$ é a ação tomada pelo jogador i. O payoff de um jogador dependerá dos tipos e das ações dos outros jogadores e é escrito como $u_i\colon A \times T \to \mathbb{R}$. Tal jogo estático com informação incompleta é denotado por (N, T, A, p, u).

No exemplo acima, teríamos:

- $N = \{F, T\}$.
- $T_F = \{t_F\}$, $T_T = \{Alto, Baixo\}$.

- $p(t_F, alto = p)$ e $p(t_F, baixo) = 1 - p$.
- $A_F = \{$contratar, não contratar$\}$ e $A_T = \{$Esforço, Não esforço$\}$.
- e as funções de utilidade u_F e u_T são definidas pelas seguintes tabelas, onde a primeira corresponde a $t_T = alto$ e $t_T = baixo$, respectivamente:

		Trabalhador	
		Esforço	Não faz esforço
Firma	Contrata	$1, 2$	$0, 1$
	Não contrata	$0, 0$	$0, 0$

		Trabalhador	
		Esforço	Não faz esforço
Firma	Contrata	$1, 1$	$-1, 2$
	Não contrata	$0, 0$	$0, 0$

É muito importante notar que os tipos de jogadores podem ser "correlacionados", o que significa que um jogador "atualiza" suas crenças sobre o tipo do outro jogador quando ele aprende seu próprio tipo. Uma vez que ele conhece seu tipo quando ele toma uma ação, ele maximiza sua utilidade esperada com relação às novas crenças depois de atualizá-las. Assumimos que ele atualiza suas crenças usando a Regra de Bayes.

Seja A e B dois eventos, então a probabilidade de que A ocorra condicional a B ocorrer é

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{6.95}$$

em que $P(A \cap B)$ é a probabilidade de A e B ocorrerem simultaneamnete.

Seja $p_i(t'_{-i}|t_i)$ a crença do jogador i de que os tipos de todos os outros jogadores são $t'_{-i} = (t'_1, t'_2, \ldots, t'_{i-1}, t'_{i+1}, \ldots, t'_n)$ dado que seu tipo é t'_i.

Por exemplo, para um jogo bayesiano de dois jogadores, seja $T_1 = T_2 = \{H, L\}$ e $p(H, H) = p(H, L) = p(L, L) = \dfrac{1}{3}$ e $p(L, H) = 0$. A distribuição é tabulada como

	H	L
H	$\frac{1}{3}$	$\frac{1}{3}$
L	0	$\frac{1}{3}$

Por exemplo,

$$p_1(H|H) = \frac{p(H,H)}{p(H,H)+p(H,L)} = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{3}} = \frac{1}{2} \tag{6.96}$$

$$p_1(H|L) = \frac{p(L,H)}{p(L,H)+p(L,L)} = \frac{0}{0+\frac{1}{3}} = 0 \tag{6.97}$$

6.8.2 Equilíbrio de Nash Bayesiano

Como de costume, uma estratégia de um jogador determina qual ação ele tomará em cada conjunto de informação. Aqui, os conjuntos de informações são identificados com os tipos $t_i \in T_i$. Portanto, uma estratégia de um jogador é uma função

$$s_i\colon T_i \to A_i, \tag{6.98}$$

mapeando seus tipos para suas ações. No exemplo acima, o trabalhador tem quatro estratégias: (Esforço, Esforço)) - querendo dizer que ele irá trabalhar independentemente de ser de alta ou baixa habilidade, (Esforço, Não faz esforço) significando que ele irá trabalhar se for de alta habilidade e não trabalhar se ele é de baixa capacidade, (Não faz esforço, Esforço), e (Não faz esforço, Não faz esforço).

Quando a probabilidade de cada tipo é positiva de acordo com p, qualquer equilíbrio de Nash de um jogo Bayesiano é chamado de equilíbrio de Nash Bayesiano. Nesse caso, em um equilíbrio de Nash, para cada tipo t_i, o jogador i tem uma melhor resposta às estratégias dos outros, considerando suas crenças sobre os tipos de outros jogadores dado t_i. Se a probabilidade da natureza escolher algum t_i é zero, então qualquer ação nesse tipo é possível de acordo com um equilíbrio (como sua ação nesse tipo não afeta seu resultado esperado.) Em um equilíbrio de Nash Bayesiano, assumimos que para cada tipo t_i, o jogador i tem uma melhor resposta às estratégias dos outros, considerando suas crenças sobre os tipos de outros jogadores, t_i, independentemente de a probabilidade desse tipo ser positiva.

Formalmente, um perfil de estratégias $s^* = (s_1^*, \ldots, s_n^*)$ é um equilíbrio de Nash Bayesiano em um estágio jogo de informação incompleta com n pessoas se e somente se para cada jogador i e tipo $t_i \in T_i$

$$s_i^*(t_i^1) \in \arg\max_{a_i} \sum u_i\left(s_i^*(t_i), \ldots, a_i, \ldots, s_N^*(t_N)\right) p_i(t'_{-i}|t_i), \tag{6.99}$$

em que u_i é a utilidade do jogador i e denota a ação a_i. Ou seja, para cada jogador i e cada tipo possível, a ação escolhida é ótima, dadas as crenças condicionais desse tipo contra as estratégias ótimas de todos os outros jogadores. Observe que a função de utilidade u_i do jogador i depende das ações e dos tipos de ambos os jogadores. Observe também que um equilíbrio de Nash Bayesiano é um equilíbrio de Nash de um jogo bayesiano com a propriedade adicional que cada tipo tem uma melhor resposta. Por exemplo, para $p = \frac{3}{4}$, considere o equilíbrio de Nash do jogo entre a firma e o trabalhador no qual a firma contrata e trabalhador trabalha, se e somente se a Natureza escolher alta habilidade. Podemos formalmente escrever este perfil de estratégia como $s^* = (s_F^*, s_T^*)$ com

$$s_F^*(t_F) = \text{contratar} \tag{6.100}$$
$$s_T^*(alto) = \text{esforço} \tag{6.101}$$
$$s_T^*(baixo) = \text{não esforço} \tag{6.102}$$

Verificamos que este é um equilíbrio Bayesiano de Nash como segue. Primeiro considere a empresa.

Em seu único tipo t_F, suas crenças sobre os outros tipos são

$$p_F(alto|t_F) = \frac{3}{4} \tag{6.103}$$
$$p_F(baixo|t_F) = \frac{1}{4} \tag{6.104}$$

A utilidade esperada da ação “contratar” é:

$$\begin{aligned}\mathbb{E}[u_F(\text{contratar}, s_T^*)|t_F] &= u_F(\text{contratar}, s_T^* \text{ alto}, \text{alto})p_F(alto|t_F)+\\ &u_F(\text{contratar}, s_T^* \text{ baixo}, \text{baixo})p_F(baixo|t_F)\\ &= u_F(\text{contratar}, \text{esforço}, \text{alto})p_F(alto|t_F)+\\ &u_F(\text{contratar}, \text{não esforço}, \text{baixo})p_F(baixo|t_F)\\ &= 1 \cdot \frac{3}{4} + (-1) \cdot \frac{1}{4} = \frac{1}{2}\end{aligned} \tag{6.105}$$

A utilidade esperada da ação “não contratar” é:

$$\begin{aligned}\mathbb{E}[u_F(\text{não contratar}, s_T^*)|t_F] &= u_F(\text{não contratar}, s_T^* \text{ alto}, \text{alto})p_F(alto|t_F)+\\ &u_F(\text{não contratar}, s_T^* \text{ baixo}, \text{baixo})p_F(baixo|t_F)\\ &= u_F(\text{não contratar}, \text{esforço}, \text{alto})p_F(alto|t_F)+\\ &u_F(\text{não contratar}, \text{não esforço}, \text{baixo})p_F(baixo|t_F)\\ &= 0 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = 0\end{aligned} \tag{6.106}$$

Dado que $\mathbb{E}[u_F(\text{contratar}, s_T^*)|t_F] > \mathbb{E}[u_F(\text{não contratar}, s_T^*)|t_F]$, “contratar” é a melho resposta.

Agora, precisamos considerar o tranalhador. Ele tem dois tipos. Nós precisamos verificar se ele joga a melhor resposta para cada tipo. Temos que $p_F(t_F|alto) = 1$ e $p_F(t_F|baixo) = 1$. Assim,

$$\mathbb{E}[u_T(s_F^*, \text{esforço})|\text{alto}] = u_T(\text{contratar,esforço,alto}) = 2 \tag{6.107}$$

$$\mathbb{E}[u_T(s_F^*, \text{não esforço})|\text{alto}] = u_T(\text{contratar,não esforço,alto}) = 1 \tag{6.108}$$

$$\mathbb{E}[u_T(s_F^*, \text{não esforço})|\text{baixo}] = u_T(\text{contratar,não esforço,baixo}) = 2 \tag{6.109}$$

$$\mathbb{E}[u_T(s_F^*, \text{esforço})|\text{baixo}] = u_T(\text{contratar,esforço,baixo}) = 1 \tag{6.110}$$

Claramente, $2 > 1$ e "esforço" é a melhor resposta a s_F^* para o tipo alto. De modo semelhante, $2 > 1$ e "não esforço" é a melhor resposta a s_F^* para o tipo baixo.

6.8.3 Aplicação: Cournot com Informação Incompleta

Considere um duopólio (Cournot) com função de demanda inversa

$$P(Q) = a - Q, \tag{6.111}$$

em que $Q = q_1 + q_2$. O custo marginal da firma 1 é $c = 0$ e é de conhecimento comum. O custo marginal da firma 2 é c_2 é uma informação privada e pode assumir os seguintes valores:

$$c_H \text{ com probabilidade } \theta \tag{6.112}$$

$$c_L \text{ com probabilidade } 1 - \theta \tag{6.113}$$

Cada empresa maximiza seu lucro esperado.

Aqui, a firma 1 tem apenas um tipo e a firma 2 tem dois tipos: c_H e c_L. Assim, uma estratégia da firma 1 é um número real q_1, enquanto uma estratégia da firma 2 são dois números reais, $q_2(c_H)$ e $q_2(c_L)$, uma para quando o custo é c_H e uma para quando o custo é c_L.

Um equilíbrio de Nash Bayesiano é um 3-upla $(q_1^*, q_2^*(c_H), q_2^*(c_L))$ de números reais, em que q_1^* é o nível de produção da firma 1, $q_2^*(c_H)$ é o nível de produção do tipo c_H da firma 2, e $q_2^*(c_L)$ é o nível de produção do tipo c_L da firma 2. Em equilíbrio, cada tipo joga uma melhor resposta. Primeiro, considere o tipo de alto custo c_H da empresa 2. Em equilíbrio, cada tipo sabe que a firma 1 produz q_1^*. Assim, seu nível de produção, $q_2^*(c_H)$, resolve o seguinte problema de maximização

$$\max_{q_2} (P - c_H) q_2 = \max_{q_2} [a - q_1^* - q_2 - c_H] q_2. \tag{6.114}$$

Portanto,

$$q_2^*(c_H) = \frac{a - q_1^* - c_H}{2}. \tag{6.115}$$

Agora, considere o tipo de baixo custo c_L da empresa 2. Em equilíbrio, cada tipo sabe que a firma 1 produz q_1^*. Assim, seu nível de produção, $q_2^*(c_L)$, resolve o seguinte problema de maximização

$$\max_{q_2}[a - q_1^* - q_2 - c_L]q_2. \tag{6.116}$$

Portanto,

$$q_2^*(c_L) = \frac{a - q_1^* - c_L}{2}. \tag{6.117}$$

O ponto importante aqui é que ambos os tipos consideram o mesmo q_1^*, como essa é a estratégia da firma 1, cujo tipo é conhecido pelos dois tipos da firma 2. Agora considere a firma 1. Ela possui um tipo. A empresa 1 conhece a estratégia da firma 2, mas, como não sabe qual tipo de firma 2 está enfrentando, não sabe o nível de produção da firma 2. Na visão da firma 1, o nível de produção da firma 2 é $q_2^*(c_H)$ com probabilidade θ e $q_2^*(c_L)$ com probabilidade $1-\theta$. Assim, o lucro esperado da firma 1 do nível de produção q_1 é

$$\begin{aligned} U_1(q_1, q_2^*) &= \theta[a - q_1 - q_2^*(c_H)]q_1 + (1-\theta)[a - q_1 - q_2^*(c_L)]q_1 \\ &= \{a - q_1 - [\theta q_2^*(c_H) + (1-\theta)q_2^*(c_L)]\}q_1. \end{aligned} \tag{6.118}$$

O termo

$$\mathbb{E}[q_2] = \theta q_2^*(c_H) + (1-\theta)q_2^*(c_L), \tag{6.119}$$

é a produção esperada da firma 2. Assim,

$$U_1(q_1, q_2^*) = (a - q_1 - \mathbb{E}[q_2])q_1. \tag{6.120}$$

Sua estratégia q_1^* resolve o seguinte problema de maximização

$$\max_{q_1} U(q_1, q_2^*). \tag{6.121}$$

Neste caso particular, a melhor resposta ao nível de produção esperado é:

$$q_1^* = \frac{a - \mathbb{E}[q_2]}{2} = \frac{a - [\theta q_2^*(c_H) + (1-\theta)q_2^*(c_L)]}{2}. \tag{6.122}$$

É importante notar que a ação de equilíbrio é a melhor resposta à estratégia esperada do outro jogador quando e somente quando a ação dos outros jogadores afeta o retorno do jogador de forma linear, como neste caso. Em particular, quando as ações dos outros jogadores têm um efeito não-linear no payoff de um jogador, sua ação pode não ser a melhor resposta à ação esperada dos outros. É um erro comum tomar a ação de um jogador como melhor resposta à ação esperada dos outros; você deve evitá-lo.

Para computar o equilíbrio de Nash Bayesiano, simplesmente computamos:

$$\begin{cases} 2q_1^* + \theta q_2^*(cH) + (1-\theta)q_2^*(c_L) & = a \\ 2q_2^*(c_H) + q_1^* & = a - c_H \\ 2q_2^*(c_L) + q_1^* & = a - c_L \end{cases} \tag{6.123}$$

Escrevendo o sistema no formato matricial, chegamos a :

$$\begin{pmatrix} q_1^* \\ q_2^*(c_H) \\ q_2^*(c_L) \end{pmatrix} \begin{pmatrix} 2 & \theta & (1-\theta) \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix} = \begin{pmatrix} a \\ a - c_H \\ a - c_L \end{pmatrix} \tag{6.124}$$

Resolvendo por regra de Cramer, obtemos:

$$q_1^* = \frac{a + \theta c_H + (1-\theta)c_L}{3} \tag{6.125}$$

$$q_2^*(c_H) = \frac{a - 2c_H}{3} + \frac{(1-\theta)(c_H - c_L)}{6} \tag{6.126}$$

$$q_2^*(c_L) = \frac{a - 2c_L}{3} - \frac{\theta(c_H - c_L)}{6} \tag{6.127}$$

$$\tag{6.128}$$

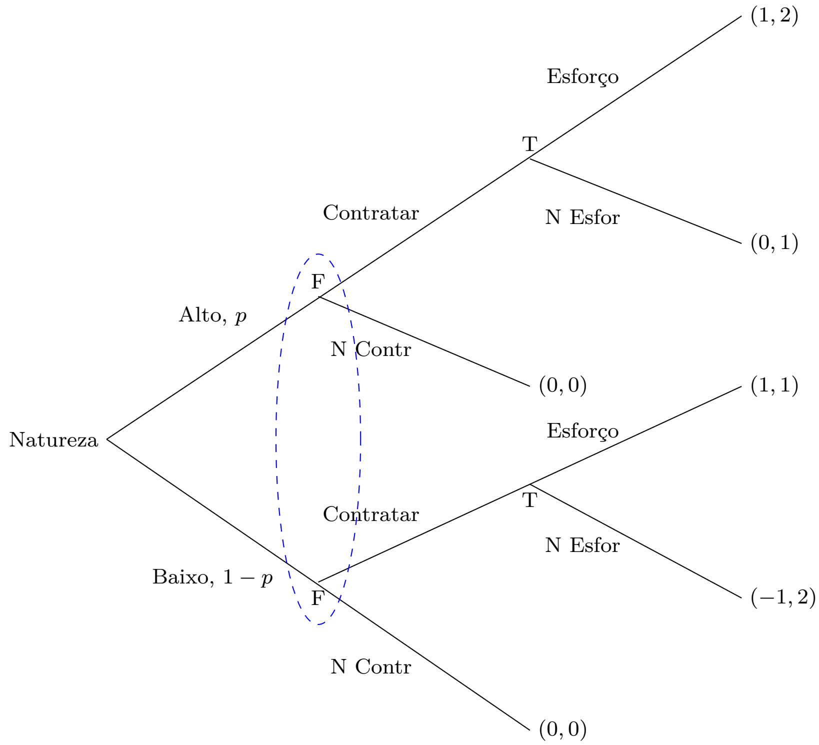
6.9 Jogos Dinâmicos com Informação Incompleta

Esta seção é dedicada aos conceitos básicos em jogos dinâmicos com informações incompletas. Como no caso de informações completas, o equilíbrio Bayesiano de Nash permite que os jogadores realizem ações sub-ótimas em conjuntos de informações que não são atingidos em equilíbrio. Esse problema é resolvido pelo equilíbrio sequencial, que requer explicitamente que os jogadores tenham uma melhor resposta em cada conjunto de informações (racionalidade sequencial) e que as crenças dos jogadores sejam “consistentes” com as estratégias dos outros jogadores. Aqui, vou definir o equilíbrio sequencial e aplicá-lo a alguns jogos importantes.

O equilíbrio sequencial está intimamente relacionado a outro conceito de solução, chamado equilíbrio de Nash Bayesiano perfeito. Equilíbrio sequencial é um conceito de solução melhor definido e mais fácil de entender. Os dois conceitos de solução são equivalentes nos jogos considerados aqui. Portanto, você deve aplicar o equilíbrio sequencial com relação ao perfeito equilíbrio de Nash

Bayesiano.

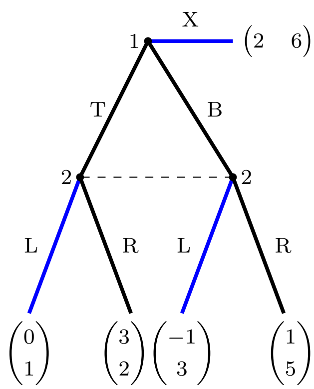
Figura 6.27 – Decisão de empregabilidade com informação incompleta



Este jogo destina-se a descrever uma situação em que uma empresa não sabe se um trabalhador é de alta habilidade, no sentido de se esforçar ao invés de não se esforçar. O trabalhador provavelmente está se esforçando. No entanto, há um equilíbrio de Nash Bayesiano em que o trabalhador não se esforçaria se fosse contratado, independente do seu tipo e antecipando isso, a empresa não contrata. Claramente, não se esforçar é contra suas preferências (que foram feitas para modelar um trabalhador que preferiria se esforçar). Isto é, no entanto, consistente com o equilíbrio bayesiano de Nash, porque toda estratégia do trabalhador é a melhor resposta à estratégia "não contratar" da firma. O trabalhador recebe 0 não importa a estratégia que ele usa. Para resolver esse problema, suponha que os jogadores sejam sequencialmente racionais, ou seja, eles respondam melhor em todas as informações, maximizando seu retorno esperado se estiverem nos conjuntos informacionais. Isto é, quando ele se mudar, o trabalhador deve saber que a Natureza escolheu "Alta" e a empresa escolheu "Contratar", e ele deve jogar "Esforçar" como a única melhor resposta sob esse conhecimento. Isso levaria ao outro equilíbrio, no qual a empresa contrata e o trabalhador se esforçar se ele for de alta habilidade duro e não se esforça caso contrário.

Observe que o último equilíbrio é o único equilíbrio perfeito para o subjogo nesse jogo. Como a perfeição do subjogo foi introduzida como um remédio para o problema exibido no antigo equilíbrio, é tentador pensar que a perfeição do subjogo resolve o problema. Como vimos nas seções anteriores, isso não acontece. Por exemplo, considere o perfil da estratégia na Figura abaixo. Este é um equilíbrio perfeito para o subjogo porque não existe um subjogo adequado e é claramente um equilíbrio de Nash. A estratégia L é a melhor resposta apenas para X. No entanto, no conjunto informacional que o jogador 2 se move, ela sabe que o jogador jogou T ou B. Dado este conhecimento, L não poderia ser a melhor resposta.

Figura 6.28 – Um equilíbrio perfeito em subjogo no qual o jogador 2 joga uma estratégia sequencialmente irracional



Para formalizar a ideia de racionalidade sequencial para jogos gerais, precisamos definir crenças:

Definição 6.9.1. *Uma crença é uma lista b de distribuições de probabilidade em conjuntos informacionais; para cada conjunto de informações I, b fornece uma distribuição de probabilidade $b(\cdot|I)$ em I.*

Para qualquer conjunto de informações I, o jogador que se move em I acredita estar no nó $n \in I$com probabilidade $b(n|I)$. Por exemplo, para o jogo na Figura acima, para definir uma avaliação de crença, precisamos atribuir uma probabilidade μ no nó após T e uma probabilidade $1 - \mu$ no nó após B. Em conjuntos de informações com nós únicos, a distribuição de probabilidade é trivial, colocando 1 no único nó. Quando o jogador 2 se move, ela acredita que o jogador 1 jogou T com probabilidade μ e B com probabilidade $1 - \mu$. Agora estamos prontos para definir a racionalidade sequencial para um perfil de estratégia:

Definição 6.9.2. *Para um dado par (s, b) dos perfis de estratégia e avaliação de crenças b, o perfil de estratégia s é dito sequencialmente racional se, em cada conjunto de informações I, o jogador que deve se mover maximiza sua utilidade esperada*

1. *dadas as suas crenças $b(\cdot|I)$ no conjunto de informações (o que implica que ele está no conjunto de informações I), e*
2. *dado que os jogadores vão jogar de acordo com s no jogo de continuação.*

Por exemplo, no jogo acima, para o jogador 2, dado qualquer crença μ, a jogada L gera

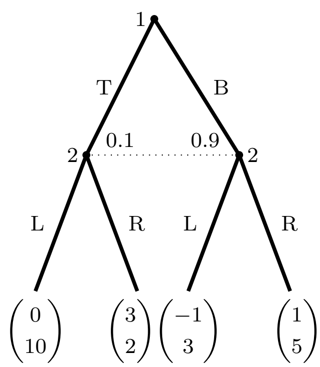
$$U_2(L; \mu) = 1 \cdot \mu + 3 \cdot (1 - \mu), \tag{6.129}$$

enquanto R gera

$$U_2(R;\mu) = 2 \cdot \mu + 5 \cdot (1-\mu). \tag{6.130}$$

Assim, a racionalidade sequencial requer que o jogador 2 jogue R. Dado que o jogador 2 joga R, a única melhor resposta para o jogador 1 é T. Portanto, para qualquer avaliação de crença b, o único perfil de estratégia sequencialmente racional é (T, R).

Para ter um equilíbrio, b também deve ser consistente com σ. Grosso modo, a consistência requer que os jogadores saibam quais estratégias (possivelmente mistas) são jogadas pelos outros jogadores. Para motivação, considere a Figura abaixo e chame o nó à esquerda de η_T e o nó à direita de η_B. Dadas as crenças $b(\eta_T|I_2) = 0.1$ e $b(\eta_B|I_2) = 0.9$, o perfil estratégico (T, R) é sequencialmente racional.



A estratégia T é a melhor resposta para R. Para verificar a racionalidade sequencial para R, basta observar que, dadas as crenças, L rende

$$0,1 \cdot 10 + 0.9 \cdot 3 = 3.7 \tag{6.131}$$

enquanto R gera

$$0.1 \cdot 2 + 0.9 \cdot 5 = 4.7 \tag{6.132}$$

Mas (T, R) não é nem mesmo um equilíbrio de Nash neste jogo. Isso ocorre porque, em um equilíbrio de Nash, o jogador conhece a estratégia do outro jogador. Ela saberia que o jogador 1 joga T e, portanto, ela atribui probabilidade 1 ao nó η_T. Em contraste, de acordo com b, ela atribui apenas probabilidade $0,1$ em η_T.

Para definir consistência formalmente, precisamos pensar com mais cuidado sobre os conjuntos de informações que são atingidos: probabilidade positiva (os conjuntos de informações que estão "no caminho") e os que não devem ser alcançados ("fora do caminho") de acordo com o perfil da estratégia.

Definição 6.9.3. *Dados quaisquer perfis de estratégia (possivelmente mistas), crença b e qualquer conjunto de informações I que seja alcançado com probabilidade positiva de acordo com s, as crenças $b(\cdot|I)$ em I são consideradas consistentes com s se e somente se $b(\cdot|I)$ é derivado usando a regra de Bayes e s. Ou seja, para cada nó η em I,*

$$b(\eta|I) = \frac{Pr(\eta|s)}{\sum_{n' \in I} Pr(\eta'|s)}, \tag{6.133}$$

em que $Pr(\eta|s)$ é a probabilidade de alcançarmos o nó η de acordo com s.

Por exemplo, para que uma avaliação de crença b seja consistente com (T, R), precisamos que

$$\mu = b(\eta_T|I) = \frac{\Pr(\eta_T|(T, R))}{\Pr(\eta_T|(T, R)) + \Pr(\eta_B|(T, R))} = \frac{1}{1+0} = 1. \tag{6.134}$$

Definição 6.9.4. *Um par (s, b) de um perfil estratégico e uma avaliação de crença b é dito ser um equilíbrio sequencial se (s, b) é sequencialmente racional e b é consistente com s.*

Note que um equilíbrio sequencial é um par, não apenas um perfil de estratégia. Assim, para identificar um equilíbrio sequencial, é preciso identificar um perfil de estratégia, que descreva o que um jogador faz em cada conjunto de informações, e uma avaliação de crenças b, que descreve o que um jogador acredita em cada conjunto de informações. Para verificar se aquele (s, b) é um equilíbrio sequencial, deve-se verificar que

1. (Racionalidade Sequencial) s é a melhor resposta para a crença $b(\cdot|I)$ e a crença de que os outros jogadores seguirão s nos jogos de continuação em cada conjunto de informações I, e

2. (Consistência) existem probabilidades que chegam a zero, de tal forma que as probabilidades condicionais derivadas de Bayes governam sob a abordagem $b(\cdot|I)$ em cada conjunto de informações I.

Referências Bibliográficas

[Cho and Kreps1987] Cho, I.-K. and Kreps, D. M. (1987). Signaling Games and Stable Equilibria. *Quarterly Journal of Economics*, 102(2):179–221.

[Fudenberg and Tirole1991] Fudenberg, D. and Tirole, J. (1991). *Game Theory*. Cambridge University Press.

[Gibbons1992] Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton University Press.

[Harsanyi1967] Harsanyi, J. C. (1967). Games with Incomplete Information Played by "Bayesian" Players, I, the basic model. *Management Science*, 14(3):159–182.

[Kreps1990] Kreps, D. M. (1990). *Game Theory and Economic Modelling*. Oxford University Press.

[Kreps et al.1990] Kreps, D. M. et al. (1990). *A Course in Microeconomic Theory*. Princeton University Press.

[Kreps and Wilson1982] Kreps, D. M. and Wilson, R. (1982). Sequential Equilibria. *Econometrica*, pages 863–894.

[Milgrom and Roberts1982] Milgrom, P. and Roberts, J. (1982). Predation, Reputation, and Entry Deterrence. *Journal of Economic Theory*, 27(2):280–312.

[Myerson2013] Myerson, R. B. (2013). *Game Theory*. Harvard University Press.

[Nash1951] Nash, J. (1951). Non-Cooperative Games. *Annals of Mathematics*, 54(2):286–295.

[Osborne et al.2004] Osborne, M. J. et al. (2004). *An Introduction to Ggame Theory*, volume 3. Oxford University Press.

[Osborne and Rubinstein1994] Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT press.

[Von Neumann et al.2007] Von Neumann, J., Morgenstern, O., and Kuhn, H. W. (2007). *Theory of Games and Economic Behavior*. Princeton University Press.

Capítulo 7

Externalidade e Bens Públicos

Contents

7.1 Bens Públicos

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 (Samuelson, 1954, p. 387). Este é comumente chamado de não-rivalidade em uso. Existem duas áreas importantes da economia em que os bens públicos desempenham um papel importante. O primeiro é o caso de gastar 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.

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 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.

7.1.1 Provisão Ótima de um Bem Público Puro

Sejam as seguintes características do modelo:

- n consumidores, indexados por $i = 1, \ldots, n$
- 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

$$u_i(x_i, G), \tag{7.1}$$

que é diferenciável, crescente em ambos os argumentos, quase-côncava e satisfaz as condições de Inada (o que garante a existência de uma solução interna)

- 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

$$G = f(z). \tag{7.2}$$

Primeiro, fazemos a pergunta normativa sobre qual é o nível ótimo de fornecimento 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 $x = (x_1, \ldots, x_n)$ para os agentes de acordo com o critério de Pareto.

Definição 7.1.1. *Uma alocação (x, G) é viável se existe um $z \geq 0$ tal que*

$$\sum_{i=1}^{n} x_i + z \leq W \tag{7.3}$$

$$G \leq f(z). \tag{7.4}$$

Alternativamente, uma alocação é viável se

$$\sum_{i=1]}^{n} x_i + f^{-1}(G) \leq W. \tag{7.5}$$

Definição 7.1.2. *Uma alocação viável (x, G) é um ótimo de Pareto se não existe oura alocação viável (x', G') tal que*

$$u_i(x_i', G') \geq u_i(x_i, G), \quad \forall i = 1, \ldots, n, \tag{7.6}$$

e para algum i

$$u_i(x_i', G') > u_i(x_i, G). \tag{7.7}$$

Ou seja, uma alocação factível (x, G) é ótima de Pareto, se não houver maneira de tornar um agente estritamente melhor sem prejudicar alguém. Agora podemos caracterizar o conjunto de alocações ótimas de Pareto. É a solução para o seguinte problema:

$$\max_{x,G,z} u_1(x_1, G) \tag{7.8}$$

$$
\begin{aligned}
\text{sujeito a} \quad & u_i(x_i, G) - \underline{u}_i \geq 0 \quad \text{para } i = 2, 3, \ldots, n && \text{com multiplicador } \gamma_i && (7.9)\\
& W - \sum_{i=1}^{n} x_i - z \geq 0 && \text{com multiplicador } \lambda && (7.10)\\
& f(z) - G \geq 0 && \text{com multiplicador } \mu && (7.11)\\
& G \geq 0, z \geq 0, x_i \geq 0 \quad \forall i = 1, 2, \ldots, n && && (7.12)
\end{aligned}
$$

em que u_i são tratados como parâmetros do problema. As condições de Inada na função de utilidade implicam que as restrições de não-negatividade podem ser ignoradas. As condições necessárias e suficientes (suficientes devido à quase-concavidade em u e f) de Kuhn-Tucker são:

$$
\frac{\partial L}{\partial x_i} = 0 \quad \Rightarrow \quad \gamma_i \frac{\partial u_i(x_i, G)}{\partial x_i} - \lambda = 0 \tag{7.13}
$$

$$
\frac{\partial L}{\partial G} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \gamma_i \frac{\partial u_i(x_i, G)}{\partial G} - \mu = 0 \tag{7.14}
$$

$$
\frac{\partial L}{\partial z} = 0 \quad \Rightarrow \quad -\lambda + \mu f'(z) = 0. \tag{7.15}
$$

em que $\gamma_1 = 1$ por convenção.

Temos que

$$
\gamma_i = \frac{\lambda}{\frac{\partial u_i(x_i, G)}{\partial x_i}}, \tag{7.16}
$$

e que

$$
\mu = \frac{\lambda}{f'(z)}. \tag{7.17}
$$

Combinando essas equações, temos:

$$
\sum_{i=1}^{n} \frac{\frac{\partial u_i(x_i, G)}{\partial G}}{\frac{\partial u_i(x_i, G)}{\partial x_i}} = \frac{1}{f'(z)}. \tag{7.18}
$$

Esta condição é referida como a condição de Samuelson, a condição de Lindahl-Samuelson, ou às vezes até mesmo a condição de Bowen-Lindahl-Samuelson.

O lado esquerdo da equação acima é a soma das taxas marginais de subtituição dos n agentes. Para ver isto, note que a partir da curva de indiferença do agente i, o termo $\frac{\frac{\partial u_i(x_i, G)}{\partial G}}{\frac{\partial u_i(x_i, G)}{\partial x_i}}$ 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). Daí a condição de Samuelson diz o seguinte: qualquer alocação ótima é tal que a soma da quantidade de consumidores que estaria disposta a desistir de 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

$$f(z_1, \ldots, z_k), \tag{7.19}$$

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

$$\sum_{i=1}^{n} \frac{\frac{\partial u_i(x_{ij}, G)}{\partial G}}{\frac{\partial u_i(x_{ij}, G)}{\partial x_{ij}}} = \frac{1}{\frac{\partial f(z_1, \ldots, z_k)}{\partial z_j}}, \quad \forall j = 1, \ldots, k. \tag{7.20}$$

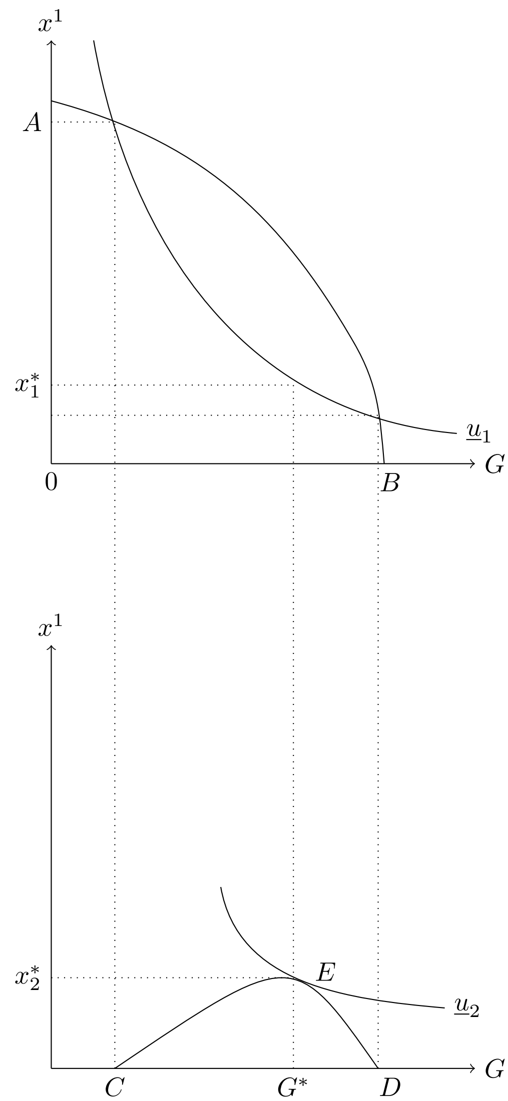
Uma ilustração gráfica da condição de Samuelson para o caso onde há dois indivíduos e dois bens é dada na Figura abaixo. Na Figura, a parte superior mostra as curvas de indiferença para o agente 1 e a restrição de produção AB. Suponha que nós fixamos o agente 1 na curva de indiferença $\underline{u}_1$; 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

$$TMgS_2 = TMgT - TMgS_1. \tag{7.21}$$

Implementação da alocação ótima: se o governo é capaz de cobrar impostos de montante fixo (lump-sum) tanto para financiar as despesas quanto para redistribuir a renda, então fica claro que 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, imposto sobre a renda do trabalho, para financiar 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. Pergunta prática: como a provisão ótima do bem público pode ser descentralizada, dadas as ferramentas políticas disponíveis e respeitando as restrições de informação?

Figura 7.1 – Provisão ótima de bens públicos



7.1.2 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. 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.

Dados os consumos de bens públicos por outros agentes, $\overline{g}_{-i} = (g_1, \ldots, g_{-i+1}, g_{i+1}, \ldots, g_n)$, a melhor resposta do agente i a $\overline{g}_{-i}$ dado o preço p é definido como

$$\beta_i(\overline{g}_{-i}, p) = \arg\max_{\{g_i\}} u_i \left(w_i - pg_i, g_i + \sum_{j \neq i} g_j \right) \tag{7.22}$$

$$\text{sujeito a} \quad g_i \geq 0 \tag{7.23}$$

$$w_i - pg_i \geq 0. \tag{7.24}$$

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

$$-\frac{\partial u_i}{\partial x_i} p + \frac{\partial u_i}{\partial G} + \lambda - \mu_p = 0 \tag{7.25}$$

$$\lambda g_i = 0 \tag{7.26}$$

$$\mu(w_i - pg_i) = 0. \tag{7.27}$$

Como a função de utilidade satisfaz as condições de Inada, $\mu = 0$. Portanto, temos:

$$p \geq \frac{\frac{\partial u_i}{\partial G}}{\frac{\partial u_i}{\partial x_i}}. \tag{7.28}$$

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

$$\max_{z \geq 0} pf(z) - z, \tag{7.29}$$

cuja condição ótima implica:

$$p = \frac{1}{f'(z)}. \tag{7.30}$$

Definição 7.1.3. *Um equilíbrio competitivo consiste de p^* e $G^* = (g_1^*, \ldots, g_n^*)$ de tal modo que:*

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

$$g_i^* \in \beta_i \left(\overline{g}_{-i}^*, p^* \right) \tag{7.31}$$

2. *a firma é otimizadora, isto é,*

$$p^* = \frac{1}{f'(f^{-1}(\sum_{i=1}^{n} g_i^*))} \tag{7.32}$$

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

$$p = \frac{\frac{\partial u_i}{\partial G}}{\frac{\partial u_i}{\partial x_i}}, \tag{7.33}$$

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

$$\frac{1}{f'(z)} = \frac{\frac{\partial u_j}{\partial G}}{\frac{\partial u_j}{\partial x_j}}. \tag{7.34}$$

No equilíbrio competitivo, chegamos a:

$$\sum_{i=1}^{n} \frac{\frac{\partial u_i}{\partial G}}{\frac{\partial u_i}{\partial x_i}} \geq \frac{1}{f'(z)}. \tag{7.35}$$

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 7.1.1. *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.*

Sabemos que:

$$\gamma_i \frac{1}{x_i} = \lambda \tag{7.36}$$

$$\sum_{i=1}^{n} \gamma_i \frac{\delta}{G} - \mu = 0 \tag{7.37}$$

$$-\lambda + \mu f'(z) = 0 \tag{7.38}$$

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

$$\begin{aligned}
\sum_{i=1}^{n} \lambda x_i \frac{\delta}{G} - \lambda =& 0 \\
\lambda \frac{\delta}{G} \sum_{i=1}^{n} x_i - \lambda =& 0 \\
\lambda \left(\frac{\delta}{G} \sum_{i=1}^{n} x_i - 1 \right) =& 0 \\
\frac{\delta}{G} \sum_{i=1}^{n} x_i =& 1 \\
\sum_{i=1}^{n} x_i =& \frac{G}{\delta} \qquad (7.39)
\end{aligned}$$

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

$$\begin{aligned}
W - \sum_{i=1}^{n} x_i - G =& 0 \\
W - \frac{G}{\delta} - G =& 0 \\
G =& \frac{\delta}{1+\delta} W \qquad (7.40)
\end{aligned}$$

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

$$\begin{aligned}
\sum_{i=1}^{n} x_i =& \frac{G}{\delta} \\
\sum_{i=1}^{n} x_i =& \frac{\frac{\delta}{1+\delta} W}{\delta} \\
n\overline{x}_i =& \frac{\frac{\delta}{1+\delta} W}{\delta} \\
\overline{x}_i =& \frac{W}{n(1+\delta)} \qquad (7.41)
\end{aligned}$$

No caso da alocação de equilíbrio competitivo, sabemos que:

$$-\frac{1}{x_i} p + \frac{\delta}{G} + \lambda - \mu p = 0 \qquad (7.42)$$

$$\lambda G = 0 \qquad (7.43)$$

$$\mu(w_i - pG) = 0 \tag{7.44}$$

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

$$\begin{aligned} -\frac{1}{x_i}p + \frac{\delta}{G} &= 0 \\ x_i &= \frac{G}{\delta} \quad [p = f'(z) \longrightarrow p = 1] \\ x_i &= \frac{G}{\delta} \end{aligned} \tag{7.45}$$

Logo,

$$\begin{aligned} W - \sum_{i=1}^{n} x_i - G &= 0 \\ W - \sum_{i=1}^{n} \frac{G}{\delta} - G &= 0 \\ G &= \frac{\delta}{n+\delta} W \end{aligned} \tag{7.46}$$

e

$$\begin{aligned} x_i &= \frac{G}{\delta} \\ &= \frac{W}{n+\delta} \end{aligned} \tag{7.47}$$

7.1.3 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 idéia é pensar na quantia comprada por cada agente como uma mercadoria distinta e fazer com que cada agente enfrente um preço personalizado p_i e tenha esses preços escolhidos de forma que todos os agentes concordem com o nível do bem público. Seja $s_i \in [0, 1]$ o share do lucro da firma que o indivíduo tem com $\sum_{i=1}^{n} s_i = 1$:

Definição 7.1.4. *Um equilíbrio de Lindhal é um vetor de preços $p^* = (p_1^*, \ldots, p_n^*)$ e uma alocação $(x_1^*, \ldots, x_n^*, G^*)$ tal que:*

- *A firma maximiza lucros, isto é,*

$$G^* = \arg\max_{G \geq 0} \left(\sum_{i=1}^{n} p_i^* \right) G - f^{-1}(G) \tag{7.48}$$

- *Cada consumidor maximiza utilidade, isto é,*

$$(x_i^*, G) = \arg\max_{x_i, G} u_i(x_i, G) \ \textit{sujeito a} \quad w_i + s_i \left(\sum_i p_i^* G^* - f^{-1}(G^*) \right) - x_i - p_i^* G \geq 0 \tag{7.49}$$

- *Equilíbrio do mercado*

$$\sum_{i=1}^{n} x_i^* + f^{-1}(G^*) \leq \sum_{i=1}^{n} w_i \tag{7.50}$$

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, isto é, os bens públicos do agente 1 através do agente n. 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, observe que a condição de primeira ordem para a maximização do lucro da empresa resulta em

$$\sum_{i=1}^{n} p_i^* = \frac{1}{f'(f^{-1}(G^*))}, \tag{7.51}$$

e a condição de primeira ordem para a maximização da utilidade do indivíduo i é:

$$\frac{\partial u_i(x_i^*, G^*)}{\partial x_i} p_i^* = \frac{\partial u_i(x_i^*, G^*)}{\partial G}, \forall i = 1, \ldots, n. \tag{7.52}$$

Portanto,

$$\sum_{i=1}^{n} \frac{\frac{\partial u_i(x_i^*, G^*)}{\partial G}}{\frac{\partial u_i(x_i^*, G^*)}{\partial x_i}} = \frac{1}{f'(f^{-1}(G^*))}, \tag{7.53}$$

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.

Diferença chave entre o equilíbrio de Lindahl e o equilíbrio padrão: nenhum mecanismo descentralizado irá gerar o vetor de preços correto.

Teorema 7.1.1. *Toda alocação de Lindhal (x^*, G^*) com vetor de preços p^* é fracamente Pareto eficiente, e sob não saciedade local é Pareto eficiente.*

7.1.4 O Problema do Free-Rider

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 necessariamente resulte em uma quantidade eficiente do bem público que está sendo produzido. 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 \$ 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	0, 0

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 órgão quer dividir o custo de produção do projeto público, mas quer aproveitar 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.

7.1.5 Mecanismo de Votação

Na prática, as pessoas votam sobre a quantidade G de bens públicos a serem fornecidos. A votação leva à solução de first-best?

O equilíbrio da votação é G^* 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 defesa).
- 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.

7.1.6 Crowd-Out: Provisão Pública com Provisão Privada Endógena

Suponha que os indivíduos resolvam o problema tradicional

$$\max u_i(x_i, g_i + \sum_{j \neq i} g_j) \quad \text{sujeito a} \quad x_i + g_i = w_i. \tag{7.54}$$

O ótimo, pelo equilíbrio de Nash, é que todos otimizem suas funções de utilidade considerando o comportamento dos outros agentes. Assim, g^* é o equilíbrio de mercado (privado).

Agora, suponha que o governo introduza impostos lump-sum τ_i para financiar o bem-público, tal que $T = \sum \tau_i$. Com isso o problema do agente se torna:

$$\max u_i(x_i, g_i + \sum_{j \neq i} g_j + \tau_i) \quad \text{sujeito a} \quad x_i + g_i = w_i - \tau_i, \tag{7.55}$$

que pode ser reescrito como

$$\max u_i(x_i, z_i + \sum_{j \neq i} z_j + \tau_i) \quad \text{sujeito a} \quad x_i + z_i = w_i. \tag{7.56}$$

Isto é isomorfo ao problema original: $z^* = g^*$. A provisão total de bens públicos é inalterada. Cada pessoa simplesmente reduz a provisão voluntária por τ.

7.2 Externalidades

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

Exemplos:

- consumo de maçã: externalidade pecuniária, internalizada em preços de mercado.
- poluição/consumo de música alta: essas externalidades entram diretamente nas funções de utilidade ou produção.

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.

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

$$u_i(x_i) \quad \text{sem preferência por externalidades} \tag{7.57}$$

$$u_i(x_1, \ldots, x_i, \ldots, x_n) \quad \text{com preferência por externalidades} \tag{7.58}$$

   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.

7.2.1 Externalidades no Consumo

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

$$u_i(x_i). \tag{7.59}$$

Neste caso, as condições de primeira ordem para o equilíbrio competitivo são dadas por

$$TMgS_{xy}^A = \frac{p_x}{p_y} = TMgS_{xy}^B, \tag{7.60}$$

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

$$TMgS_{xy}^A = TMgS_{xy}^B. \tag{7.61}$$

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. Os seguintes materiais são baseados principalmente de Tian e Yang (2009).

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

$$u_A(x_A, x_B, y_A), \tag{7.62}$$
$$u_B(x_A, x_B, y_B), \tag{7.63}$$

que são consideradas estritamente crescentes em seu próprio consumo de bens, quase-côncavas, e satisfazem as condições de Inada, $\frac{\partial u}{\partial x_i}(0) = +\infty$ e $\lim_{x_i \to 0} \frac{\partial u}{\partial x_i} x_i = 0$, resultando em soluções interiores. Além disso, assumimos que o gradiente 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

$$TMgS_{xy}^{A} = \frac{p_x}{p_y} = TMgS_{xy}^{B}. \tag{7.64}$$

Agora encontramos as condições de primeira ordem para alocações eficientes de Pareto em economias de troca com externalidades. Assim, alocações eficientes de Pareto x^* podem ser completamente determinadas pelas condições de primeira ordem do seguinte problema:

$$\begin{aligned}
&\max_{x \in \mathbb{R}_{++}} u_B(x_A, x_B, y_B) && (7.65)\\
\text{sujeito a} \quad & x_A + x_B \leq w_x && (7.66)\\
& y_A + y_B \leq w_y && (7.67)\\
& u_A(x_A, x_B, y_A) \geq u_A(x_A^*, x_B^*, y_A^*). && (7.68)
\end{aligned}$$

O Lagrangeano é

$$\begin{aligned}
L =& u_B(x_A, x_B, y_B) + \lambda_x(w_x - x_A - x_B) + \lambda_y(w_y - y_A - y_B) + \\
&+ \mu(u_A(x_A, x_B, y_A) - u_A(x_A^*, x_B^*, y_A^*)). && (7.69)
\end{aligned}$$

As condições de primeira ordem são:

$$\begin{aligned}
\frac{\partial L}{\partial x_A} = 0 \quad &\longrightarrow \quad \frac{\partial u_B}{\partial x_A} - \lambda_x + \frac{\partial u_A}{\partial x_A} = 0 && (7.70)\\
\frac{\partial L}{\partial y_A} = 0 \quad &\longrightarrow \quad -\lambda_y + \frac{\partial u_A}{\partial y_A} = 0 && (7.71)\\
\frac{\partial L}{\partial x_B} = 0 \quad &\longrightarrow \quad \frac{\partial u_B}{\partial x_B} - \lambda_x + \frac{\partial u_A}{\partial x_B} = 0 && (7.72)\\
\frac{\partial L}{\partial y_B} = 0 \quad &\longrightarrow \quad \frac{\partial u_B}{\partial y_B} - \lambda_y = 0 && (7.73)\\
\frac{\partial L}{\partial \lambda_x} = 0 \quad &\longrightarrow \quad w_x - x_A - x_B \geq 0 && (7.74)\\
& \qquad \lambda_x \geq 0 && (7.75)\\
& \qquad \lambda_x(w_x - x_A - x_B) = 0 && (7.76)\\
\frac{\partial L}{\partial \lambda_y} = 0 \quad &\longrightarrow \quad w_y - y_A - y_B \geq 0 && (7.77)\\
& \qquad \lambda_y \geq 0 && (7.78)\\
& \qquad \lambda_y(w_y - y_A - y_B) = 0 && (7.79)\\
\frac{\partial L}{\partial \mu} = 0 \quad &\longrightarrow \quad u_A(x_A, x_B, y_A) - u_A(x_A^*, x_B^*, y_A^*) \geq 0 && (7.80)\\
& \qquad \mu \geq 0 && (7.81)
\end{aligned}$$

$$\mu(u_A(x_A, x_B, y_A) - u_A(x_A^*, x_B^*, y_A^*)) = 0 \tag{7.82}$$

Por (7.73) temos que $\lambda_y = \dfrac{\partial u_B}{\partial y_B} > 0$ e, portanto, por (7.79),

$$y_A + y_B = w_y, \tag{7.83}$$

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

De (7.71) e (7.73), temos:

$$\mu = \frac{\dfrac{\partial u_B}{\partial y_B}}{\dfrac{\partial u_A}{\partial y_A}}. \tag{7.84}$$

Então, por (7.70) e (7.71), chegamos a:

$$\frac{\lambda_x}{\lambda_y} = \frac{\dfrac{\partial u_A}{\partial x_A}}{\dfrac{\partial u_A}{\partial y_A}} + \frac{\dfrac{\partial u_B}{\partial x_A}}{\dfrac{\partial u_B}{\partial y_B}}. \tag{7.85}$$

E por (7.72) e (7.73), chegamos a:

$$\frac{\lambda_x}{\lambda_y} = \frac{\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_B}{\partial y_B}} + \frac{\dfrac{\partial u_A}{\partial x_B}}{\dfrac{\partial u_A}{\partial y_A}}. \tag{7.86}$$

Logo,

$$\frac{\dfrac{\partial u_A}{\partial x_A}}{\dfrac{\partial u_A}{\partial y_A}} + \frac{\dfrac{\partial u_B}{\partial x_A}}{\dfrac{\partial u_B}{\partial y_B}} = \frac{\dfrac{\partial u_B}{\partial x_B}}{\dfrac{\partial u_B}{\partial y_B}} + \frac{\dfrac{\partial u_A}{\partial x_B}}{\dfrac{\partial u_A}{\partial y_A}}, \tag{7.87}$$

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 $\dfrac{\frac{\partial u_i}{\partial x_i}}{\frac{\partial u_i}{\partial y_i}} + \dfrac{\frac{\partial u_j}{\partial x_i}}{\frac{\partial u_j}{\partial y_j}}$.

Resolvendo (7.70) e (7.72) para μ e λ_x, obtemos:

$$\mu = \frac{\frac{\partial u_B}{\partial x_B} - \frac{\partial u_B}{\partial x_A}}{\frac{\partial u_A}{\partial x_A} - \frac{\partial u_A}{\partial x_B}} > 0, \tag{7.88}$$

e

$$\lambda_x = \frac{\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_A}{\partial x_A} - \frac{\partial u_A}{\partial x_B}}. \tag{7.89}$$

Quando a externalidade no consumo é positiva, a partir de (7.85) ou (7.86), podemos facilmente ver que λ_x é sempre positivo, dado que $\lambda_y = \frac{\partial u_B}{\partial y_B} > 0$. Além disso, quando não existe externalidade ou uma externalidade unilateral (somente um consumidor impõe uma externalidade sobre o outro), por (7.85) ou (7.86), λ_x é positivo. Assim, a condição de igualdade marginal (7.87) 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 (7.89) ou indiretamente dado por (7.85) ou (7.86) 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,

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \geq 0, \tag{7.90}$$

e

$$\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} \geq \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}. \tag{7.91}$$

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 conjunto de consumir o bem x.

Assim, as seguintes condições

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \geq 0 \tag{7.92}$$

$$y_A + y_B = w_y \tag{7.93}$$

$$x_A + x_B \leq w_x \tag{7.94}$$

$$\left(\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}\right)(w_x - x_A - x_B) = 0, \tag{7.95}$$

constituem um sistema que permite obter todas as alocações eficientes de Pareto. Podemos fazer isso considerando três casos.

1. Quando $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} > \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}$ ou $\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} > 0$, $\lambda_x > 0$ 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 (7.87), 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:

$$\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} = \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}, \tag{7.96}$$

então,

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} = 0, \tag{7.97}$$

e, portanto, $\lambda_x = 0$. Neste caso, quando $x_A + x_B \leq w_x$, a necessidade de destruição é indeterminada. No entanto, mesmo quando a destruição é necessária, ainda podemos determinar o conjunto de alocações eficientes de Pareto usando $y_A + y_B = w_y$ e as condições (7.97). De fato, depois de substituir $y_A + y_B = w_y$ em (7.97), 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 (7.87), as taxas marginais sociais de

substituição devem ser negativas. Assim, a alocação não será Pareto eficiente. Nesse caso, deve haver uma destruição do bem x para atingir a eficiência de Pareto e uma alocação eficiente de Pareto irá satisfazer (7.97).

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:

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \tag{7.98}$$

$$x_A + x_B = w_x \tag{7.99}$$

$$y_A + y_B = w_y \tag{7.100}$$

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:

$$\frac{\frac{\partial u_A}{\partial x_A}}{\frac{\partial u_A}{\partial y_A}} + \frac{\frac{\partial u_B}{\partial x_A}}{\frac{\partial u_B}{\partial y_B}} = \frac{\frac{\partial u_B}{\partial x_B}}{\frac{\partial u_B}{\partial y_B}} + \frac{\frac{\partial u_A}{\partial x_B}}{\frac{\partial u_A}{\partial y_A}} \tag{7.101}$$

$$x_A + x_B \leq w_x \tag{7.102}$$

$$y_A + y_B = w_y \tag{7.103}$$

Exemplo 7.2.1. *Seja a seguinte função de utilidade:*

$$u_i(x_A, x_B, y_i) = \sqrt{x_i y_i} - x_j, \quad i \in \{A, B\}, j \in \{A, B\}, j \neq i \tag{7.104}$$

Pela condição (7.87), temos que

$$\frac{y_A}{x_A} = \frac{y_B}{x_B} \tag{7.105}$$

Vamos supor que $x_A + x_B \equiv \overline{x}$. Substituindo $\overline{x}$ e $y_A + y_B = w_y$ na condição acima, obtemos:

$$\begin{aligned} \frac{y_A}{x_A} &= \frac{y_B}{x_B} \\ \frac{y_A}{x_A} &= \frac{w_y - y_A}{\overline{x} - y_A} \\ y_A\overline{x} - y_A x_A &= w_y x_A - y_A x_A \end{aligned}$$

$$\frac{y_A}{x_A} = \frac{w_y}{\overline{x}} \tag{7.106}$$

Temos que

$$\begin{aligned}
\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} &= \frac{1}{4}\sqrt{\frac{y_A}{x_A}}\sqrt{\frac{y_B}{x_B}} \\
&= \frac{1}{4}\sqrt{\frac{y_A}{x_A}}\sqrt{\frac{y_A}{x_A}} \qquad [por\ (7.105)] \\
&= \frac{1}{4}\frac{y_A}{x_A} \\
\frac{1}{4}\frac{y_A}{x_A} &= \frac{1}{4}\frac{w_y}{\overline{x}} \qquad (7.107)
\end{aligned}$$

Como $\frac{\partial u_A}{\partial x_A}\frac{\partial u_B}{\partial x_B} = \frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A}$ *e* $\frac{\partial u_A}{\partial x_B}\frac{\partial u_B}{\partial x_A} = 1$, *temos que*

$$\overline{x} = \frac{w_y}{4} \tag{7.108}$$

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* $w_x = \frac{w_y}{4}$.

7.2.2 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.

O lucro das duas firmas é

$$\pi_1 = p_x x - c_x(x), \tag{7.109}$$

$$\pi_2 = p_y y - c_y(y) - e(x), \tag{7.110}$$

em que p_x e p_y são os preços de x e de y, respectivamente. Então, pelas condições de primeira ordem, temos quantidades positivas de produtos:

$$p_x = c'_x(x), \tag{7.111}$$

$$p_y = c'_y(y). \tag{7.112}$$

No entanto, o produto maximizador de lucro, x_c, 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_c e y_c que satisfazem as duas condições de primeira ordem. Se as duas empresas se fundiram de modo a internalizar a externalidade

$$\max_{x,y} = p_x x + p_y y - c_x(x) - c_y(y) - e(x), \tag{7.113}$$

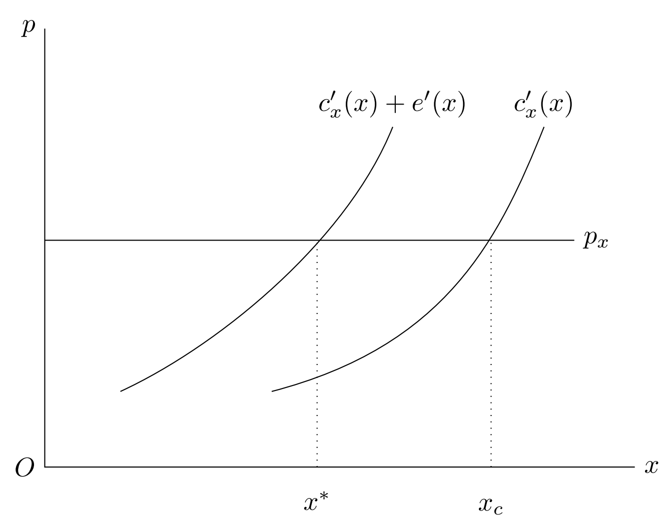
que resulta nas seguintes condições de primeira ordem:

$$p_x = c'_x(x^*) + e'(^*), \tag{7.114}$$

$$p_y = c'_y(y^*), \tag{7.115}$$

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 7.2 – O PRODUTO EFICIENTE x^* É MENOR DO QUE O PRODUTO EM UM AMBIENTE COMPETITIVO



7.2.3 Soluções para as Exetrnalidades

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.

7.2.3.1 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 função lucro é dada por

$$\pi_1 = p_x x - c_x(x) - tx. \tag{7.116}$$

A condição de primeira ordem é

$$p_x = c'_x(x) + t = c'_x(x) + e'(x^*), \tag{7.117}$$

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.

7.2.3.2 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 negoociação 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 $50.000. Se os produtos químicos poluem o lago, o peixe não pode ser comido. Como se resolve a externalidade? O método 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 $c_f = \$50.000$. 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 $c_f > \$50.000$. 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 \$ 50.000.

(a) Seja $c_f = \$50.000$. A fábrica compra o filtro para que o produto químico não possa poluir o lago.

(b) $c_f > \$50.000$. A firma paga \$50.000 ao pescador e o produto químico é descartado no lago.

Como no exemplo acima, os exemplos de Coase apóiam 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 consumidorese 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,

$$u_i(x_i^1, x_i^L, h). \tag{7.118}$$

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 reproduzi-la, 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

$$\frac{\partial u_2}{\partial h} \neq 0. \tag{7.119}$$

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:

$$\begin{aligned} v_i(w_i, h) &= \max_{x_i} u_i(x_i, h) \\ \text{sujeito a} \quad & px_i \leqq w_i. \end{aligned} \tag{7.120}$$

Assumimos que as preferências são quase-lineares em relação a algum bem numerário. Assim, a função de utilidade indireta do consumidor assume a forma:

$$v_i(w_i, h) = \phi_i(h) + w_i. \tag{7.121}$$

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. Para maximizar a utilidade, 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:

$$\max_h \phi_1(h) + \phi_2(h). \tag{7.122}$$

A condição de primeira ordem para um máximo interior é:

$$\phi_1(h^{**}) + \phi_2(h^{**}) = 0. \tag{7.123}$$

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), o nível de $h^* > h^{**}$. 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:

$$\phi_1(h_2 - T_2) \geqq \phi(0). \tag{7.124}$$

Dada essa restrição no conjunto de ofertas aceitáveis, o consumidor 2 escolherá (h_2, T_2) para resolver o seguinte problema:

$$\begin{aligned} &\max_{h_2, T_2} \phi_2(h_2) + T_2 \\ \text{sujeito a} \quad &\phi_1(h_2) - T_2 \geqq \phi_1(0). \end{aligned} \tag{7.125}$$

Como o consumidor 2 prefere T_2 mais alto, a restrição será binding no ótimo. Assim, o problema se torna:

$$\max_{h_2} \phi_1(h_2) + \phi_2(h_2) - \phi_1(0). \tag{7.126}$$

A condição de primeira ordem para este problema é dada por

$$\phi_1'(h_2) + \phi_2'(h_2) = 0. \tag{7.127}$$

Mas essa é a mesma condição que define o nível socialmente ótimo de h_2. Assim, o consumidor 2 escolhe $h_2 = h^{**}$, e usando a restrição, $T_2 = \phi(h^{**}) - \phi_1(0)$. E 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 2 é:

$$\begin{aligned} &\max_{h_1, T_1} \phi_1(h_1) + T_1 \\ \text{sujeito a} \quad &\phi_1(h_1) + T_1 \geqq \phi_1(0). \end{aligned} \tag{7.128}$$

Mais uma vez, sabemos que a restrição será binding, e assim o consumidor 2 escolhe h_1 e T_1 para maximizar

$$\max_{h_1} \phi_1(h_1) + \phi_2(h_1) - \phi_1(h^*). \tag{7.129}$$

A condição de primeira ordem para este problema é dada por

$$\phi_1'(h_1) + \phi_2'(h_1) = 0, \tag{7.130}$$

que é ótima em $h_1 = h^{**}$. A única diferença aqui é que $T_1 = \phi_1(h^*) - \phi_1(h^{**})$.

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.

7.2.3.3 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:

$$\pi_1 = p_x x_1 + r x_1 - c_1(x_1), \tag{7.131}$$

$$\pi_2 = p_y y - r x_2 - e_2(x_2) - c_y(y). \tag{7.132}$$

As respectivas condições de primeira ordem são:

$$p_x + r = c_1'(x_1), \tag{7.133}$$

$$p_y = c_y'(y) \quad \text{ou} \quad -r = e'(x_2). \tag{7.134}$$

No equilíbrio de mercado, $x_1^* = x_2^* = x^*$. Logo,

$$p_x = c_1'(x^*) + e'(x^*). \tag{7.135}$$

o que resulta em um resultado social ótimo.

7.2.3.4 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)[1] 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 de 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.

[1] Varian, H. "*A Solution to the Problem of Externalities when Agents Are Well Informed*". American Economic Review, 84(1994), 1278–1293.

O mecanismo tem dois estágios:

1. Fase de anúncio: as empresas 1 e 2 dizem as taxas de imposto Pigouvianas, $t_i, i = 1, 2$, 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 $t_2 x$ para firmar 2. Assim, cada firma considera a taxa de imposto como dada. A empresa 2 recebe as unidades $t_1 x$ como compensação. Cada empresa paga uma multa, $(t_1 - t_2)^2$, se anunciar taxas de impostos diferentes.

Portanto, os lucros das duas firmas são dados por:

$$\pi_1^* = \max_x p_x x - c_x(x) - t_2 x - (t_1 - t_2)^2, \tag{7.136}$$

$$\pi_2^* = \max_y p_y y - c_y(y) + t_1 x - e(x) - (t_1 - t_2)^2. \tag{7.137}$$

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:

$$p_x - c_x'(x) - t_2 = 0. \tag{7.138}$$

Note que, pela convexidade de c_x, ou seja, $c_x''(x) > 0$, temos

$$x'(t_2) = -\frac{1}{c''(x)} < 0. \tag{7.139}$$

A firma 2 escolherá y que satisfaça $p_y = c_y'(y)$.

Estágio 1

Cada empresa escolherá a alíquota t_1 e t_2 que maximize seus lucros.

Para a firma 1,

$$\max_{t_1} p_x x - c_x(x) - t_2 x(t_2) - (t_1 - t_2)^2, \tag{7.140}$$

o que nos dá a condição de primeira ordem:

$$2(t_1 - t_2) = 0. \tag{7.141}$$

Então a solução ótima é $t_1^* = t_2$.

Para a firma 2,

$$\max_{t_2} p_y y - c_y(y) - t_1 x(t_2) - e(x(t_2)) - (t_1 - t_2)^2, \tag{7.142}$$

o que nos dá a condição de primeira ordem:

$$t_1 x'(t_2) - e'(x(t_2))x'(t_2) + 2(t_1 - t_2) = 0 \quad \Longrightarrow \quad [t_1 - e'(x(t_2))]x'(t_2) + 2(t_1 - t_2) = 0. \tag{7.143}$$

Disso decorre que

$$t^* = e'(x(t^*)) \quad \text{com} \quad t^* = t_1^* = t_2^*. \tag{7.144}$$

Substituindo a taxa de imposto de equilíbrio, $t^* = e'(x(t^*))$, que maximiza o lucro da firma 1 no segundo estágio, obtemos:

$$p_x = c_x'(x^*) + e'(x^*) \tag{7.145}$$

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.

Referências Bibliográficas

[Aivazian and Callen1981] Aivazian, V. A. and Callen, J. L. (1981). The Coase Theorem and the Empty Core. *The Journal of Law and Economics*, 24(1):175–181.

[Arrow1979] Arrow, K. J. (1979). The Property Rights Doctrine and Demand Revelation under Incomplete Information. In *Economics and Human Welfare*, pages 23–39. Elsevier.

[Chipman1998] Chipman, J. S. (1998). A Close Look to the Coase Theorem. *The Economists' Vision: Essays in Modern Economic Perspectives*, pages 131–162.

[Chipman and Tian2012] Chipman, J. S. and Tian, G. (2012). Detrimental Externalities, Pollution Rights, and the "Coase Theorem". *Economic Theory*, 49(2):309–327.

[Coase1960] Coase, R. H. (1960). The Problem of Social Cost. In *Classic Papers in Natural Resource Economics*, pages 87–137. Springer.

[Foley1970] Foley, D. K. (1970). Lindahl's Solution and the Core of an Economy with Public Goods. *Econometrica:*, 38:66–72.

[Hurwicz1995] Hurwicz, L. (1995). What is the Coase Theorem? *Japan and the World Economy*, 7(1):49–74.

[Laffont et al.1988] Laffont, J.-J. et al. (1988). Fundamentals of Public Economics. *MIT Press Books*, 1.

[Leach2004] Leach, J. (2004). *A Course in Public Economics*. Cambridge University Press.

[Mas-Colell et al.1995] Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic Theory*, volume 1. Oxford university press New York.

[Milleron1972] Milleron, J.-C. (1972). Theory of Value with Public Goods: A Survey Article. *Journal of Economic Theory*, 5(3):419–477.

[Muench1972] Muench, T. J. (1972). The Core and the Lindahl Equilibrium of an Economy with a Public Good: An Example. *Journal of Economic Theory*, 4(2):241–255.

[Roberts1974] Roberts, D. J. (1974). The Lindahl Solution for Economies with Public Goods. *Journal of Public Economics*, 3(1):23–42.

[Salanié2000] Salanié, B. (2000). *Microeconomics of Market Failures.* MIT Press.

[Tian2003] Tian, G. (2003). A Solution to the Problem of Consumption Externalities. *Journal of Mathematical Economics*, 39(8):831–847.

Capítulo 8

Equilíbrio Geral

Contents

8.1 Introdução

As principais ideias da teoria do equilíbrio geral têm uma longa história, remontando às descrições evocativas de Adam Smith de como a competição canaliza o interesse próprio individual no interesse social e como um senso de coerência entre o vasto número de indivíduos e decisões aparentemente separadas (Arrow, 1972) pode surgir na economia sem desenho explícito. A teoria do equilíbrio geral aborda como essa "coerência" agregada emerge das interações individuais e pode potencialmente levar a alocações socialmente desejáveis de bens e serviços na economia. O mecanismo pelo qual essa coerência emerge é, naturalmente, o mecanismo de preços. Indivíduos que enfrentam os mesmos preços, determinados adequadamente, acabarão tomando decisões que são bem coordenadas no nível da economia.

8.2 Economia de Troca Pura

Um modelo de equilíbrio geral descreve três atividades básicas que ocorrem na economia: produção, troca e consumo. Como sempre, a exposição de um modelo econômico especifica uma descrição completa do ambiente econômico (jogadores, ações e preferências) e o conceito de solução que será usado para derivar prescrições e previsões.

Formalmente, uma economia de troca pura é uma economia em que não há oportunidades de produção. Há I consumidores, $i \in \mathcal{I} = \{1, \ldots, I\}$, que compram, vendem e consomem L commodities, $\ell \in \mathcal{L} = \{1, \ldots, L\}$. Uma cesta de consumo do consumidor i é um vetor $x_i = (x_{1,i}, \ldots, x_{L,i})$, com $x_i \in \mathcal{X}_i$, que descreve o conjunto de consumo do consumidor i, definindo apenas suas cestas de consumo viáveis. Vamos supor que $\mathcal{X}_i$ contém o vetor 0 e é um conjunto convexo. O consumidor i tem uma dotação das mercadorias L, que é descrita por um vetor $\omega = (\omega_{1,i}, \ldots, \omega_{L,i})$ e tem preferências sobre as cestas de consumo, que podem ser representadas por uma função de utilidade $u_i \colon \mathcal{X}_i \to \mathbb{R}$. Uma economia de troca pura é, portanto, um conjunto $\varepsilon = ((u_i, \omega_i)_{i \in \mathcal{I}})$, que descreve completamente a base do modelo: o conjunto de jogadores, suas preferências e suas dotações.

Cada consumidor se defronta com um vetor de preços $p = (p_1, \ldots, p_L)$ e resolve seu problema de maximização, qual seja:

$$\begin{aligned} & \max_{x_i \in \mathcal{X}} \; u_i(x_i) \\ \text{sujeito a} \quad & p x_i \leq p \omega_i, \end{aligned} \tag{8.1}$$

em que a restrição orçamentária do consumidor é sua riqueza, medida pelo valor de mercado de sua dotação a preços p. As ações viáveis do consumidor são, portanto, $x_i \in \mathcal{B}_i(p) \equiv \{x_i \in \mathcal{X}_i \colon p x_i \leq p\omega_i\}$, em que nos referimos a $\mathcal{B}_i(p)$ como seu orçamento definido a preços p. Dados os preços p e as dotações ω_i, nos referiremos a escolha ótima do consumidor i como sua correspondência de demanda marshalliana e a denotamos por $x_i(p, p\omega_i)$. Simplificando, todos os consumidores neste modelo devem escolher suas cestas de consumo favoritas em seu conjunto orçamentário.

Descrevemos agora os jogadores e suas ações, mas nenhuma descrição de modelo é completa sem um conceito de solução. Aqui, o conceito de solução será um equilíbrio walrasiano, que especificará um conjunto de preços e uma cesta de consumo para cada consumidor (nos referiremos a uma coleção de cestas de consumo para cada consumidor como uma alocação) que satisfazem duas propriedades: otimização do consumidor e equilíbrio de mercado. Dados os preços p, cada consumidor escolhe otimamente sua cesta de consumo, e a demanda total por cada mercadoria é igual à oferta total.

Definição 8.2.1. *Um equilíbrio walrasiano para uma economia de troca pura ε é um vetor $(p^*, (x_i^*)_{i\in\mathcal{I}})$ que satisfaz:*

1. *Otimização das escolhas do consumidor: para todos os consumidores $i \in \mathcal{I}$,*

$$x_i^* \in \arg\max_{x_i \in \mathcal{B}_i(p^*)} u_i(x_i). \tag{8.2}$$

2. *Market clearing: para todas os bens $\ell \in \mathcal{L}$,*

$$\sum_{i\in\mathcal{I}} x_{\ell,i}^* = \sum_{i\in\mathcal{I}} \omega_{\ell,i}. \tag{8.3}$$

Nós agora especificamos completamente o modelo, mas antes de começarmos a discutir mais detalhadamente as propriedades do equilíbrio walrasiano, há algumas definições mais importantes a serem introduzidas. A primeira é a noção de uma alocação viável, que é apenas uma coleção de cestas de consumo para cada consumidor, para a qual a quantidade total consumida para cada bem não exceda a dotação total desse bem.

Definição 8.2.2. *Uma alocação $(x_i)_{i\in\mathcal{I}} \in \mathbb{R}_+^{L\cdot I}$ é viável se para todo $\ell \in \mathcal{L}$, $\sum_{i\in\mathcal{I}} x_{\ell,i} \leq \sum_{i\in\mathcal{I}} \omega_{\ell,i}$.*

A próxima definição vai descrever como estaremos pensando sobre otimalidade na economia. Para muitos problemas de otimização que você viu em seus outros cursos, a noção apropriada de otimalidade é direta. Por exemplo, se um consumidor tem uma função de utilidade bem definida, é fácil pensar sobre o que é ideal para ela, dado seu orçamento definido. Quando começamos a pensar em ambientes com mais de um consumidor, gostaríamos, em certo sentido, de maximizar múltiplas funções objetivas (ou seja, a utilidade de cada consumidor) simultaneamente. Em geral, não há alocações que simultaneamente maximizem a utilidade de todos os objetivos do consumidor - os consumidores estão tipicamente em conflito uns com os outros - então a noção apropriada de otimalidade não é tão simples. A noção que usaremos é a da otimização de Pareto, o que significa que tudo o que estamos fazendo é excluir as alocações dominadas por outras alocações possíveis.

Definição 8.2.3. *Dada uma economia ε, uma alocação viável $(x_i)_{i\in\mathcal{I}}$ é Pareto ótimo (ou Pareto eficiente) se não houver outra alocação viável $(\widehat{x}_i)_{i\in\mathcal{I}}$ tal que $u_i(\widehat{x}_i) \geq u_i(x_i)$ para todo $i \in \mathcal{I}$ e com estrita desigualdade para algum $i \in \mathcal{I}$.*

Em outras palavras, todas as regras de otimalidade de Pareto são alocações para as quais alguém poderia estar melhor sem prejudicar ninguém. Essa noção de otimalidade é, portanto, omissa quanto a questões de distribuição, já que pode ser ótimo para Pareto que um consumidor consuma tudo na economia e que todos não consumam nada.

Durante as próximas seções, invocaremos diferentes conjuntos de suposições para resultados diferentes. Vou coletar essas suposições aqui e seremos explícito ao se referir a elas quando elas forem necessárias para um resultado.

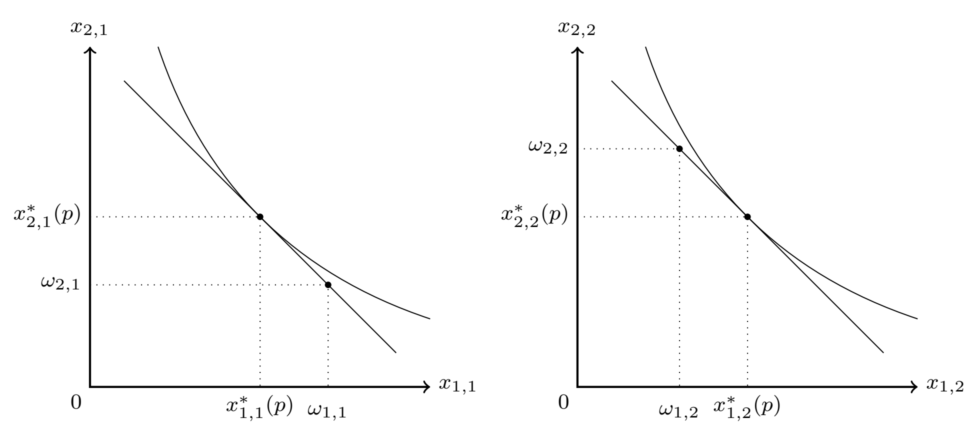
- Continuidade: para todos os consumidores $i \in \mathcal{I}$, u_i é contínua.
- Monotonicidade: para todos os consumidores $i \in \mathcal{I}$, u_i é crescente.
- Concavidade: para todos os consumidores $i \in \mathcal{I}$, u_i é côncava.
- Dotações interiores: para todos os consumidores $i \in \mathcal{I}$, $\omega_{\ell,i} > 0$ para todo $\ell \in \mathcal{L}$.

Os resultados que estaremos estabelecendo nas próximas seções serão mantidos sob premissas mais fracas - por exemplo, monotonicidade pode tipicamente ser relaxado para não-saciedade local, e concavidade pode tipicamente ser relaxado para quasiconcavidade. A última hipótese é uma forte suposição que se mostrará suficiente para descartar alguns casos patológicos em que não exista um equilíbrio walrasiano.

Muitas das principais ideias da teoria do equilíbrio geral podem ser entendidas em uma economia de troca pura de dois consumidores e dois bens. Podemos obter a maioria dos resultados graficamente no que é referido como uma caixa de Edgeworth. Caixas de Edgeworth são densamente informativas, então deixe-me apresentar os elementos constituintes separadamente.

A Figura 8.1 mostra a informação relevante para o consumidor 1. No eixo horizontal está seu consumo do bem 1 e no eixo vertical está seu consumo do bem 2. Sua doação é $\omega_1 = (\omega_{1,1}, \omega_{2,1})$. Aos preços $p = (p_1, p_2)$, ela pode comprar qualquer cesta de consumo no conjunto $\mathcal{B}_1(p)$. A inclinação de sua linha orçamentária é $-\dfrac{p_1}{p_2}$. Dadas suas preferências, que são representadas por sua curva de indiferença, e dados os preços p e a dotação ω_1, o consumidor optou otimamente por consumir $x_1^*(p) = (x_{1,1}^*(p), x_{2,1}^*(p))$. Em outras palavras, a preços p, ela optaria otimamente por vender $\omega_{1,1} - x_{1,1}^*(p)$ unidades do bem 1 em troca de $x_{2,1}^*(p) - \omega_{2,1}$ unidades do bem 2. A Figura 8.1 também mostra as mesmas informações para o consumidor 2.

Figura 8.1 – Problemas de otimização do consumidor

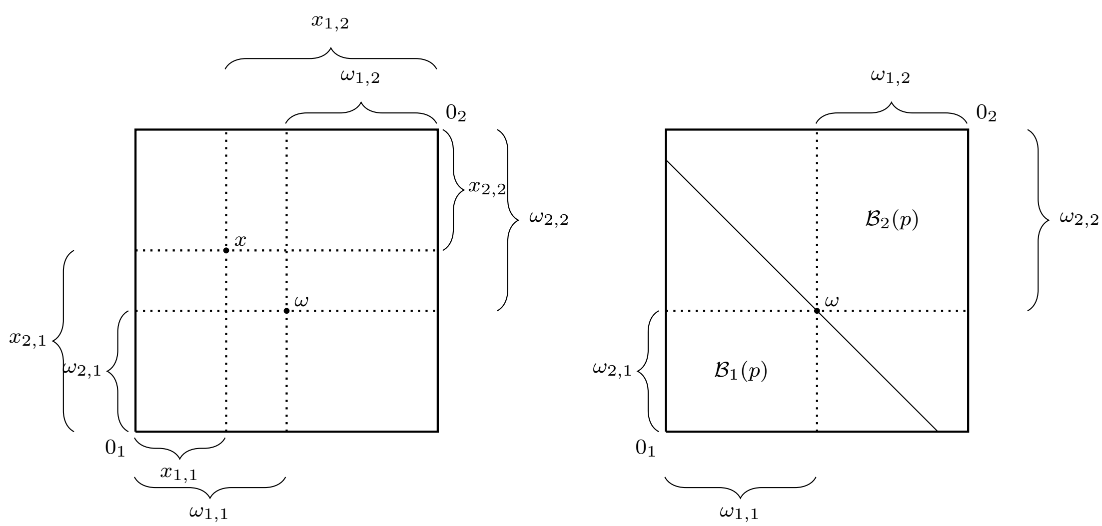


A caixa de Edgeworth representa as dotações dos consumidores e suas escolhas ótimas em função dos preços p, portanto, incorporará todas as informações na Figura 8.1. Para construir esse objetivo, a Figura 8.2 mostra todas as alocações não-perdulárias na economia: alocações $(x_i)_{i\in\{1,2\}}$ para as quais $x_{\ell,1} + x_{\ell,2} = \omega_{\ell,1} + \omega_{\ell,2}$ para todo $\ell \in \{1, 2\}$. O comprimento do eixo horizontal é igual à dotação total da mercadoria 1, e o comprimento do eixo vertical é igual à dotação total de mercadoria 2.

O eixo horizontal, lido da esquerda para a direita, representa o consumo 1 da mercadoria 1, e lido da direita para a esquerda, representa o consumo 2 da mercadoria 1. O eixo vertical, lido de baixo para cima, representa o consumo do consumidor 1 do bem 2 2, e lido de cima para baixo, representa o consumo do consumidor 2 do bem 2. A dotação ω é um ponto na caixa de Edgeworth, e representa uma alocação não-perdulária, uma vez que $\omega_{\ell,1}+\omega_{\ell,2} = \omega_{\ell,1}+\omega_{\ell,2}$ para todo $\ell \in \{1, 2\}$. A alocação x também representa uma alocação não-perdulária, uma vez que $x_{\ell,1}+x_{\ell,2} = \omega_{\ell,1}+\omega_{\ell,2}$ para todo $\ell \in \{1, 2\}$.

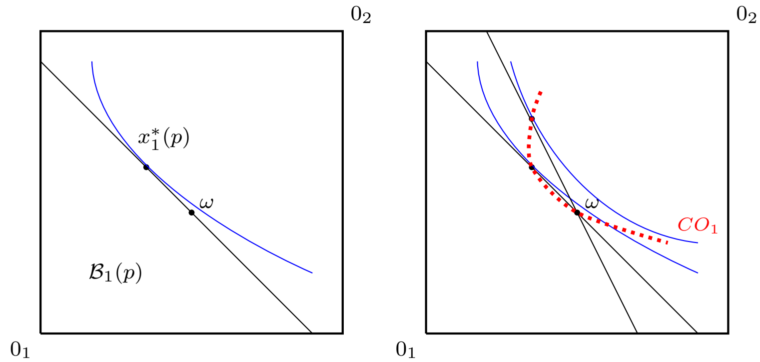
A Figura 8.2 (direita) adiciona preços à imagem e mostra que, dado qualquer vetor de preço p, a caixa de Edgeworth pode ser particionada nos conjuntos de orçamento para os dois consumidores. Dados esses preços, o consumidor 1 pode escolher qualquer cesta de consumo no canto inferior esquerdo da linha diagonal, e o consumidor 2 pode escolher qualquer cesta de consumo no canto superior direito da linha diagonal.

Figura 8.2 – Representação da dotação na caixa de Edgeworth



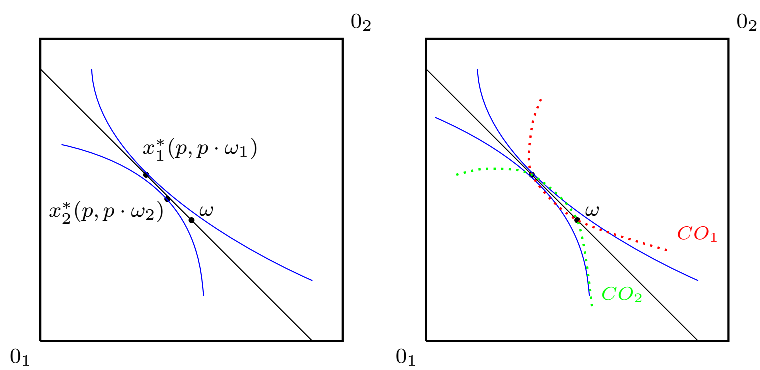
Dados estes preços, a Figura 8.3 (à esquerda) mostra que o consumidor 1 optará otimamente pela cesta $x_1^*(p, p \cdot \omega)$, e a Figura 8.3 (à direita) mostra como $x_1^*(p, p \cdot \omega)$ varia quando a relação de preços varia. Note que, em termos de determinar a escolha ótima do consumidor 1, a relação de preço $\dfrac{p_1}{p_2}$ é uma estatística suficiente para o vetor de preços p. Isso ocorre porque as correspondências de demanda marshallianas são homogêneas de grau zero nos preços (isto é, $x_1^*(p, p \cdot \omega) = x_1^*(\lambda p, \lambda p \cdot \omega)$ para todo $\lambda \in \mathbb{R}_{++}$). A curva traçada na 8.3 (à direita) é referida como curva de oferta do consumidor 1.

Figura 8.3 – Curva de demanda Marshalliana para um p fixo e sua curva de oferta



Lembre-se que em um equilíbrio Walrasiano, $\left(p^*, (x_i^*)_{i\in\{1,2\}}\right)$, o consumidor i escolhe de forma ótima x_i^* dado os preços de equilíbrio p^*. Isso significa que, em qualquer equilíbrio Wwalrasiano, as escolhas ótimas de ambos os consumidores estão em suas curvas de oferta. A Figura 8.4 (à esquerda) representa, para um dado vetor de preços p, as escolhas ótimas de ambos os consumidores. Neste vetor de preços, o consumidor 1 gostaria de vender $\omega_{1,1} - x_{1,1}^*(p, p\cdot\omega_1)$ unidades do bem 1 em troca de $x_{2,1}^*(p, p\cdot\omega_1) - \omega_{2,1}$ unidades do bem 2, e o consumidor 2 gostaria de comprar $x_{1,2}^*(p, p\cdot\omega_2) - \omega_{1,2}$ unidades do bem 1 e vender $\omega_{2,2} - x_{2,2}^*(p, p\cdot\omega_2)$ unidades do bem 2. A alocação associada, $(x_1^*(p, p\cdot\omega_1), x_2^*(p, p\cdot\omega_2))$, não é um equilíbrio Walrasiano, dado que o consumidor 1 gostaria de vender mais unidades do bem 1 do que o consumidor 2 gostaria de comprar, então o mercado para o bem 1 não está em equilíbrio: $\omega_{1,1} - x_{1,1}^*(p, p\cdot\omega_1) > x_{1,2}^*(p, p\cdot\omega_2) - \omega_{1,2}$.

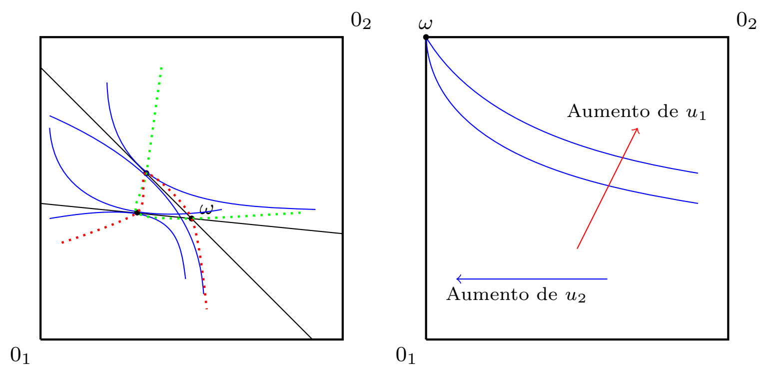
Figura 8.4 – Desequilíbrio e equilíbrio Walrasiano



Deve ficar claro a partir do argumento acima, então, que qualquer alocação de equilíbrio Walrasiano tem que ocorrer em um ponto onde as curvas de oferta de ambos os consumidores se cruzam. A Figura 8.4 (à direita) ilustra tal ponto. O ponto superior esquerdo no qual as duas curvas de oferta se cruzam é uma alocação de equilíbrio Walrasiano, e o vetor de preço que garante que ambos os consumidores escolham de forma otimizada as cestas de consumo associados é um vetor de preço de equilíbrio Walrasiano. Este exemplo também ilustra que, em relação à alocação de equilíbrio Walrasiana, não há outras alocações possíveis que possam melhorar um dos consumidores sem prejudicar o outro consumidor. A alocação de equilíbrio Walrasiana é, portanto, ótima de Pareto. Note que as curvas de oferta também se cruzam no ponto de dotação, mas o ponto de dotação não é uma alocação de equilíbrio Walrasiano neste exemplo em particular.

A caixa Edgeworth também é útil para ilustrar por que pode haver múltiplos equilíbrios walrasianos e por que um equilíbrio walrasiano pode falhar em existir. A Figura 8.5 (à esquerda) ilustra uma situação na qual existem vários equilíbrios walrasianos. Aqui, as duas curvas do consumidor se cruzam várias vezes.

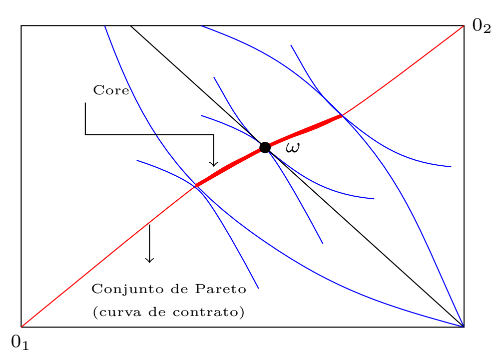
Figura 8.5 – Múltiplos equilíbrio vs. inexistência de equilíbrio



A Figura 8.5 (à dreita) ilustra uma situação em que não há equilíbrio walrasiano. No exemplo, o consumidor 1 não é dotado de unidades do bem 1 e tão somente unidades do bem 2. O consumidor 2 é dotado de todas as unidades do bem 1 e não possui unidades do bem 2. Para qualquer preço p com $p_1 > 0$, o mercado do bem 1 não está equilíbrio, dado que o consumidor 2 sempre escolhe $x^*_{1,2}(p, p{\cdot}\omega_2) = \omega_{1,2}$, e o consumidor 1 sempre escolhe consumir $x^*_{1,1}(p, p{\cdot}\omega) > 0$, a menos que $p_2 = 0$. E se $p_2 = 0$, então $x^*_{2,1}(p, p \cdot \omega_1) = +\infty$ e, portanto, o mercado do bem 2 não estará em equilíbrio. Este exemplo ilustra porque as coisas podem dar errado quando a suposição de dotações interiores não é satisfeita. Esses exemplos nos dizem que as respostas para as duas questões importantes a seguir são "não": (i) há sempre um equilíbrio Walrasiano? (ii) se há um equilíbrio Walrasiano, é único?

Finalmente, a Figura 8.6 mostra que podemos usar a caixa Edgeworth para ilustrar todo o conjunto de alocações ótimas de Pareto.

Figura 8.6 – Curva de contrato e o core



O conjunto de Pareto é o conjunto de todas as alocações viáveis para as quais tornar um consumidor melhor significa necessariamente piorar o consumo do outro. Ele também ilustra a curva de contrato, que é o conjunto de alocações ótimas de Pareto que ambos os jogadores preferem à dotação. Se os dois consumidores negociassem um acordo, considerando suas dotações como opções externas, provavelmente atingiriam um ponto na curva do contrato. As alocações de equilíbrio Walrasianas são tipicamente um pequeno subconjunto da curva de contrato e, em particular, estão no conjunto de Pareto. Esse resultado é conhecido como o primeiro teorema do bem-estar, e agora vamos estabelecer esse resultado.

Nos equilíbrios Walrasianos nos exemplos que acabamos de ver, não há alocações possíveis que melhorem ambos os atores: a alocação de equilíbrio Walrasiana é ótima de Pareto. Acontece que isso é um resultado geral e talvez um dos resultados mais importantes da teoria do equilíbrio geral. Esse resultado é conhecido como o primeiro teorema do bem-estar. Antes de afirmar e provar o primeiro teorema do bem-estar, primeiro estabeleceremos um resultado intermediário conhecido como Lei de Walras, que é uma implicação direta da otimização do consumidor quando as preferências do consumidor são monotônicas (ou mais geralmente, satisfazem a não-saciedade local).

Lema 8.2.1 (Lei de Walras)**.** *Dada uma economia ε e preços p, se a monotonicidade é válida, então*
$p \cdot \left(\sum_{i\in\mathcal{I}} x_i(p, p\cdot\omega_i) = p \cdot \left(\sum_{i\in\mathcal{I}} \omega_i\right)\right).$

Note que a Lei de Walras vale para qualquer conjunto de alocações que sejam ótimas para o consumidor - o resultado não requer que a alocação $(x_i(p, p\cdot\omega_i))_{i\in\mathcal{I}}$ seja uma alocação de equilíbrio Walrasiana.

Agora podemos provar uma versão do primeiro teorema do bem-estar.

Teorema 8.2.1 (Primeiro Teorema do Bem Estar)**.** *Suponha que $\left(p^*, (x_i^*)_{i\in\mathcal{I}}\right)$ é um equilíbrio Walrasiano para a economia ε. Se a monotonicidade é válida, a alocação $(x_i^*)_{i\in\mathcal{I}}$ é Pareto-ótima.*

Demonstração. Suponha que a alocação $(x_i^*)_{i\in\mathcal{I}}$ não é Pareto-ótima. Então há alguma outra alocação $(\widehat{x}_i)_{i\in\mathcal{I}}$ para o qual $u_i(\widehat{x}_i) \geq u_i(x_i^*)$ para todo $i \in \mathcal{I}$ e $u_{i'}(\widehat{x}_{i'}) > u_{i'}(x_{i'}^*)$ para algum i'. Dado que $(x_i^*)_{i\in\mathcal{I}}$ é uma alocação de equilíbrio de Walrasiana, e as preferências dos consumidores satisfazem a hipótese de monotonicidade, por meio da preferência revelada, deve ser o caso em que $p^* \cdot \widehat{x}_i \geq p^* \cdot x_i^*$ para todo $i \in \mathcal{I}$ e $p^* \cdot \widehat{x}_{i'} \geq p^* \cdot x_{i'}^*$. Disso decorre que

$$p^* \cdot \left(\sum_{i\in\mathcal{I}} \widehat{x}_i\right) > p^* \left(\sum_{i\in\mathcal{I}} x_i^*\right) = p^* \left(\sum_{i\in\mathcal{I}} \omega_i\right), \tag{8.4}$$

em que a igualdade é válida pelo Lema acima. Como os preços de equilíbrio p são não-negativos, essa desigualdade implica que há algum bem ℓ tal que $\sum_{i\in\mathcal{I}} \widehat{x}_{i,\ell} > \sum_{i\in\mathcal{I}} \omega_{i,\ell}$, e portanto, $(\widehat{x}_i)_{i\in\mathcal{I}}$ não é uma alocação viável. ■

O primeiro teorema do bem-estar é um resultado notável porque (i) sua conclusão é intelectualmente importante e poderosa, (ii) suas suposições explícitas são simples, e (iii) tem uma prova simples, no sentido de que envolve apenas alguns passos e cada passo é completamente transparente.

Deixe-me comentar um pouco mais sobre cada um desses três pontos. O primeiro teorema do bem-estar fornece uma declaração formal de uma versão do argumento de Adam Smith de que a "mão invisível" dos mercados descentralizados leva os consumidores egoístas a tomar decisões que levam a resultados socialmente eficientes. Apesar de não haver coordenação explícita entre os consumidores, a alocação de equilíbrio resultante é ótima de Pareto. Segundo, a única suposição explícita que fizemos para provar o primeiro teorema do bem-estar foi que os consumidores têm preferências monotônicas e até essa suposição pode ser relaxada. Mas, no fundo, há várias suposições fortes e importantes. Primeiro, assumimos que todos os consumidores enfrentam os mesmos preços entre si para todas os bens. Em segundo lugar, assumimos que todos os consumidores são tomadores de preços e consideram que suas decisões de consumo não afetam esses preços. Terceiro, existem mercados para cada bem e todos os consumidores podem participar livremente em cada mercado. Em quarto lugar, assumimos que cada consumidor se importa apenas com seu próprio consumo e não com o consumo de mais ninguém na economia - portanto descartamos as externalidades. Finalmente, assumimos que há um número finito de bens e consumidores. Quando há um número infinito de bens e consumidores, as alocações de equilíbrio Walrasianas não precisam ser ótimas de Pareto.

O primeiro teorema de bem-estar estabelece que as alocações de equilíbrio Walrasianas são ótimas de Pareto. O segundo teorema do bem-estar, em certo sentido, estabelece um inverso. Ele diz que, sob algumas suposições, qualquer alocação ótima de Pareto pode ser "descentralizada" como uma alocação de equilíbrio Walrasiana, dados os preços e dotações corretos.

Teorema 8.2.2 (Segundo Teorema do Bem Estar)**.** *Suponha um economia ε que satisfaz os axiomas da continuidade, monotonicidade, concavidade e dotações interiores, e $\mathcal{X}$=$\mathbb{R}^L_+$. Se $(\omega_i)_{i\in\mathcal{I}}$ é Pareto-ótimo, então existe um vetor de preços $p \in \mathbb{R}^L_+$ tal que $\left(p, (\omega_i)_{i\in\mathcal{I}}\right)$ é um equilíbrio Walrasiano para ε.*

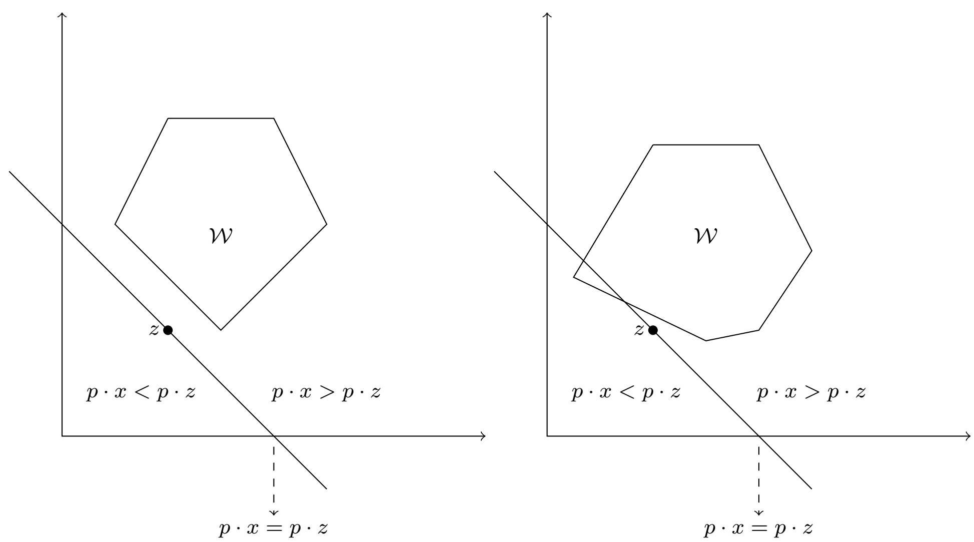
Antes de provar o segundo teorema do bem-estar, vamos apresentar uma versão de um teorema importante na análise convexa, que é usada na etapa chave da prova do segundo teorema do bem-estar social.

Teorema 8.2.3 (Teorema do Hiperplano Separador)**.** *Se $\mathcal{W} \subseteq \mathbb{R}^n$ é um conjunto convexo aberto, e $z \notin \mathcal{W}$ um ponto que não está em $\mathcal{W}$, então existe um ponto $p \neq 0$ tal que $p \cdot x \geq p \cdot z$ para todo $x \in \mathrm{cl}(\mathcal{W})$.*

A Figura 8.7 (à esquerda) ilustra essa versão do teorema do hiperplano de separação em duas dimensões. O conjunto $\mathcal{W}$ é aberto e convexo, e $z \notin \mathcal{W}$. O ponto z está sobre a linha (que é um hiperplano em um espaço bidimensional) caracterizado pela equação $p \cdot x = p \cdot z$ (isto é, p é o vetor normal para a linha). Todos os pontos no canto superior direito dessa linha satisfazem $p \cdot x > p \cdot z$, e todos os pontos à esquerda inferior dessa linha satisfazem $p \cdot x < p \cdot z$. E, em particular, $\mathcal{W}$ está totalmente no canto superior direito desta linha. No caso ilustrado na Figura 8.7 (à esquerda), há, é claro, muitos outros hiperplanos de separação satisfazendo $p \cdot x \geq p \cdot z$ para todo $x \in \mathcal{W}$ correspondendo a linhas inclinadas que atravessam z, mas não intersectam $\mathscr{W}$. A Figura 8.7 (à direita) mostra por que a suposição de que $\mathcal{W}$ é um conjunto convexo é importante para este resultado. Se $z \notin \mathcal{W}$, mas $z \in \text{conv}(\mathcal{W})$[1] então não há um vetor $p \neq 0$ para o qual $p \cdot x \geq p \cdot z$ para todo $x \in \mathcal{W}$.

O hiperplano separa o ponto z do conjunto $\mathcal{W}$. Como isso vale para todos os $z \notin \mathcal{W}$ posso "separar" todos os pontos que não pertencem a $\mathcal{W}$ dos pontos que efetivamente pertencem a $\mathcal{W}$: uma vez excluÌdos os pontos que não pertencem a $\mathcal{W}$ só me restará o conjunto $\mathcal{W}$. Note como isso pode nos ajudar na nossa tarefa de identificar as preferências. Identificar as preferências significa que para toda cesta x consigo construir os conjuntos do tipo $\succeq(x)$, isto é, o conjunto das cestas preferíveis a x.

Figura 8.7 – Teorema do hiperplano separador e convexidade



A idéia do segundo teorema do bem-estar é mostrar que, se a dotação $(\omega_i)_{i\in\mathcal{I}}$ é ótima de Pareto, sempre podemos encontrar um vetor de preços que separe o conjunto de alocações preferidas por todos os consumidores na economia a partir de $(\omega_i)_{i\in\mathcal{I}}$ e, portanto, mostrar que $\left(p, (\omega_i)_{i\in\mathcal{I}}\right)$ é um equilíbrio Walrasiano.

Demonstração. Pela afirmação do teorema, $(\omega_i)_{i\in\mathcal{I}}$ é ótima de Pareto. Vamos definir o conjunto de cestas de consumo agregado que podem ser alocadas de tal forma entre os consumidores para torná-los todos estritamente melhores do que em $(\omega_i)_{i\in\mathcal{I}}$. Para fazer isso, defina o conjunto de cestas de consumo que o consumidor i prefere a ω_i:

$$\mathcal{A}_i = \left\{a \in \mathbb{R}^L : a + \omega_i \geq 0 \text{ e } u_i(a + \omega_i) > u_i(\omega_i)\right\}. \tag{8.5}$$

Dado que u_i é côncava, o conjunto $\mathcal{A}_i$ é convexo. A soma de Minkowski[2] dos conjuntos $\mathcal{A}_i$ também é um conjunto convexo. Isto é, se nós definirmos

[1] O conjunto conv($\mathcal{W}$) é definido como sendo o menor conjunto convexo e fechado contendo $\mathcal{W}$, que é o envoltório convexo de $\mathcal{W}$.

[2] A soma de Minkowski de dois conjuntos $\mathcal{A}$ e $\mathcal{B}$ é o conjunto de vetores x que pode ser escrito como a soma dos vetores $x = a + b$ para os quais $a \in \mathcal{A}$ e $b \in \mathcal{B}$.

$$\mathcal{A} = \sum_{i\in\mathcal{I}} \mathcal{A}_i = \left\{ a \in \mathbb{R}^L \colon \exists a_{\in}\mathcal{A}_1, \ldots, \exists a_I \in \mathcal{A}_I \text{ com } a = \sum_{i\in\mathcal{I}} a_i \right\}, \tag{8.6}$$

então $\mathcal{A}$ é um conjunto convexo. O conjunto $\mathcal{A}$ não contém o vetor 0 porque $(\omega_i)_{i\in\mathcal{I}}$ é Pareto ótimo. Para ver porque este é o caso, note que se $0 \in \mathcal{A}$, então existiria $(a_i)_{i\in\mathcal{I}}$ com $\sum_{i\in\mathcal{I}} = 0$ e $u_i(a_i + \omega_i) > u_i(\omega_i)$ para todo i. Isto é, poderíamos essencialmente realocar a dotação $(\omega_i)_{i\in\mathcal{I}}$ entre os $\mathcal{I}$ consumidores e torná-los todos estritamente melhores, mas isso contradiz a suposição de que $(\omega_i)_{i\in\mathcal{I}}$ é Pareto ótimo.

Note que, pelo teorema do hiperplano separador, existe um vetor de preços $p^* \neq 0$ tal que $p^* \cdot a \geq 0$ para todo $a \in \text{cl}(\mathcal{A})$ ($\mathcal{A}$ é o menor conjunto fechado que contém $\mathcal{A}$). Além disso, temos que $p^*_\ell \geq 0$. Para ver porque, suponha que $p^*_\ell < 0$ para algum ℓ. Tome algum a para o qual a_ℓ é arbitrariamente grande e todos os outros $a_{\ell'}$ são arbitrariamente pequenos, mas positivos. Pela monotonicidade das preferências do consumidor, $a \in \mathcal{A}$, mas se a é escolhido dessa forma, então $p \cdot a < 0$. Nós devemos, portanto, ter $p^* > 0$ (isto é, $p^*_\ell > 0$ para todo $\ell \in \mathcal{L}$ com no mínimo uma desigualdade estrita).

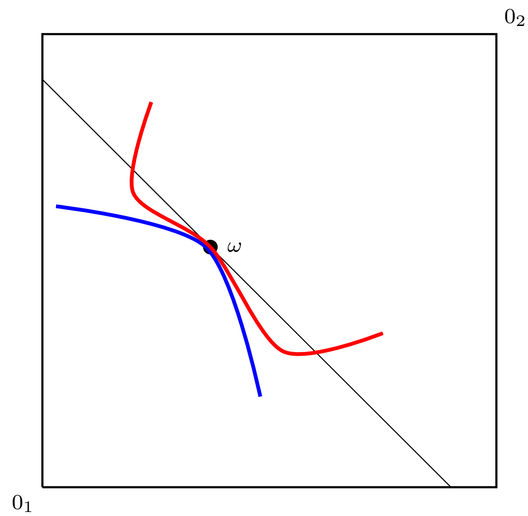
Vamos agora mostrar que $\left(p^*, (\omega_i)_{i\in\mathcal{I}}\right)$ é um equilíbrio Walrasiano. Para fazer isso, precisamos mostrar que, em p^*, os consumidores consomem otimamente suas dotações e os mercados estão ajustados. A segunda condição é imediata. Resta mostrar que, neste p^*, os consumidores consomem otimamente suas dotações. Para fazer isso, suponha que exista algum $\widehat{x}_i \in \mathbb{R}^L_+$ para o qual $u_i(\widehat{x}_i) > u_i(\omega_i)$. Temos que mostrar que $\widehat{x}_i \notin \mathcal{B}_i(p^*)$. Pela definição de $\mathcal{A}$, a alocação $(x_i)_{i\in\mathcal{I}} - (\omega_i)_{i\in\mathcal{I}}$ com $x_i = \widehat{x}_i$ e $x_j = \omega_j$ para todo $j \neq i$ está em $\text{cl}(\mathcal{A})$. Pela definição de p^*, necessariamente temos que ter $p^* \cdot (\widehat{x}_i - \omega) + p^* \cdot \sum_{j\neq i}(\omega_j - \omega_i) \geq 0$, o que implica que $p^*\widehat{x}_i \geq p^*\omega_i > 0$, em que a última desigualdade é verdadeira pois todos os consumidores têm dotações positivas de todos os bens.

Finalmente, temos que mostrar que a últiam desigualdade é estrita. Aqui é que entra a continuidade das preferências. Dado que $u_i(\widehat{x}_i) > u_i(\omega_i)$, isto implica que para $\lambda < 1$, $u_i(\lambda\widehat{x}_i) > u_i(\omega_i)$, o que, por sua vez, implica que $\lambda p^* \cdot \widehat{x}_i \geq p^*\omega_i > 0$. Este não pode ser o caso se $p^* \cdot \widehat{x}_i = p^*\omega_i$ e, então, devemos ter $p^* \cdot \widehat{x}_i > p^*\omega_i$. Portanto, $\widehat{x}_i \in \mathcal{B}_i(p^*)$ – ou seja, qualquer alocação preferida pelo consumidor i à sua dotação é inacessível e, portanto, sua cesta de consumo ótima é a sua dotação. ■

O segundo teorema do bem-estar não mostra que toda alocação ótima de Pareto é um equilíbrio Walrasiano dada uma dotação particular. Em vez disso, diz que se começássemos a partir de uma dotação específica $(\omega_i)_{i\in\mathcal{I}}$ e uma alocação $(x_i)_{i\in\mathcal{I}}$ fosse Pareto ótimo, então poderíamos realocar as dotações dos consumidores de tal maneira que $(x_i)_{i\in\mathcal{I}}$ é uma alocação de equilíbrio Walrasiana. A versão do teorema que acabamos de provar executa este exercício usando uma realocação particularmente rígida de dotações (isto é, $(\omega_i)_{i\in\mathcal{I}} = (x_i)_{i\in\mathcal{I}}$). Existem versões do teorema que envolvem a realização de transferências de riqueza em massa em vez de se movimentarem diretamente em torno de dotações. Como é de se esperar, a descentralização de uma alocação específica de Pareto na prática requer potencialmente redistribuição de riqueza em larga escala. Podemos ver o resultado acima mais como o estabelecimento de uma equivalência entre o equilíbrio Walrasiano e as alocações ótimas de Pareto, em vez de um guia prático para determinar como alcançar uma distribuição particular do consumo na sociedade.

Vale a pena lembrar que a convexidade das preferências dos consumidores foi fundamental para estabelecer o resultado de que $\mathcal{A}$ era um conjunto convexo, que, por sua vez, é necessário para usar o teorema do hiperplano separador. A Figura 8.8 mostra um exemplo onde a conclusão do segundo teorema do bem-estar falha se as preferências dos consumidores não forem convexas.

Figura 8.8 – Preferências não-convexas

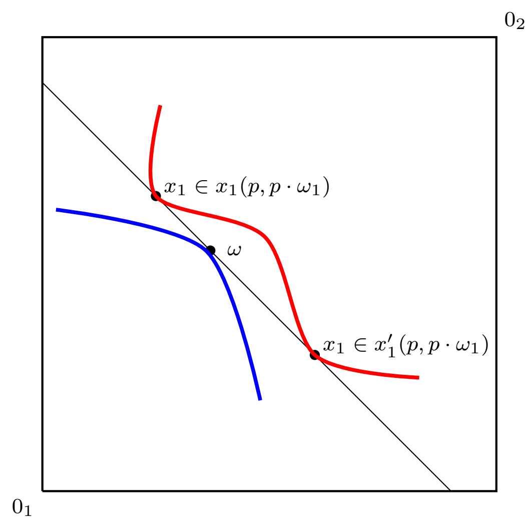


Nessa figura, a dotação é uma alocação ótima de Pareto. Mas não há preços que possam tornar ω_1 o consumo ideal para o consumidor 1. No entanto, uma versão do segundo teorema do bem-estar continua a valer quando os consumidores não têm preferências convexas se você replicar a economia um grande número de vezes. Pense na economia de 2 consumidores como uma metáfora para uma grande economia com dois tipos de consumidores: os consumidores do tipo 1 têm preferências u_1 e dotações ω_1, e os consumidores do tipo 2 têm preferências u_2 e dotações ω_2. Se replicarmos a economia um grande número de vezes, para que haja N consumidores do tipo 1 e N consumidores do tipo 2, em que N é grande, então podemos ver ω como uma alocação de equilíbrio walrasiana, pelo menos em média. Este resultado decorre de uma aplicação do lema de Shapley-Folkman, que diz aproximadamente que a média de conjuntos de Minkowski converge para a envoltória convexa daquele conjunto. Você não precisa saber a matemática por trás desse resultado, mas é um resultado útil de se estar ciente.

A Figura 8.9 ilustra uma economia de replicação para a economia descrita na Figura 8.8. Ela mostra que pode haver um p para o qual existe um $x_1 \in x_1(p, p\cdot\omega_1)$ e um $x_1' \in x_1(p, p\cdot\omega_1)$, de modo que, se alocarmos uma fração λ de consumidores do tipo 1 para consumir x_1 e uma fração $1-\lambda$ de consumidores do tipo 1 para consumir x_1', em média eles estão consumindo $\omega_1 = \lambda x_1 + (1-\lambda)x_1'$.

Esta figura ilustra a ideia de que grandes números "convexizam" a economia. Há um tema recorrente em toda a teoria do equilíbrio geral que muitas das patologias que surgem parecem desaparecer em economias suficientemente grandes. As não-convexidades parecem esotéricas, já que geralmente pensamos nas preferências dos consumidores como tendo utilidade marginal decrescente e os consumidores preferem variedade. As não-convexidades tornam-se especialmente relevantes quando pensamos em empresas. Quando há custos fixos, por exemplo, temos não-convexidades, já que o conjunto de níveis de produção pode incluir fechamento e/ou produção, mas em uma escala muito maior.

Figura 8.9 – Ecomomia replicada



Finalmente, fizemos uso do pressuposto de dotações interiores de uma maneira um pouco opaca na prova. O que essa suposição descarta é um caso como o ilustrado na Figura 8.5 (à direita), no qual não havia equilíbrio Walrasiano. A falha da existência de equilíbrio ilustrada na Figura 8.5 (à direita) surge devido a um problema de divisão por zero: a dotação como uma alocação de equilíbrio exigiria que o consumidor 2 comprasse apenas uma quantidade finita de uma mercadoria com um preço zero quando ele tem zero riqueza.

Exemplo 8.2.1. *Considere uma economia com dois bens e dois consumidores, tal que $\omega_1 = (1, 0)$ e $\omega_2 = (0, 1)$. Seja*

$$u_1(x_1) = (x_{1,1})^a (x_{2,1})^{1-a}, \quad a \in (0, 1) \tag{8.7}$$

$$u_2(x_2) = (x_{1,2})^b (x_{2,2})^{1-b}, \quad b \in (0, 1) \tag{8.8}$$

Seja $p_1 = 1$ por normalização, o consumidor 1 resolve

$$\max_{x_1} u_1(x_1) \tag{8.9}$$

$$\textit{sujeito a} \quad x_{1,1} + p_2 x_{2,1} = 1 \tag{8.10}$$

que resulta em

$$x_{1,1}(1,p_2) = a \tag{8.11}$$
$$x_{2,1}(1,p_2) = \frac{1-a}{p_2} \tag{8.12}$$

De forma semelhante,

$$x_{1,2}(1,p_2) = bp_2 \tag{8.13}$$
$$x_{2,2}(1,p_2) = 1-b \tag{8.14}$$

A condição de market clearing é

$$z_1(1,p_2) = x_{1,1}(1,p_2) + x_{1,2}(1,p_2) - 1 = a + bp_2 - 1 = 0 \quad \Longrightarrow \quad p_2 = \frac{1-a}{b} \tag{8.15}$$

Sendo $p_2 = \frac{1-a}{b}$, *o mercado do bem 2 está em equilíbrio, isto é,* $z_2(1,p_2) = 0$, *como é esperado pela lei de Walras.*

Agora, vamos calcular as alocações Pareto eficiente. Par atanto, seja:

$$L = (x_{1,1})^a(x_{2,1})^{1-a} + \delta_2[(x_{1,2})^b(x_{2,2})^{1-b} - \overline{u}_2] + \mu_1[1 - x_{1,1} - x_{1,2}] + \mu_2[1 - x2,1 - x_{2,2}] \tag{8.16}$$

cujas condições de priemira ordem, assumindo soluções interiores, são:

$$a(x_{1,1})^{a-1}(x_{2,1})^{1-a} = \mu_1 = \delta_2[b(x_{1,2})^{b-1}(x_{2,2})^{1-b}] \tag{8.17}$$
$$(1-a)a(x_{1,1})^a(x_{2,1})^{-a} = \mu_2 = \delta_2[(1-b)(x_{1,2})^b(x_{2,2})^{-b}] \tag{8.18}$$
$$x_{1,1} + x_{1,2} = 1 = x2,1 + x_{2,2} \tag{8.19}$$

Dividindo as duas primeira condições de primeira ordem, temos:

$$\frac{a}{1-a}\frac{x_{2,1}}{x_{1,1}} = \frac{b}{1-b}\frac{x_{2,2}}{x_{1,2}} = \frac{b}{1-b}\frac{1-x_{2,1}}{1-x_{1,1}} \tag{8.20}$$

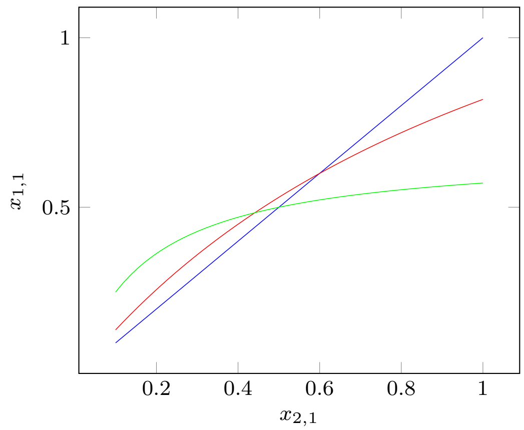
em que a última igualdade deriva da terceira condição de primeira ordem.

Rearranjando,

$$x_{1,1} = \frac{a(1-b)x_{2,1}}{b(1-a) + (a-b)x_{2,1}} \tag{8.21}$$

Assim, o conjunto de alocações Pareto-eficiente (ou curva de contrato) contém todas os pontos na caixa de Edgeworth que satisfazem a equação acima.

Figura 8.10 – Curvas de contrato para diferentes valotes de a e b



8.3 Caracterizando Alocações Pareto-Ótimas

Os teoremas do bem-estar fornecem uma conexão estreita entre o conjunto de alocações ótimas de Pareto e o conjunto de alocações de equilíbrio Walrasianas. Esta seção fornecerá uma breve nota sobre como encontrar alocações ótimas de Pareto em ambientes particularmente bem comportados. Defina o conjunto de possibilidades de utilidade:

$$\mathcal{U} = \left\{(u_1, \ldots, u_I) \in \mathbb{R}^I : \text{ há uma alocação viável } (x_i)_{i \in \mathcal{I}} \text{ com } u_i(x_i) \geq u_i \ \forall i\right\}. \tag{8.22}$$

Se os conjuntos $\mathcal{X}_i$ são conjuntos convexos e as preferências dos consumidores são côncavas, então $\mathcal{U}$ é um conjunto convexo. Quando este é o caso, o problema de encontrar alocações ótimas de Pareto pode ser reduzido ao problema de resolver os problemas de Pareto da forma

$$\max_{u \in \mathcal{U}} \lambda \cdot u, \tag{8.23}$$

para algum vetor de pesos de Paerto $\lambda \geq 0$. A função objetivo deste problema é algumas vezes chamada de função linear de bem-estar social de Bergson-Samuelson. Diremos que u^* é um vetor de utilidade Pareto-ótimo se há uma alocação Pareto-ótima $(x_i)_{i \in \mathcal{I}}$ para o qual $u_i(x_i) = u_i^*$ para todo $i \in \mathcal{I}$. O próximo teorema estabelece esse resultado.

Teorema 8.3.1. *Se u^* for uma solução para o problema de Pareto descrito acima para algum vetor de pesos de Pareto $\lambda \gg 0$, então u^* é um vetor de utilidade Pareto-ótimo. Por outro lado, se o conjunto de possibilidades de utilidade $\mathcal{U}$ é convexo, então qualquer vetor de utilidade Pareto-ótimo u^* é uma solução para o problema de Pareto para algum vetor $\lambda \geq 0$ diferente de zero.*

O teorema mostra que, quando o conjunto de possibilidades de utilidade é um conjunto convexo, o problema de encontrar alocações ótimas de Pareto se resume a resolver uma classe de problemas de Pareto. Se assumirmos ainda que as funções de utilidade dos consumidores são diferenciáveis, então as alocações ótimas de Pareto podem ser caracterizadas por condições de primeira ordem. Por exemplo, suponha que as funções de utilidade sejam diferenciáveis com $\nabla u_i(xi) \gg 0$ para todo x_i, e nós temos uma solução interior, podemos encontrar alocações ótimas de Pareto resolvendo o problema:

$$\max_{(x_i)_{i\in\mathcal{I}}} \sum_{i\in\mathcal{I}} \lambda_i u_i(x_i), \tag{8.24}$$

sujeito a viabilidade para cada bem:

$$\sum_{i\in\mathcal{I}} x_{\ell,i} \leq \sum_{i\in\mathcal{I}} \omega_{\ell,i} \ \forall \ell \in \mathcal{L}. \tag{8.25}$$

Então, usamos o teorema de Kuhn-Tucker para verificar que qualquer alocação Pareto-ótima $(x_i)_{i\in\mathcal{I}}$ com $x_i \gg 0$ para todo $i \in \mathcal{I}$ deve satisfazer

$$\frac{\dfrac{\partial u_i}{\partial x_{\ell,i}}}{\dfrac{\partial u_i}{\partial x_{\ell',i}}} = \frac{\dfrac{\partial u_i'}{\partial x_{\ell,i'}}}{\dfrac{\partial u_i'}{\partial x_{\ell',i'}}} = \frac{\mu_\ell}{\mu_{\ell'}}, \quad \forall i, i', \ell, \ell', \tag{8.26}$$

para $\mu_\ell > 0$ e $\mu_{\ell'} > 0$. Esta condição diz que a taxa marginal de substituição entre quaisquer duas mercadorias deve ser equalizada entre os consumidores em qualquer alocação ótima de Pareto. Se esta condição falha, haveria uma melhoria de Pareto por meio da troca de bens ℓ e ℓ' entre os consumidores i e i'. Os valores de μ_ℓ e $\mu_{\ell'}$ correspondem aos multiplicadores de Lagrange da restrição de viabilidade $\sum_{i\in\mathcal{I}} x_{\ell,i} = \sum_{i\in\mathcal{I}} \omega_{\ell,i}$.

8.4 Existência do Equilíbrio Walrasiano

Nos concentramos até agora nas propriedades normativas e de eficiência do equilíbrio Walrasiano. Agora, vamos nos concentrar em algumas propriedades positivas. Em particular, começaremos por perguntar o que parece ser uma questão direta: existe um equilíbrio Walrasiano? E então faremos algumas outras questões importantes relacionadas à unicidade de equilíbrio, estabilidade de equilíbrio e a estática comparativa do equilíbrio Walrasiano.

A questão de saber se um equilíbrio walrasiano realmente se resume a: em que condições de preferências e dotações existe um equilíbrio walrasiano? Sabemos do exemplo da Figura 8.5 (à direita) que um equilíbrio Wwalrasiano nem sempre existe. E sabemos a partir do segundo teorema do bem-estar que, quando as suposições acima são satisfeitas, então, se a dotação é uma alocação

Pareto-ótima, há um equilíbrio Walrasiano para o qual é a alocação de equilíbrio. Em certo sentido, o segundo teorema do bem-estar fornece uma resposta mundana à questão da existência, uma vez que fornece condições sob as quais nenhum comércio é ideal para cada consumidor. A questão mais interessante é a mais difícil: quando um equilíbrio Walrasiano existe se a própria dotação já não é ótima de Pareto? Ou seja, quando há um equilíbrio Walrasiano que realmente envolve o comércio?

Essa foi uma questão em aberto desde a formulação de Walras do modelo da teoria do equilíbrio geral nos anos 1870, até que Arrow, Debreu e McKenzie produziram as primeiras provas rigorosas de existência nos anos 50. A questão básica é, dadas as funções de demanda agregada $\sum_{i\in\mathcal{I}} x_{\ell,i}(p, p\cdot\omega_i)$ para cada bem, quando existe um vetor de preços p^* tal que $\sum_{i\in\mathcal{I}} x_{\ell,i}(p^*, p^*\cdot\omega_i) = \sum_{i\in\mathcal{I}} \omega_{\ell,i}$ para todo $\ell \in \mathcal{L}$. Os primeiros argumentos equivaliam a contar apenas o número de equações e incógnitas, mas essas abordagens não eram satisfatórias, pois não ficaria claro o que aconteceria se a solução das equações envolvesse preços ou quantidades negativos. O avanço veio na década de 1950, quando Arrow e Debreu (1954) provaram o seguinte resultado de existência.

Teorema 8.4.1 (Existência do Equilíbrio Walrasiano)**.** *Daoo uma economia ε satisfazendo os axiomas da continuidade, monotonicidade, concavidade e dotações interiores, existe um equilíbrio Walrasiano*
$\left(p^*, (x_i^*)_{i\in\mathcal{I}}\right)$.

O principal insight nos anos 1950 foi reformular a questão da existência do equilíbrio Walrasiano como uma questão de ponto fixo, seguindo a prova de John Nash (1951) da existência do equilíbrio de Nash usando uma abordagem relacionada. Um ponto fixo de uma correspondência $f\colon Z \rightrightarrows Z$ é um ponto z tal que $z \in f(z)$, e teoremas de ponto fixo fornecem condições razoavelmente gerais sob as quais funções ou correspondências têm ponto fixo.

O passo importante em fazer uso dos teoremas de ponto fixo é descobrir como mapear a questão da existência de equilíbrio dentro da questão de se uma correspondência adequadamente escolhida tem um ponto fixo: trata-se de escolher a correspondência correta. Suponha que a correspondência f mapeie uma alocação $(x_i)_{i\in\mathcal{I}}$ e um vetor de preços p para uma nova alocação $(x_i')_{i\in\mathcal{I}}$ e vetor de preços p', em que a nova alocação é o conjunto de escolhas ótimas para os consumidores, dado o vetor de preço p, e o novo vetor de preços p' é aquele que eleva os preços dos produtos sobre demandados e reduz o preço dos produtos sub-demandados sob a alocação $(x_i)_{i\in\mathcal{I}}$ e, por outro lado, não altera os preços. Então um ponto fixo de f será um equilíbrio Walrasiano se pudermos mostrar que f satisfaz as condições requeridas para um teorema de ponto fixo se aplicar. Com isso, podemos concluir que existe um equilíbrio Walrasiano.

Primeiro, passaremos por um argumento intuitivo para a existência do equilíbrio no caso especial de dois bens, e então passaremos pelo resultado mais geral descrito no Teorema de Existência do Equilíbrio Walrasiano. O argumento no caso de dois bens também desenvolverá algumas ferramentas que serão úteis quando falamos de unicidade e estabilidade do equilíbrio Walrasiano. Para esta parte, vamos reforçar a condição de monotonicidade para uma condição de monotonicidade forte.

A condição de monotonicidade estrita implica que para todo $i \in \mathcal{I}$, u_i é estritamente crescente: $u_i(x_i') > u_i(x_i)$ se $x_{\ell,i}' \geq x_{\ell,i}$ para todo $\ell \in \mathcal{L}$ com no mínimo uma desigualdade estrita.

Como ponto de partida, vamos introduzir a ideia de uma função excesso de demanda para uma economia $\varepsilon = (u_i, \omega)_{i \in \mathcal{I}}$. A função excesso de demanda para o consumidor i é definida como $z_i(p) = x_i(p, p \cdot \omega_i) - \omega_i$. A função excesso de demanda agregada é definida como $z(p) = \sum_{i \in \mathcal{I}} z_i(p)$. Deve ficar claro, a partir da definição da função excesso de demanda agregada, que se houver um p^* que satisfaça $z(p^*) = 0$, então $(p^*, x_i^*)_{i \in \mathcal{I}}$ com $x_i^* = x_i(p^*, p^* \cdot \omega_i)$ é um equilíbrio Walrasiano. Se $(p^*) = 0$, então o mercado de cada bem está em equilíbrio. Neste caso, a existência de prova resume-se a estabelecer que existe uma solução para $z(p) = 0$, dadas as suposições acima. A função excesso de demanda agregada herda muitas das propriedades das funções de demanda marshallianas, como o próximo lema ilustra.

Lema 8.4.1. *Suponha que ε satisfaz continuidade, monotonicidade forte, concavidade e dotações interiores e que $\mathcal{X}_i = \mathbb{R}_+^L$ para todo i. Então a função excesso de demanda agregada satisfaz:*

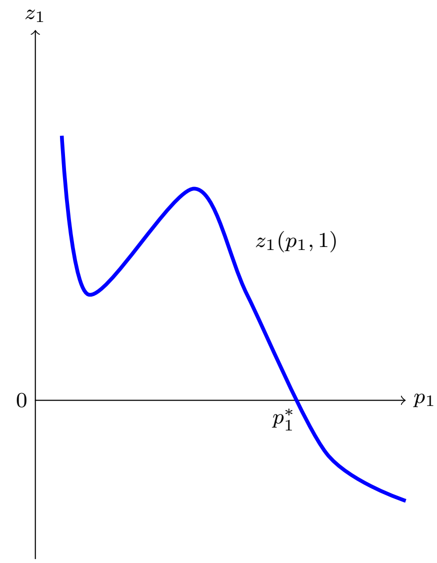
1. *z é contínua*
2. *z é homogênea de grau zero*
3. *$p \cdot (p) = 0$ para todo p (Lei de Walras)*
4. *existe um $Z > 0$ tal que $z_\ell(p) > -Z$ para todo $\ell \in \mathcal{L}$ e para todo p*
5. *se $p^n \to p$, em que $p \neq 0$ e $p_{ell} = 0$ para algum ℓ, então* $\max\{z_1(p^n), \ldots, z_L(p^n)\} \to \infty$.

A última propriedade tem algum comentário. Está dizendo que, como alguns, mas não todos, os preços tendem a zero, deve haver algum consumidor cuja riqueza não seja zero. Como ele tem preferências fortemente monótonas, ele deve exigir mais de um dos bens cujo preço está indo para zero.

Para ganhar intuição para a prova de existência geral, vamos considerar o caso onde há apenas dois bens na economia, e vamos supor ainda que as preferências do consumidor são estritamente côncavas, de modo que $x_i(p, p \cdot \omega_i)$ é uma correspondência única para todo p (vamos permitir que $x_i(p, p \cdot \omega_i)$ seja uma correspondência para que haja múltiplas alocações ótimas para um determinado consumidor em um determinado vetor de preço). Nosso objetivo é encontrar um vetor de preço $p = (p_1, p_2)$ para o qual $z(p) = 0$. Como $z(\cdot)$ é homogênea de grau zero, podemos normalizar um dos preços, digamos p_2, para um. Isso reduz nossa pesquisa para vetores de preço da forma $(p_1; 1)$. Além disso, a Lei de Walras implica que, se o mercado para a mercadoria 1 se equilibra, o mesmo ocorre com o mercado da mercadoria 2, então é suficiente encontrar um preço p_1 tal que $z_1(p_1; 1) = 0$.

O problema de encontrar p_1 tal que $z_1(p_1; 1) = 0$ é um problema unidimensional, então podemos representá-lo graficamente. A Figura 8.11 plota $z_1(p_1; 1)$ em função de p_1. A figura destaca três propriedades importantes de $z_1(p_1; 1)$. Primeiro, a função excesso de demanda é contínua. Segundo, para p_1 muito pequeno, $z_1(p_1; 1) > 0$ e terceiro, para p_1 muito grande, $z_1(p_1; 1) < 0$. Dadas essas três propriedades, pelo teorema do valor intermediário – o mais simples dos pontos fixos teoremas – existe necessariamente algum $p = (p_1; 1)$ tal que $z_1(p) = 0$ e, portanto, existe um equilíbrio Walrasiano. As sutilezas em fazer este argumento estão em estabelecer que $z_1(p_1; 1) > 0$ para p_1 pequeno e $z_1(p_1; 1) < 0$ para p_1 grande. A primeira propriedade segue da condição (5) no lema acima. A segunda propriedade segue porque se $p_1 \to \infty$, então a demanda de cada consumidor pelo bem 1 irá convergir para algo menor do que sua dotação do bem 1, já que a continuidade e a monotonicidade das preferências implicam que ele gostaria de vender pelo menos uma parte do bem 1 para uma quantidade infinitamente grande do bem 2.

Figura 8.11 – Existência do equilíbrio Walrasiano com dois bens



Antes de provar o teorema da existência, primeiro nos lembraremos de alguns importantes teoremas matemáticos que usaremos na prova. O primeiro é o teorema do ponto fixo de Kakutani. O segundo é o teorema do máximo ou o teorema do máximo de Berge.

Teorema 8.4.2 (Teorema do Ponto Fixo de Kakutani)**.** *Suponha que $\mathcal{Z}$ seja um subconjunto não-vazio, compacto e convexo do $\mathbb{R}^n$ e que $f\colon \mathcal{Z} \rightrightarrows \mathcal{Z}$ é uma correspondência não-vazia, convexa e hemi contínua superior. Então f tem um ponto fixo.*

O teorema do ponto fixo de Kakutani é uma generalização do teorema do ponto fixo de Brouwer, mas para funções com valor definido. A idéia básica do teorema é que um ponto fixo é uma interseção do gráfico de f com a linha de 45º, e as condições para o teorema asseguram que o gráfico de f não pode ‘ťsaltar” através da linha de 45º. No caso especial, quando $n = 1$ e quando f é um escalar-valorado, este teorema se resume ao teorema do valor intermediário. A prova da existência de equilíbrio vai fazer uso do teorema do ponto fixo de Kakutani para uma correspondência apropriadamente definida, e precisaremos ser capazes de estabelecer que a correspondência tem as propriedades que são requeridas pelo teorema. O teorema a seguir será útil para estabelecer essas propriedades.

Teorema 8.4.3 (Teorema do Máximo de Berge)**.** *Se $f\colon \mathcal{X} \times \Theta \to \mathbb{R}$ é uma função contínua e $C\colon \Theta \rightrightarrows \mathcal{X}$ é uma correspondência contínua e compacta, então $V(\theta) = \max\{f(x,\theta)\colon x \in C(\theta)\}$ é contínua em θ, e $X^*(\theta) = \arg\max\{f(x,\theta)\colon x \in C(\theta)\}$ é contínua, não-vazia e hemi contínua superior.*

Seja a seguinte notação:

$$x_i^{\mathcal{T}} = \max_{x_i \in \mathcal{B}_i(p) \cap \mathcal{T}_i} u_i(x_i) \tag{8.27}$$

$$a^*(x) = \arg\max_{\widetilde{p} \in \Delta} \widetilde{p} \cdot \left(\sum_{i \in \mathcal{I}} x_i - \sum_{i \in \mathcal{I}} \omega_i \right) \tag{8.28}$$

$$f(x, p) = \left(\prod_{i \in \mathcal{I}} x_i^{\mathcal{T}}(p, p \cdot \omega_i) \right) \times a^*(x) \tag{8.29}$$

em que

1. $x_i^{\mathcal{T}}$ é o conjunto de consumo se o agente i possuísse todas as dotações
2. a^* é o leiloeiro Walrasiano que escolhe o vetor de preços, dentro das opções Δ, que maximiza o valor do excesso de demanda agregada

Suponha que $(x^*, p^*) \in f(x^*, p^*)$, isto é, (x^*, p^*) é um ponto fixo de f. A monotonicidade implica que x_i^* é a escolha que maximiza a utilidade do agente i dentro de seu conjunto orçamentário $\mathcal{B}_i(p^*)$. A lei de Walras garantem que, dado que x_i^* é ótimo, p^* implica que $a^* = 0$. Logo, (x^*, p^*) é um equilíbrio Walrasiano.

Resumindo, ou (x^*, p^*) é um equilíbrio Walrasiano ou a alocação resultante da alocação arbitrária de qualquer mercadoria em oferta excedente para os consumidores (juntamente com o vetor de preço p^*) é um equilíbrio Walrasiano. Em ambos os casos, um equilíbrio Walrasiano.

8.5 Unicidade, Estabilidade e Testabilidade

Agora fornecemos uma introdução a algumas das propriedades positivas mais importantes da teoria do equilíbrio geral. Perguntaremos quando um equilíbrio walrasiano é único, se é estável no sentido em que pode ser alcançado por um simples processo de ajuste de preços, e veremos se o equilíbrio walrasiano impõe restrições substantivas a dados observáveis. Esta seção será menos formal do que as anteriores. Já aludimos às respostas a algumas dessas questões: não, o equilíbrio Walrasiano não precisa ser único, e não, não é o caso de um processo de ajuste de preço simples sempre convergir para um equilíbrio walrasiano. Vamos primeiro estabelecer esses resultados sob preferências gerais. Vamos então focar em uma classe especial de economias nas quais as preferências do consumidor satisfazem a propriedade de substitutos brutos quando essa propriedade é satisfeita, o modelo é particularmente bem comportado: haverá um equilíbrio Walrasiano único, e haverá um simples ajuste de preço processo que sempre converge para ele.

Vamos primeiro olhar para a questão de saber se existe um equilíbrio Walrasiano globalmente único. Lembre-se da seção anterior a definição da função excesso de demanda agregada: $z(p) = \sum_{i \in \mathcal{I}} z_i(p)$, em que $_i(p) = x_i(p, p \cdot \omega_i) - \omega_i$.

Vamos pensar em uma economia com dois bens e dois consumidores e normalizar $p_2 = 1$. Informalmente, um equilíbrio Walrasiano existe se $z_1(p_1, 1)$ é contínua em p_1, $z_1(p_1, 1) > 0$ para p_1

pequeno e $z_1(p_1, 1) < 0$ para p_1 grande. Pelo teorema do valor intermediário, existe um p_1^* tal que $z_1(p_1^*, 1) = 0$ e, portanto, $(p_1*, 1)$ é um vetor de preços de equilíbrio Walrasiano.

Há alguma razão para pensar que há somente um p_1^* no qual $z_1(p_1^*, 1) = 0$? Sim, se $z_1(p_1, 1)$ é decrescente. Mas, certamente, há situações em que $z_1(p_1; 1)$ nem sempre é negativamente inclinada. Existe o caso do bem 1 ser um bem de Giffen, de modo que $x1, i(p, \omega)$ é crescente em p_1 mantendo ω fixo. Mesmo que nenhum bem seja de Giffen, no entanto, $x1, i(p, \omega)$ pode ser crescente em p_1 porque a riqueza do consumidor i pode estar aumentando em p_1, então uma região inclinada para cima de $z_1(p_1; 1)$ não é particularmente implausível.

Suponha que as preferências e dotação dos consumidores sejam:

$$u_1(x_{1,1}, x_{2,1}) = \left(x_{1,1}^{-2} + \left(\frac{12}{37}\right)^3 x_{2,1}^{-2} \right)^{-1/2} \tag{8.30}$$

$$u_2(x_{1,2}, x_{2,2}) = \left(\left(\frac{12}{37}\right)^3 x_{1,2}^{-2} + x_{2,2}^{-2} \right)^{-1/2} \tag{8.31}$$

$$\omega_1 = (1, 0) \tag{8.32}$$

$$\omega_2 = (0, 1) \tag{8.33}$$

Se normalizarmos $p_2 = 1$, a demanda Marshalliana pelo bem 1 é:

$$x_{1,1}((p_1, 1), p_1) = \frac{p_1}{p_1 + \frac{12}{37} p_1^{1/3}} \tag{8.34}$$

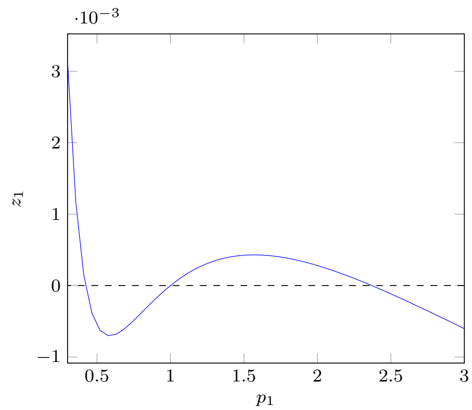
$$x_{1,2}((p_1, 1), p_1) = \frac{1}{p_1 + \frac{37}{12} p_1^{1/3}} \tag{8.35}$$

e o excesso de demanda agregada pelo bem 1 é

$$z_1(p_1, 1) = \frac{1}{p_1 + \frac{37}{12} p_1^{1/3}} - \frac{\frac{12}{37} p_1^{1/3}}{p_1 + \frac{12}{37} p_1^{1/3}} \tag{8.36}$$

A Figura 8.12 apresenta a função $z_1(p_1, 1)$ e mostra que há três soluções para $z_1(p_1, 1) = 0$, a saber: $\left\{ \frac{27}{64}, 1, \frac{64}{27} \right\}$.

Figura 8.12 – MÚLTIPLOS EQUILÍBRIOS WALRASIANOS



Temos que notar o seguinte. Em primeiro, se você perturbar a economia ligeiramente, alterando as preferências ou dotações dos consumidores em uma pequena quantidade, isso não afetaria o fato de que existem três equilíbrios Walrasianos.

O segundo ponto geral que esses exemplos ilustram é que, embora o equilíbrio Walrasiano possa não ser globalmente único, ele pode ser chamado de localmente único no sentido de que não há outro vetor de preço de equilíbrio walrasiano dentro de uma faixa suficientemente pequena em torno do original. Um equilíbrio não é localmente único se o vetor de preço p é o limite de uma sequência de outros vetores de preço de equilíbrio. Isto pode acontecer, mas somente se $z_1(p_1; 1)$ for "horiontal" e igual a zero em algum intervalo de preços $[p_1^*; p_1^{**}]$. O ponto importante a ser observado é que qualquer pequena perturbação de $z_1(\cdot, 1)$ que surgiria de, digamos, uma mudança nas dotações, restauraria a propriedade de haver um número finito de equilíbrios.

Um aspecto importante do equilíbrio walrasiano que aludimos ao longo do curso, mas que ainda não abordamos, é: de onde vêm os preços de equilíbrio walrasianos? A teoria do equilíbrio geral é bastante fraca nos tipos de processos de ajuste de preços que podem levar a resultados de equilíbrio walrasianos.

Walras propôs um processo que ele chamou de "tatonnement", segundo o qual um leiloeiro walrasiano fictício eleva gradualmente o preço dos vens com excesso de demanda e reduz os preços daqueles em excesso de oferta até que os mercados estejam ajustados.

Formalmente, considere o seguinte processo de ajuste de preços em tempo contínuo $p(t)$:

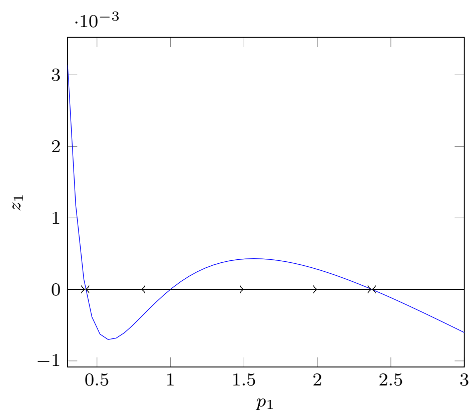
$$\frac{dp(t)}{dt} = \alpha z(p(t)), \tag{8.37}$$

para alguma constante $\alpha > 0$. Dado um vetor de preço inicial $p(0)$, o processo eleva os preços de quaisquer mercadorias ℓ para as quais $z_\ell(p(t)) > 0$ e reduz os preços daqueles para os quais $z_\ell(p(t)) < 0$.

Os pontos estacionários deste processo são os preços p em que $z(p) = 0$, isto é, os preços de equilíbrio walrasianos. Um vetor de preço de equilíbrio p é dito localmente estável se o processo de ajuste de preços convergir para p a partir de qualquer vetor de preço "próximo", e é globalmente estável se o processo convergir para p a partir de qualquer vetor inicial de preço. Esse processo converge para um vetor de preço de equilíbrio walrasiano? Quando há apenas dois bens, e a economia satisfaz as propriedades acima, esse processo de fato converge, como a Figura 8.13 destaca. Aqui, também podemos ver que $p_1 = \frac{27}{64}$ e $p_1 = \frac{64}{27}$ são localmente estáveis, e $p_1 = 1$ não é, e nenhum dos vetores de preço de equilíbrio é globalmente estável.

Esse processo de ajuste de preços nos dá uma maneira de estudar como os preços de equilíbrio podem ser alcançados, mas tem várias desvantagens. Primeiro, o processo em si é um exercício conceitual e não prático – o modelo de equilíbrio geral prevê que ninguém negociará a preços de não-equilíbrio. Segundo, se alguém tentasse implementar esse processo perguntando aos consumidores o quanto eles exigiriam em diferentes níveis de preço, seria improvável que eles quisessem revelar suas demandas com sinceridade. Finalmente, a principal desvantagem desse procedimento é que ele não converge, em geral, para um vetor de preço de equilíbrio. Em um famoso artigo, Scarf (1960) forneceu vários exemplos em que o processo não converge quando há mais de duas mercadorias. Mostraremos na próxima seção, no entanto, que existem classes de economias para as quais ele converge.

Figura 8.13 – Processo de tatonnement para dois bens



8.6 A Estabilidade do Equilíbrio Competitivo: O Processo Walrasiano

Seja x_i o excesso de demanda pelo bem i ($i = 1, 2, \ldots, n+1$). Se D_i e S_i denotam, respectivamente, a demanda de mercado e a oferta de mercado pelo bem i, então, $x_i \equiv D_i - S_i$. Seja P_i o preço do bem i e assuma que x_i depende somente dos preços $P_i's$, em que:

$$x_i = F_i(P_1, P_2, \ldots, P_n, P_{n+1}), \tag{8.38}$$

ou mais compactamente, $x = F(P)$, em que $x = (x_1, \ldots, x_{n+1})$. Para lançar luz à diferença entre estabilidade walrasiana e estabilidade marshalliana, assumimos que as $D_i's$ dependem de P, mas que S_i é constante para cada i[3]. Um equilíbrio competitivo pode então ser definido por

$$F_i(P_1^*, \ldots, P_n^*, P_{n+1}^*) = 0, \tag{8.39}$$

ou mais compactamente, $F(P^*) = 0$. As $n+1$ equações em (8.39) determinam os valores de equilíbrio dos $n+1$ preços, $P_1^*, \ldots, P_{n+1}^*$. Assuma que exista um preço P^* tal que $F(P^*) = 0$.

Por outro lado, a soma ponderada dos preços dos excessos de demanda para todos os bens deve ser idêntico a zero, ou seja,

$$P_1 F_1(P) + \ldots + P_n F_n(P) + P_{n+1} F_{n+1}(P) \equiv 0. \tag{8.40}$$

A equação acima é obtida pela soma de todas as condições orçamentárias de todos os agentes na economia e é chamada Lei de Walras. Então uma das equações em (8.39) torna-se supérflua e, portanto, o número de equações torna-se menor do que o número de variáveis. Entretanto, este problema é evitado ao simplesmente notar que a função excesso de demanda por cada bem é homogênea de grau zero nos preços. Para cada $i = 1, 2, \ldots, n+1$, temos $F_i(\alpha P_1, \ldots, \alpha P_{n+1}) = F_i(P_1, \ldots, P_{n+1})$, $\alpha > 0$. Isto pode ser justificado em termos de racionalidade do comportamento. Em outras palavras, se os $P_i' s$ são mensurados em termos de alguma unidade de medida, então mudanças na unidade de medida não deve afetar o comportamento de cada indivíduo.

Assim, assumindo $p_i \equiv {}^{P_i}/_{P_{n+1}}$, $i = 1, 2, \ldots, n$, podemos definir a função f_i por

$$f_i(p_1, p_2, \ldots, p_n) = F_i(p_1, p_2, \ldots, p_n, 1) = 0. \tag{8.41}$$

O $n+1$-ésimo bem é chamado numerário. O equilíbrio competitivo é então definido em termos de $f_i' s$ como:

$$f_i(p_1^*, p_2^*, \ldots, p_n^*) = 0, \tag{8.42}$$

ou mais compactamente, $f(p^*) = 0$. As n equações em (8.42) determinam os n valores de equilíbrio dos preços $p_1^*, p_2^*, \ldots, p_n^*$. Assumimos que existe um $p^* > 0$ tal que $f(p^*) = 0$, que chamamos de um vetor de preços de equilíbrio (normalizado). Se p não é um vetor de preços de equilíbrio, então $f_i(p) \neq 0$ para algum ou todos os $i' s$. O valor do excesso de demanda para o bem $n+1$-ésimo bem, x_{n+1} é obtido usando (8.40):

$$x_{n+1} = -[p_1 f_1(p) + p_2 f_2(p) + \ldots + p_n f_n(p)]. \tag{8.43}$$

[3] Como Peter Newman (1965, pp. 106–108) apontou, a confusão comum deste ponto reside na falha em discutir claramente a teoria da troca da teoria da produção. Marshall e Walras teorizam sobre a produção bem como sobre a troca pura e ambos concordam que o ajuste marshalliano é explicitamente elaborado para a teoria da produção, enquanto o ajuste de preços walrasiano é mais adequado à teoria da troca.

Note que se p^* é um vetor de preços de equilíbrio, então (8.43) mostra que $x_{n+1} = 0$. Em outras palavras, se os primeiros n mercados estão em equilíbrio, então o $n+1$-ésimo mercado está automaticamente em equilíbrio.

O problema de estabilidade de um equilíbrio competitivo está preocupado com a questão de se o vetor de preços p, quando desviado do ponto de equilíbrio p^*, retornará a p^*. A premisa fundamental para estudar este problema é que um excesso de demanda pelo bem i aumenta o preço do bem i e um excesso de oferta do bem i reduz o preço do bem i. Seguindo a reformulação de Samuelson do problema de estabilidade walrasiano, podemos especificar tal premisa por meio do seguinte sistema de equações diferenciais:

$$\dot{p}_i(t) = k_i f_i[p_1(t), p_2(t), \ldots, p_n(t)], i = 1, 2, \ldots, n, \quad k_i > 0, \tag{8.44}$$

em que $\dot{p}_i$ denota a derivada de p_i em relação ao tempo e k_i é a velocidade de ajustamento do i-ésimo mercado. Assuma que a condição inicial do sistema acima seja especificada por $p = p^0$ quando $t = 0$ e seja a solução indicada como $p(t; p^0)$. Assumimos que $p(t; p^0)$ existe e é única e que p^* para o qual $f(p^*) = 0$ é um ponto de equilíbrio isolado do sistema (8.44). Então, o problema de estabilidade do equilíbrio competitivo pode ser apresentado como:

$$\text{Dado um } p^* \text{ tal que } f(p^*) = 0, \tag{8.45}$$

$$p(t; p^0) \Rightarrow p^* \text{ quando } t \Rightarrow \infty? \tag{8.46}$$

Esta questão torna-se um problema de estabiliade local assintótica (do sistema de equações diferenciais) quando permitimos que p^0 esteja na vizinhança de p^*. Se p^0 não é restrito a vizinhança de p^*, então temos um problema de estabilidade global assintótica.

Para facilitar a exposição, vamos assumir que há dois bens (agrícolas e manufaturados ou exportáveis e não-exportáveis) e o segundo bem é o numerário. Então, assumindo $f_1 = p_1$, temos:

$$\dot{p}(t) = kf[p(t)], \quad p \equiv {}^{p_1}/_{p_2}, \tag{8.47}$$

em que p é um escalar. Assuma que p^* é um preço relativo de equilíbrio em que $f(p^*) = 0$. Com isso, obtemos o seguinte resultado.

Proposição 8.6.1. *Em um mundo com dois bens, p^* é local e assintoticamente estável se e somente se $f'(p^*) < 0$. Se $f'(p^*) < 0$ para qualquer p^* para o qual $f(p^*) = 0$, então p^* é único e estável global e asisntoticamente.*

Observe que $f'(p) < 0$ significa que um aumento no preço do bem 1 vis-à-vis o bem 2, reduz o excesso de demanda pelo bem 1. Se isto é válido para um intervalo relevante de p, então o preço de equilíbrio é único e estável global e assintoticamente.

Para um caso de n bens, temos:

$$\dot{p}(t) = Kf[p(t)], \tag{8.48}$$

em que K é uma matriz diagonal $n \times n$ cujos i-ésimo elementos da diagonal principal são $k_i > 0$. Sendo p^* o ponto de equilíbrio, a aproximação linear do sistema pode ser escrita como

$$\dot{p} = KA(p - p^*), \tag{8.49}$$

em que $A = [a_{ij}]$, $a_{ij} \equiv \dfrac{\partial f_i(p^*)}{\partial p_j}$, $i, j = 1, 2, \ldots, n$. A questão da estabilidade local repousa sobre as propriedades da matriz A, que está intimamente relacionada ao estudo de Hicks sobre a estabilidade (1939, 1946). Ele demonstra que a "estabilidade perfeita" é estabelecida em termos de sinais alternados dos menores principais de A. Samuelson em 1947 mostrou que tal condição não é nem necessária e nem suficiente para a estabilidade do sistema dinâmico (8.49) para o caso de n-bens. Em 1958, Hahn e Negishi e Arrow e Hurwicz mostraram que p^* é um equilíbrio estável global e assintoticamente se $a_{ij} > 0$ para $i \neq j$. Isto implica que p^* é um equilíbrio estável local e assintoticamente para o sistema de equações diferenciais exposto em (8.44).

Em 1959, Arrow, Block e Hurwicz obtiveram um novo resultado: p^* é um equilíbrio estável global e assintoticamente para (8.44) se $f_{ij} \left(\equiv \dfrac{\partial f_i}{\partial p_j}\right) > 0$ para todo p e todo $i \neq j$. Este pressuposto é chamado de substitutibilidade bruta global, enquanto em contraste o pressuposto $a_{ij} \equiv \dfrac{\partial f_i(p^*)}{\partial p_j} > 0$ para $i \neq j$ é chamado de substitubilidade bruta local.

8.7 O Equilíbrio Competitivo para Três Bens

Agra vamos trabalhar com três bens (exportáveis, importáveis e nontradabels). Nós não devemos considerar a aproximação linear do sistema. Ao invés disso, assumindo substitutibilidade bruta, nós devemos provar a estabilidade global do equilíbrio competitivo.

Seja $x_i = F_i(P_1, P_2, P_3)$, $i = 1, 2, 3$, o excesso de demanda pelo bem i. A lei de Walras pode ser escrita como

$$P_1F_1(P) + P_2F_2(P) + P_3F_3(P) \equiv 0, \tag{8.50}$$

em que $P = (P_1, P_2, P_3)$. Usando a homogeneidade de grau zero das funções excesso de demanda, obtemos:

$$F_i(P_1, P_2, P_3) = F_i(p_1, p_2, 1) \quad [\equiv f_i(p_1, p_2)], \quad i = 1, 2, 3, \tag{8.51}$$

em que $p_i \equiv {}^{P_1}/_{P_3}$. A dinâmica de ajustamento pode ser descrito por meio do seguinte sistema de equações diferenciais:

$$\dot{p}_i = k_i F_i(p_1, p_2, 1) \quad [\equiv k_i f_i(p_1, p_2)], \quad i = 1, 2, \tag{8.52}$$

em que $k_i, i = 1, 2$ são constantes maiores que zero, representando a velocidade de ajustamento do i-ésimo mercado. O vetor de preços de equilíbrio competitivo (p_1^*, p_2^*) é definido por:

$$F_i(p_1^*, p_2^*, 1) = 0 \quad \text{ou} \quad f_i(p_1^*, p_2^*) = 0. \tag{8.53}$$

O vetor de preços de equilíbrio competitivo significa um ponto de equilíbrio do sistema de equações diferenciais. Assumimos que existe e é único. Observe que a lei de Walras é válida.

Uma vez que $F_i(P), i = 1, 2, 3$, são homogêneas de garu zero, usando a equação de Euler, temos que:

$$F_{i1}(P)(P_1) + F_{i2}(P)P_2 + F_{i3}(P)P_3 \equiv 0, \quad i = 1, 2, 3, \tag{8.54}$$

em que $F_{ij} = \dfrac{\partial F_i}{\partial P_j}, i, j = 1, 2, 3$. A partir disso, podemos obter:

$$F_{i1}(p_1) + F_{i2}p_2 + F_{i3} \equiv 0, \quad i = 1, 2, 3, \tag{8.55}$$

em que $F_{ij} = F_{ij}(p_1, p_2, 1), i, j = 1, 2$. De (8.55), obtemos:

$$\frac{p_1}{p_2} = -\frac{F_{12}}{F_{11}} - \frac{F_{13}}{F_{11}p_2} \tag{8.56}$$

$$\frac{p_1}{p_2} = -\frac{F_{22}}{F_{21}} - \frac{F_{23}}{F_{21}p_2}. \tag{8.57}$$

Assumimos substitutibilidade bruta global no sentido de que:

$$F_{ij}(p, 1) > 0, \quad \forall p = (p_1, p_2), \quad i \neq j, \quad i, j = 1, 2. \tag{8.58}$$

Isto e (8.55) implicam:

$$F_{ii} < 0, \quad i = 1, 2, 3, \quad \forall p. \tag{8.59}$$

$F_{ij} > 0, i \neq j$ e $F_{ii} < 0$ podem ser interpretados como segue. Um aumento no preço de j desloca a demanda do bem j para outros bens $\left(\dfrac{\partial D_j}{\partial P_j} < 0, \dfrac{\partial D_i}{\partial P_j} > 0, i \neq j\right)$. Em termos da equação de Hicks-Slutsky, $\dfrac{\partial D_i}{\partial P_j} > 0$ exclui a possibilidade do efeito renda ser suficientemente grande, bem como a possibilidade do efeito adverso devido a complementariedade.

A saber, o efeito preço deve ser sempre negativo. Combinando (8.58) e (8.59), chegamos a:

$$\frac{p_1}{p_2} > -\frac{F_{12}}{F_{11}}, \quad \frac{p_1}{p_2} < -\frac{F_{22}}{F_{21}}, \quad \forall p. \tag{8.60}$$

Isso implica que:

$$-\frac{F_{22}}{F_{21}} > -\frac{F_{12}}{F_{11}}, \quad \forall p. \tag{8.61}$$

Assim, temos a seguinte proposição.

Proposição 8.7.1. *Sob o pressuposto de substitutibilidade bruta, o equilíbrio competitivo* (p_1^*, p_2^*) *é estável global e asintoticamente sob o processo dinâmico de ajustamento dado pelo sistema de equações diferenciais.*

Do teorema de Olech, p^* é estável global e asintoticamente se:

1. $k_1F_{11} + k_2F_{22} < 0, \forall p$: traço < 0
2. $k_1k_2(F_{11}F_{22} - F_{12}F_{21}) > 0, \forall p$: determinante > 0
3. ou $F_{11}F_{22} \neq 0, \forall p$ ou $F_{12}F_{21} \neq 0, \forall p$

Toda essa discussão pode ser obtida graficamente por meio do diagrama de fase. Vamos considerar o plano $p_1 \times p_2$. Defina a curva $F_i = 0$ como o locus do par (p_1, p_2) no qual $F_i(p_1, p_2, 1) = 0$ com $i = 1, 2$. Para obter a inclinação da curva $F_i = 0$, diferenciamos $F_i = 0$ para obtermos:

$$F_{i1}dp_1 + F_{i2}dp_2 = 0, \quad i = 1, 2. \tag{8.62}$$

Disso, a inclinação das duas curvas são obtidas como:

$$\left.\frac{dp_1}{dp_2}\right|_{F_1=0} = -\frac{F_{12}}{F_{11}} > 0, \quad \left.\frac{dp_1}{dp_2}\right|_{F_2=0} = -\frac{F_{22}}{F_{21}} > 0. \tag{8.63}$$

Portanto, ambas as curvas têm inclinação positiva no plano (p_1, p_2). Manipulando os resultados encontramos que:

$$\left.\frac{dp_1}{dp_2}\right|_{F_1=0} < \frac{p_1}{p_2}, \quad \left.\frac{dp_1}{dp_2}\right|_{F_2=0} > \frac{p_1}{p_2}. \tag{8.64}$$

Isso significa que a inclinação do segmento a partir da origem para qualquer ponto sobre $F_1 = 0$ é maior do que a inclinação de $F_1 = 0$ neste ponto. Para o caso de $F_2 = 0$ temos o opostp. Isto é ilustrado nas figuras abaixo. A figura da esquerda representa a curva $F_1 = 0$ e a figura da direita a curva $F_2 = 0$.

O equilíbrio competitivo definido por

$$F_i(p_1^*, p_2^*, 1) = 0, \quad i = 1, 2, \tag{8.65}$$

pode ser descrito pela intersecção das curvas $F_1 = 0$ e $F_2 = 0$. Isto é ilustrado na figura abaixo, em que o ponto A representa o equilíbrio competitivo. Note que o segmento $0A$ deve estar entre as curvas $F_1 = 0$ e $F_2 = 0$, ou seja, a curva $F_1 = 0$ deve cortar a curva $F_2 = 0$ exatamente acima do ponto A. Isto estabelece uma prova gráfica da existência de um único equilíbrio competitivo sob o pressuposto de substitubilidade bruta.

Figura 8.14 – Curvas $F_1 = 0$ e $F_2 = 0$

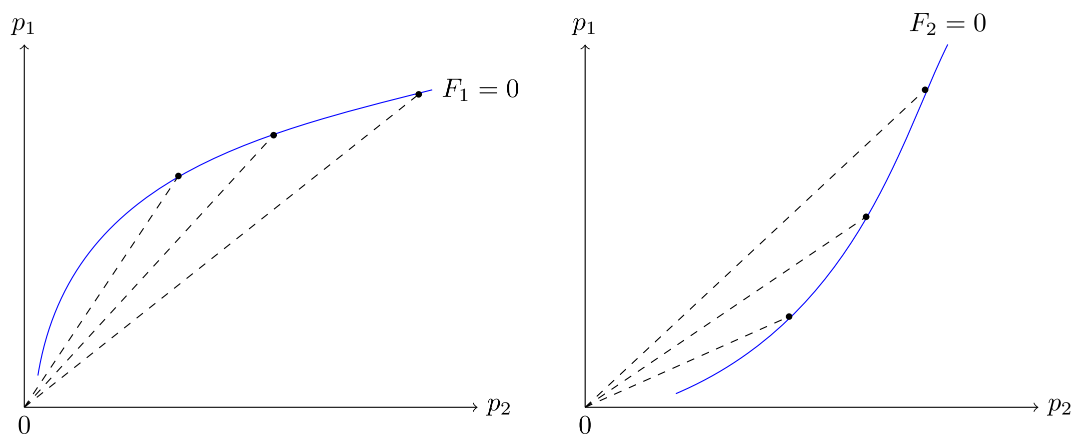
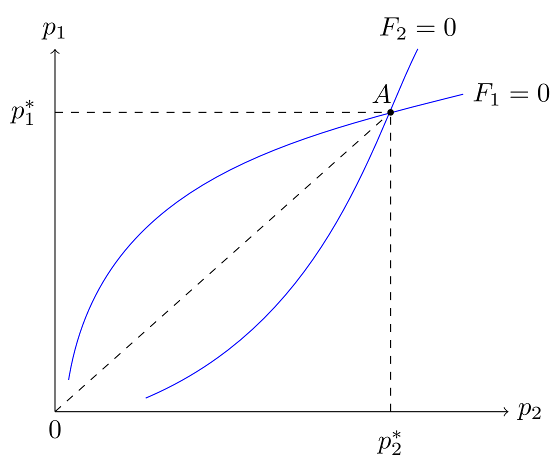
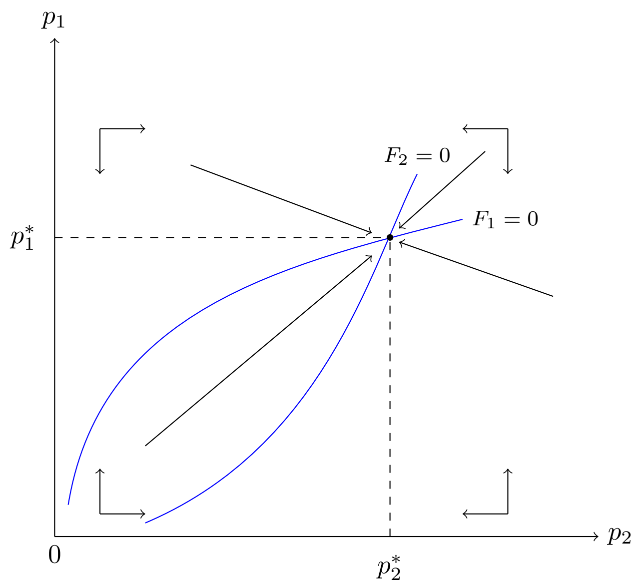


Figura 8.15 – Existência e unicidade do equilíbrio competitivo



Nós podemos agora estabelecer um representação gráfica do comportamento dinâmico de (p_1, p_2), conforme o sistema de equações diferenciais estabelecido acima. Note que as curvas $F_1 = 0$ e $F_2 = 0$ dividem o ortante não-negativo do espaço de (p_1, p_2) dentro de quatro regiões. Uma vez que $F_{12} > 0$, temos que $F_1 > 0$ à direita da curva $F_1 = 0$ e $F_1 < 0$ à esquerda de $F_1 = 0$. Portanto, $\dot{p}_1 > 0$ ou p_1 cresce ao longo do tempo à direita da curva $F_1 = 0$ e $\dot{p}_1 < 0$ ou p_1 decresce ao longo do tempo à esquerda de $F_2 = 0$. De forma similar, uma vez que $F_{22} < 0$, temos que $F_2 < 0$ à direita da curva $F_2 = 0$ e $F_2 > 0$ à esquerda de $F_2 = 0$. Portanto, $\dot{p}_2 < 0$ ou p_2 decresce ao longo do tempo à direita da curva $F_2 = 0$ e $\dot{p}_2 > 0$ ou p_2 cresce ao longo do tempo à esquerda de $F_2 = 0$.

Figura 8.16 – Comportamento dinâmico de (p_1, p_2)



O ponto de equilíbrio A é conhecido como um nó impróprio. Para um sistema não-linear, sabe-se que o comportamento dinâmico da solução das equações diferenciais na vizinhança de um ponto de equilíbrio pode, com algumas exceções, ser aproximado pelo comportamento dinâmico da solução de sua aproximação linear. Este resultado é conhecido como Teorema de Poincaré. A técnica diagramática utilizada aqui é muito útil para obter a representação gráfica do comportamento da solução de sistemas não lineares para o caso de duas dimensões. Esta técnica, conhecida com técnica do diagrama de fase, é amplamente utilizada em economia.

8.8 Empiria

Como acabamos de ver, a existência e a estabilidade de um equilíbrio Walrasiano depende criticamente da estrutura da função excesso de demanda agregada. O que sabemos em geral sobre a estrutura da função excesso de demanda agregada? Nós provamos que sob as suposições de continuidade, concavidade, monotonicidade (forte) e dotações interiores, $z(\cdot)$ é contínua, homogênea de grau zero em p, satisfaz a Lei de Walras e $\lim_{p \to 0} z(p) \to \infty$. Mas, como Sonnenschein (1973) mostrou para o caso de dois bens, e Mantel (1974) e Debreu (1974) mostraram de forma mais geral, a hipótese de maximização do consumidor sozinha não impõe mais restrições a $z(\cdot)$. Este é um resultado muito negativo, pois implica que, mesmo se observarmos uma economia em um equilíbrio walrasiano com o vetor de preço p, é possível que a mesma economia tenha um número arbitrário de equilíbrios walrasianos com propriedades de estabilidade arbitrárias.

Mas para garantir unicidade é necessário algo mais. Será possível delimitar ainda mais a classe de funções? Pode-se responder a essa dúvida invertendo a questão. Será possível encontrar um conjunto de fundamentos para qualquer função pertencente à classe acima? Se isto for verdade, então, uma função que possibilite múltiplos equilíbrios - e, no limite, um número infinito ou arbitrariamente grande de equilíbrios - não seria descartada pelo conjunto de axiomas da teoria.

A resposta a esta última questão é afirmativa: qualquer função de excesso de demanda da classe acima é passível de ser obtida de uma economia ε, isto é, existe uma economia de agentes bem comportados que gera tal função. A sequência de teoremas que garantem esse resultado deve-se a Sonnenschein (1973), Mantel (1974) e Debreu (1974). O teorema responde definitivamente à questão proposta. Observe-se que ele tem o estatuto de um teorema de impossibilidade, ou seja, com os axiomas adotados não se pode descartar a possibilidade de múltiplos equilíbrios. Porém, nada, ainda, pode ser afirmado em relação a quais são as economias que permitem unicidade ou multiplicidade de equilíbrios. Dito de outra forma, não haverá um conjunto de fundamentos - embora arbitrário, é verdade, mas talvez plausível em determinadas circunstâncias - que garanta a unicidade? A questão pode ser discutida investigando-se quais são as hipóteses econômicas necessárias ou pelo menos suficientes para descartar a possibilidade de múltiplos equilíbrios.

A tentativa de investigar os motivos que ocasionam a multiplicidade de equilíbrios foi, de certo modo, infrutífera. É possível decompor a matriz de efeitos preços de diversas formas, mas isto só revela que a multiplicidade está associada tanto às preferências dos agentes quanto às distribuições iniciais da riqueza. É por isso que casos clássicos de unicidade ou fazem hipóteses extremamente fortes, como no exemplo de substituição bruta, ou fazem hipóteses mais fracas sobre preferências e hipóteses adicionais sobre a distribuição - por exemplo, preferências homotéticas e dotações iniciais proporcionais para todos os agentes.

Em resumo: (i) os teoremas de Sonnenschein, Mantel e Debreu demonstram que qualquer função de excesso de demanda agregada que atenda a poucos requisitos tem microfundamentos, no sentido de que é possível encontrar uma economia de troca de agentes bem comportados que gere tal função; (ii) a investigação sobre condições necessárias e suficientes para unicidade demonstra que esta propriedade só pode ser obtida em casos muitos especiais que exigem hipóteses fortes ou ad hoc sobre preferências e/ou distribuição de renda. Logo, o caso geral do modelo é o de múltiplos equilíbrios. Para retratar esse problema, os economistas matemáticos afirmam que "falta estrutura" no sistema de equações. Poderia a introdução da produção fornecer tal "estrutura"? A resposta, no entanto, é negativa. A produção não elimina, embora possa atenuar, o problema da multiplicidade de equilíbrio que advém do lado da demanda (Mas-Colell, 1985:259).

Brown e Matzkin (1996) provam um resultado importante, mostrando que, se um economista é capaz de observar as dotações, bem como os preços, então o modelo walrasiano é, em princípio, testável. Ou seja, há pares de dotação e preço $\left(p, (\omega_i)_{i\in\mathcal{I}}\right)$ e $\left(p', (\omega_i')_{i\in\mathcal{I}}\right)$ tal que se é um vetor de preço de equilíbrio Walrasiano dado um conjunto de consumidores com dotações $(\omega_i)_{i\in\mathcal{I}}$, então, se o mesmo conjunto de consumidores tivesse as dotações $(\omega_i')_{i\in\mathcal{I}}$, p' não poderia ser um vetor de preço de equilíbrio Walrasiano.

Dito de outra forma, o teorema de Brown-Matzkin afirma que os dados de mercado são consistentes com o modelo walrasiano se e somente se as desigualdades de equilíbrio walrasianas são solucionáveis para os níveis de utilidade não observados, utilidades marginais da renda e demandas individuais.

O primeiro teorema do bem-estar diz que o equilíbrio do mercado livre é a solução para o problema acima. Simplesmente por permitir o comércio irrestrito entre os atores do mercado, a solução de mercado - isto é, o vetor de preços e as escolhas de equilíbrio resultantes - satisfará as três restrições acima.

Este é um resultado importante e não óbvio. Implica que o mercado descentralizado continuamente "resolve" um problema de maximização complexo, multi-pessoas, multi-bens que seria difícil para qualquer indivíduo (ou grande agência governamental) resolver por si só devido aos requisitos de informação

Enquanto as suposições acima forem cumpridas, existirá um equilíbrio competitivo simplesmente porque cada pessoa está maximizando auto-interessadamente seu próprio bem-estar. O Segundo Teorema do Bem-Estar implica, portanto, que não há um trade-off intrínseco entre equidade e eficiência.

8.9 Equilíbrio em uma Economia com Produção

Assumimos que já J produtores, em que:

1. $\mathcal{J} = \{1, \ldots, J\}$ é o conjunto de produtores.

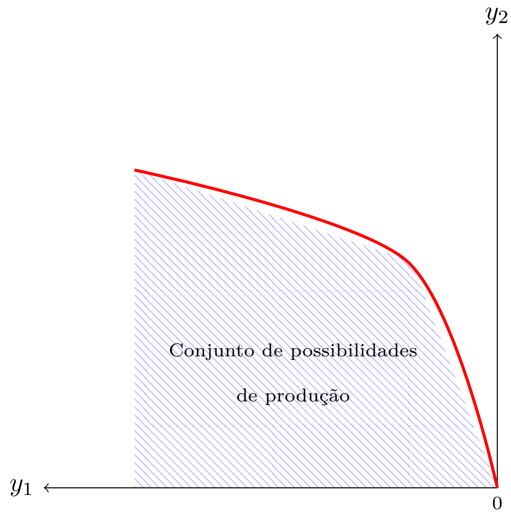
2. $y_{\ell,j}$ é o produto (ou insumo) da firma j para o ℓ-ésimo bem e $y_j = (y_{1,j}, \ldots, y_{\ell,j}) \in \mathbb{R}^L$ é o plano de produção da firma j. Se $y_{\ell,j}$ é positivo o bem ℓ é utilizado como insumo.

3. $Y^j \subset \mathbb{R}^L$ é o conjunto de possibilidade de produção da firma j (plano de produção viável). Qualquer plano de produção $y_j \in Y^j$ é viável. Assumimos que existe uma função $F^j : \mathbb{R}^L \to \mathbb{R}$ tal que $Y^j = \{y \in \mathbb{R}^L | F^j(y) \leq 0\}$. Observe que $\{y \in \mathbb{R}^L | F^j(y) = 0\}$ é a fronteira de possibilidades da firma j. Além disso, definimos o conjunto agregado de possibilidade de produção como $Y \equiv \{y | y = \sum_{j\in\mathcal{J}} y_j,$ em que $y_j \in Y^j, \forall j \in \mathcal{J}\}$.

4. Para qualquer $j \in \mathcal{J}$, temos que:
   (a) $0 \in Y^j$
   (b) Y^j é fechado e limitado
   (c) Y^j é estritamente convexo

A Figura 8.17 retrata o conjunto de possibilidade de produção para um bem e um insumo.

Figura 8.17 – Conjunto de possibilidade de produção



Uma alocação nesta economia denotada por $(x_1, \ldots, x_I, y_1, \ldots, y_J) \in \mathbb{R}_+^{LI} \times \mathbb{R}^{LJ}$ é viável se

$$\sum_{i \in \mathcal{I}} x_{\ell,i} \leq \overline{\omega}_\ell + \sum_{j \in \mathcal{J}} y_{\ell,j}. \tag{8.66}$$

Definição 8.9.1. *Dada uma economia ε, uma alocação viável $(x_i, y_j)_{i \in \mathcal{I}, j \in \mathcal{J}}$ é Pareto ótimo (ou Pareto eficiente) se não houver outra alocação viável $(\widehat{x}_i, \widehat{y}_j)_{i \in \mathcal{I}, j \in \mathcal{J}}$ tal que $u_i(\widehat{x}_i) \geq u_i(x_i)$ para todo $i \in \mathcal{I}$ e com desigualdade estrita para algum $i \in \mathcal{I}$.*

Com os níveis de utilidade fixa, $\overline{u}_2, \ldots, \overline{u}_I$, resolvemos o seguinte problema de maximiação:

$$\max_{(x,y) \in \mathbb{R}_+^{LI} \times \mathbb{R}^{LJ}} u_1(x_{1,1}, \ldots, x_{\ell,1}) \tag{8.67}$$

$$\text{sujeito a} \quad u_i(x_{i,1}, \ldots, x_{i,\ell}) \geq \overline{u}_i \tag{8.68}$$

$$\sum_{i \in \mathcal{I}} x_{\ell,i} \leq \overline{\omega}_\ell + \sum_{j \in \mathcal{J}} y_{\ell,j} \tag{8.69}$$

$$F^j(y_{1,j}, \ldots, y_{\ell,j}) \leq 0 \tag{8.70}$$

As condições de primeira ordem implicam que as taxas marginais de substituição para quaisquer pares de bens devem ser igualadas entre os consumidores, as taxas marginais de transformação para quaisquer pares de bens devem ser igualadas entre os produtores e as taxas marginais de substituição e de transformação devem ser igualadas entre consumidores e produtores.

Dado um vetor de preços $p \gg 0$, cada firma $j \in \mathcal{J}$ procura maximizar os lucros

$$\max_{y_j \in Y^j} p \cdot y_j. \tag{8.71}$$

Assuma que $y_j \equiv \arg\max_{y_j \in Y^j} p \cdot y_j$ e $\pi_j(p) = p \cdot y_j(p)$.

Assumimos que as firmas são de propriedade dos consumidores, de tal forma que $\theta_{ij} \in [0,1]$ é a participação do consumidor i na forma j, isto é, $\sum_{i \in \mathcal{I}} \theta_{ij} = 1, \forall j \in \mathcal{J}$.

Com isso, uma economia ε com produção é caracterizada pelo vetor $(u_i, \omega_i, \theta_{ij}, Y^j)_{i \in \mathcal{I}, j \in \mathcal{J}}$.

Dados os preços p, cada consumidor i tem duas fontes de renda, a saber:

1. Renda das dotações: $p \cdot \omega_i$
2. Participação nos lucros: $\sum_{j \in \mathcal{J}} \theta_{ij} \pi_j(p)$

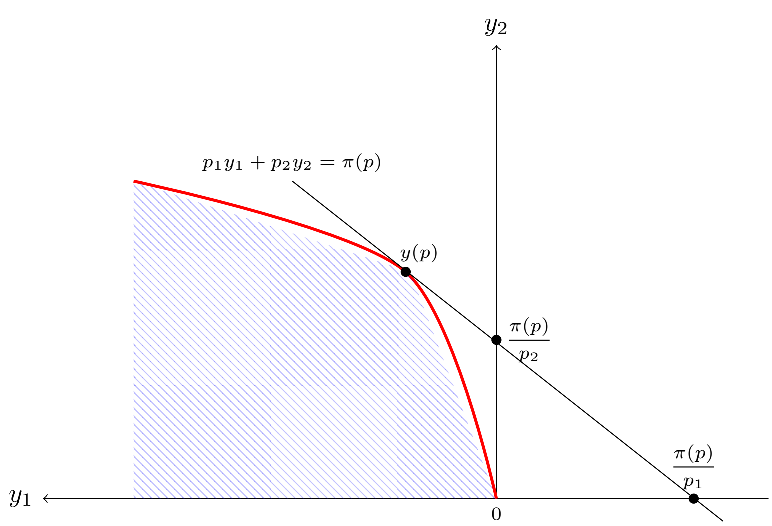
Assim, a renda total do consumidor i é denotada por $m_i(p) \equiv p \cdot \omega_i + \sum_{j \in \mathcal{J}} \theta_{ij} \pi_j(p)$. O consumidor resolve o seguinte problema de maximização da utilidade:

$$\max_{x_i \in \mathbb{R}^L_+} u_i(x_i) \tag{8.72}$$

$$\text{sujeito a} \quad p \cdot x_i \leq m_i(p) \tag{8.73}$$

cuja solução é dada por $x_i(p, m_i(p))$.

Figura 8.18 – Problema de maximização de lucro



Com isso em mente, podemos definir o excesso de demanda agregada como:

$$z_\ell(p) = \sum_{i \in \mathcal{I}} x_{\ell,i}(p, m_i(p)) - \sum_{i \in \mathcal{I}} \omega_{\ell,i} - \sum_{j \in \mathcal{J}} y_{\ell,j}(p), \tag{8.74}$$

$$z(p) = (z_1(p), \ldots, z_\ell(p)). \tag{8.75}$$

Teorema 8.9.1 (Existência do equilíbrio competitivo)**.** *Sob os pressupostos que fundamentam as escolhas do consumidor e do produtor, considere uma economia* $(u_i, \omega_i, \theta_{ij}, Y^j)_{i\in\mathcal{I}, j\in\mathcal{J}}$ *e* $\overline{\omega} + y \gg 0$ *para algum* $y = \sum_{j\in\mathcal{J}} y_j \in Y$*. Então existe no mínimo um vetor de preços* p^* *tal que* $z(p^*) = 0$*.*

Exemplo 8.9.1. *Considere um indivíduo que trabalhe de dia e escolhe entre lazer e consumo no tempo restante. A função de utilidade desse agente é* $U(h, y) = h^{1-\beta}y^{\beta}$ *com* $\beta \in (0, 1)$*, em que* h *denota o tempo de lazer e* y *o consumo. As dotações iniciais de tempo e consumo são, respectivamente,* T *e* 0*. Ele é o proprietário da firma e o único fornecedor de trabalho. A função de produção é* $y = \ell^{\alpha}$ *com* $\alpha \in (0, 1)$*, em que* $\ell \geq 0$ *denota a oferta de trabalho. Sendo* p *e* w *o preço do consumo e do salário, respectivamente, o agente, enquanto produtor, resolve*

$$\max_{\ell\geq 0} p\ell^{\alpha} - w\ell \tag{8.76}$$

e, no papel de consumidor, resolve

$$\max_{h,y} h^{1-\beta}y^{\beta} \tag{8.77}$$

$$\text{sujeito a} \quad py + wh = wT + \pi(w, p) \tag{8.78}$$

Resolvendo o problema do consumidor, obtemos:

$$\ell(w, p) = \left(\frac{\alpha p}{w}\right)^{\dfrac{1}{1-\alpha}} \tag{8.79}$$

$$\pi(w, p) = w\left(\frac{1-\alpha}{\alpha}\right)\left(\frac{\alpha p}{w}\right)^{\dfrac{1}{1-\alpha}} \tag{8.80}$$

Resolvendo o problema do consumidor, obtemos:

$$h(w, p) = \frac{(1-\beta)(wT + \pi(w, p))}{w} \tag{8.81}$$

Usando a normalização $p = 1$*, temos:*

$$\begin{aligned} &\ell(w,1) + h(w,1) = T \\ \left(\frac{\alpha}{w}\right)^{\dfrac{1}{1-\alpha}} + \frac{(1-\beta)(wT + \pi(w,1))}{w} &= T \quad \Longrightarrow \quad w = \alpha\left(\frac{\alpha\beta + (1-\beta)}{\alpha\beta}\right)^{1-\alpha} T^{1-\alpha} \end{aligned} \tag{8.82}$$

Teorema 8.9.2 (Primeiro Teorema do Bem-Estar com Produção)**.** *Se cada* u_i *é estritamente crescente sobre o* $\mathbb{R}^L_+$*, então toda alocação de equilíbrio competivo deve ser Pareto eficiente.*

Demonstração. Suponha que (x, y) é um equilíbrio competitivo, mas não é Pareto eficiente. Então, $\exists(\widehat{x}, \widehat{y}) \in F(\omega, Y)$ tal que $u_i(\widehat{x}_i) > u_i(x_i), \forall i \in \mathcal{I}$ com no mínimo uma desigualdade estrita, o que implica que assumindo que p^* denota o vetor de preços de equilíbrio competitivo, devemos ter

$$\begin{aligned} p^* \widehat{x}_i &\geq p^* x_i \\ p^* \sum_{i \in \mathcal{I}} \widehat{x}_i &> p^* \sum_{i \in \mathcal{I}} x_i \\ p^* \left(\sum_{i \in \mathcal{I}} \omega_i + \sum_{j \in \mathcal{J}} \widehat{y}_j \right) &> p^* \left(\sum_{i \in \mathcal{I}} \omega_i + \sum_{j \in \mathcal{J}} y_j \right) \\ p^* \sum_{j \in \mathcal{J}} \widehat{y}_j &> p^* \sum_{j \in \mathcal{J}} y_j \end{aligned} \tag{8.83}$$

o que implica que $p^* \widehat{y}_j > p^* y_j$ para algum $j \in \mathcal{J}$, contrariando a maximização do lucro. ■

Teorema 8.9.3 (Segundo Teorema do Bem-Estar com Produção)**.** *Suponha que (i) os axiomas acerca do consumidor e do produtor são satisfeitos, (ii)* $y + \sum_{i \in \mathcal{I}} \omega_i \gg 0$ *para algum* $y \in Y$*, e (iii)* (x, y) *é Pareto-eficiente. Então,* $\exists T_1, \ldots, T_I$ *com* $\sum_{i \in \mathcal{I}} T_i = 0$ *e um* p^* *tal que*

1. $x_i \in \arg\max_{x_i \in \mathbb{R}^I_+} u_i(x_i)$ *tal que* $p^* x_i < m_i(p^*) + T_i$, $\forall i \in \mathcal{I}$
2. $y_j \in \arg\max_{y_j \in Y^j} p^* y_j$, $\forall j \in \mathcal{J}$

Podemos pensar no problema do Equilíbrio Geral como uma maximização da utilidade sujeita a três restrições:

1. Nenhum agente está em pior situação no equilíbrio de mercado do que na alocação inicial. Isso sempre será válido porque um agente sempre pode se recusar a negociar e consumir sua dotação original. Assim, nenhuma das partes pode ser prejudicada pelo comércio em relação a sua dotação inicial.

2. Em equilíbrio, nenhuma das partes pode ficar melhor sem prejudicar outra parte (de outro modo, há ganhos não esgotados do comércio).

3. Não há excesso de oferta. Isso não é verdadeiramente uma restrição - é simplesmente uma propriedade de qualquer equilíbrio, que decorre da não-saciedade.

8.10 Implicações

A maior parte do trabalho empírico usando microdados, particularmente na economia do desenvolvimento, envolve comparações parciais de equilíbrio. Dependendo da magnitude de vários efeitos, as interações de equilíbrio geral podem compensar ou até mesmo reverter as conclusões de equilíbrio parcial. No entanto, a maioria das estratégias empíricas não estima diretamente os efeitos gerais de equilíbrio.

A teoria econômica, no entanto, fornece algumas orientações para avaliar a importância dos efeitos gerais de equilíbrio. Três tipos de efeitos de equilíbrio geral, que geralmente não são estimados em comparações de equilíbrio parcial, são potencialmente importantes. Em primeiro lugar, em resposta a grandes intervenções políticas ou choques, a substituição imperfeita entre fatores e os retornos decrescentes implicam que as produtividades e os preços dos fatores mudarão. Em segundo lugar, as mesmas intervenções políticas ou choques podem levar a respostas tecnológicas endógenas. Terceiro, pode haver efeitos de composição resultantes da substituição de equilíbrio de alguns fatores ou produtos por outros (pelo qual a composição das micro unidades muda de forma diferente em resposta a diferentes tipos de intervenções). A teoria geralmente implica que o primeiro e o terceiro efeitos tenderão a compensar parcialmente ou mesmo reverter os efeitos de equilíbrio direto, enquanto as respostas tecnológicas endógenas poderiam atenuá-las ou amplificá-las.

Como exemplo de mudanças nos preços dos fatores, considere o problema de estimar os retornos da escolaridade. Isso geralmente é feito focalizando um pequeno grupo de indivíduos que são induzidos a permanecer na escola por mais tempo e comparando-os a outros indivíduos no mesmo mercado (enfrentando assim os mesmos preços) que abandonaram a escola. A suposição implícita aqui é que alterar as decisões de escolaridade não gerará mudanças nos preços de mercado. Mas para muitas das questões relevantes para a economia do desenvolvimento, desejamos pensar em contrafactuais nos quais uma grande fração da população adquire mais escolaridade. Nesse caso, não é mais plausível supor que os preços permanecerão necessariamente constantes. A substituição imperfeita entre diferentes níveis de habilidade implicará tipicamente que um aumento no nível de escolaridade de uma fração significativa da população pode reduzir o retorno à escola. Por exemplo, Angrist (1995) mostra que os grandes programas de construção de escolas nos territórios palestinos levaram a uma queda acentuada no prêmio de qualificação.

Como exemplo de respostas tecnológicas endógenas, considere o grande aumento na oferta relativa de trabalhadores com formação universitária nos Estados Unidos a partir do final da década de 1960. Dada a tecnologia, essa mudança na oferta relativa deveria ter reduzido o prêmio da faculdade. Como é sabido, o oposto aconteceu na prática, e o prêmio da faculdade aumentou acentuadamente a partir do final dos anos 70 em diante. Acemoglu (1998) argumenta, por exemplo, que isso foi uma consequência da resposta endógena da tecnologia à relativa abundância de trabalhadores mais qualificados. O mesmo raciocínio implica que, ao avaliar o efeito da abertura comercial, não se pode simplesmente confiar em estimativas de equilíbrio parcial derivadas da variação no nível de firma no acesso a mercados estrangeiros, uma vez que a abertura comercial é uma mudança de equilíbrio geral que também afeta as escolhas tecnológicas.

Como exemplo de efeitos de composição, considere o problema de estimar a importância das imperfeições do mercado de crédito. Banerjee e Duflo (2005) pesquisam uma grande quantidade de evidências de que pequenas e médias empresas em economias menos desenvolvidas têm restrições de crédito e uma extensão do crédito a essas empresas fará com que elas aumentem a produção. Agora, considere o efeito de uma política de expansão de crédito em larga escala para pequenas e médias empresas. Esta política poderia levar a um tipo diferente de efeito de composição do que aquele que opera em equilíbrio parcial. Por exemplo, pode ser que, na estimativa de equilíbrio parcial com foco na variação do nível da firma, constatamos que as empresas com melhor acesso ao crédito se expandiram, mas isso às custas de outras empresas que não tiveram acesso ao crédito (isto é, em parte roubando negócios dos outros). E, no entanto, a mesma resposta não pode ocorrer no equilíbrio geral. Como consequência, quando o crédito adicional se torna disponível para uma grande fração de empresas, a produção total pode não aumentar tanto ou totalmente. Poder-se-ia, portanto, imaginar uma situação em que estimativas de equilíbrio parcial de restrições de crédito relaxantes são grandes, enquanto os efeitos de equilíbrio geral seriam pequenos.

8.11 Appropriation and Efficiency: A Revision of the First Theorem of Welfare Economics

O Primeiro Teorema do Bem-Estar fornece um conjunto de condições suficientes para que um sistema de preços coordene eficientemente a atividade econômica. É um belo resultado, com uma prova incrivelmente simples. Mas a sua dependência das hipóteses de mercados completos e agentes tomadores de preçso contribui para uma falta de ênfase explícita nas questões estratégicas e de incentivo.

O artigo de Makowski e Ostroy (1995) oferece um conjunto alternativo e complementar de condições suficientes para uma coordenação eficiente, que enfatiza a importância da ideia de *full appropriation* em vez do comportamento de tomador de preços. Uma vez que a apropriação é dada ao centro do palco, nossa compreensão das razões para a eficiência econômica se aprofunda. Por exemplo, o ajuste entre a teoria das falhas de mercado (que já enfatiza problemas de apropriabilidade na forma de externalidades) e a teoria do sucesso do mercado se torna mais rígido. Para dar uma segunda ilustração, no Primeiro Teorema, o conjunto de mercados disponíveis deve ser completo; portanto, a inovação de produto não pode ocorrer. Em contraste, em um modelo para atingir a eficiência econômica baseada na apropriação, a atividade inovadora pode ser considerada endógena. A medida em que o inovador pode apropriar-se plenamente das consequências de suas inovações torna-se a questão central, no que se refere à eficiência.

Como isso sugere, mercados completos não serão uma suposição mantida no que segue. Há, no entanto, uma limitação importante: o modelo dos autores não aborda questões associadas ao risco moral e seleção adversa, porém mostram como estes também podem ser considerados como problemas de apropriação.

O primeiro teorema é baseado no modelo walrasiano de coordenação econômica, em que os indivíduos não têm capacidade de alterar os preços. Em contraste, o artigo dos autores é baseado em uma extensão que chamam de modelo de escolha ocupacional. Está relacionado ao design do mecanismo, na medida em que o resultado do mercado pode ser descrito como um mecanismo walrasiano no qual os preços, bem como os produtos comercializados, respondem às escolhas ocupacionais dos indivíduos. Assim, no modelo de escolha ocupacional há:

1. price-making: os indivíduos podem ser capazes de influenciar os preços de mercado pela escolha de suas ocupações; e

2. market-making: os indivíduos determinam o conjunto de mercados disponíveis por sua escolha de ocupações.

Para ilustrar seu funcionamento, a produção em diferentes escalas pode ser modelada como a escolha de diferentes ocupações, de modo que o produtor pode ser capaz de influenciar os preços optando por operar em menor escala. Ou para ilustrar a criação de mercado, diferentes ocupações podem envolver a introdução de novas mercadorias diferentes. No modelo, os mercados são abertos apenas para mercadorias que podem ser realmente fornecidas dadas as escolhas ocupacionais dos indivíduos; daí a escolha das ocupações ter um papel de mercado. Um equilíbrio no modelo é chamado de equilíbrio ocupacional.

O principal resultado identifica as condições sob as quais os equilíbrios ocupacionais serão eficientes, apesar do maior escopo para escolha individual e comportamento de interesse próprio. Mostram que, se as duas condições a seguir forem atendidas, a alocação de recursos será Pareto eficiente:

1. *full appropriation*: o benefício privado de cada indivíduo de qualquer escolha ocupacional coincide com sua contribuição social nessa ocupação; e

2. não complementaridade: uma condição de subaditividade é satisfeita entre as escolhas ocupacionais feitas por diferentes indivíduos.

A condição central, *full appropriation*, representa uma extensão da "lógica de apropriação" de Pigou, subjacente tanto ao sucesso quanto às falhas de mercado. Especificamente, a condição de *full appropriation* requer que os benefícios privados e sociais sejam alinhados. Seu papel é dar aos indivíduos os incentivos corretos em suas escolhas ocupacionais e, portanto, tanto em sua "criação" de preços quanto em sua "criação" de mercado.

Como a criação de mercado é endógena, mesmo com apropriação total, podem ocorrer falhas de coordenação. Os primeiros exemplos desse fenômeno são armadilhas de subinoviação apontadas por Tibor Scitovsky (1954). A condição de não-complementaridade governa esses casos. Este resultado é baseado em condições identificadas por Oliver Hart (1980) e Makowski (1980) como suficientes para inovação eficiente de produto sob competição perfeita. Embora ambos os estudos contenham heurísticas que apontam para a importância da apropriação, seu foco principal é em uma aplicação particular, ao invés de incorporar suas descobertas na economia padrão de bem-estar (ou seja, o Primeiro Teorema).

Em vez de enfatizar os mercados completos e a ideia de tomador de preços, *full appropriation* e não complementaridade descrevem diretamente a estrutura dos payoffs individuais que dão bons incentivos. Tais retornos podem surgir tanto em um mercado denso quanto em um cenário de mercado pequeno. Em ambos os cenários, a concorrência perfeita é um ingrediente essencial para a eficiência: leva a *full appropriation.* Por "concorrência perfeita" queremos dizer não apenas o preço, mas algo mais forte: que os indivíduos realmente enfrentam demandas e ofertas perfeitamente elásticas (PEDS).

A importância dessas condições para a competição perfeita se torna evidente quando estudamos a propriedade de eficiência da competição perfeita. Antes, porém, preciso mencionar alguns problemas relativos à informação assimétrica. Aprendemos nos livros textos que há dois tipos de informação privada: informação oculta e ação oculta. Em outras palavras, existem problemas de adverse selection e moral hazard.

Qual é a vantagem de se redefinir competição perfeita como FA com relação aos problemas de informação assimétrica? Mais uma vez, Ostroy recorre aos antigos economistas, desta vez Leonid Hurwicz, o pai da moderna teoria de mechanism design. Os termos adverse selection e moral hazard ganharam fama nos anos 70 com os artigos seminais de Stiglitz, Oliver Hart e outros. Mas Hurwicz, nos anos 60, dividia os problemas informacionais de modo diferente. Em vez da dicotomia informação oculta versus ação oculta, ele os dividia em problemas de privacy (privacidade) e problemas de delivery (entrega). Problemas de privacidade são aqueles problemas alocativos que resultam do fato de o indivíduo conhecer suas próprias valorações e suas possibilidades de produção melhor que qualquer um. Problemas de entrega são aqueles em que um indivíduo é mais bem informado sobre a qualidade ou as características de um bem que ele comercia numa transação, como a qualidade do seu carro, por exemplo, que pode ser um lemon ou um peach, ou ainda se o manager exercerá esforço alto ou baixo. Assim, problemas de privacidade são um subconjunto dos problemas de informação oculta. Tudo o mais são problemas de entrega, ou seja, problemas de entrega englobam todos os problemas de ação oculta (ou moral hazard) e alguns problemas de informação oculta (ou adverse selection).

O fato é que Ostroy mostrou que a competição perfeita resolve todos os problemas de privacidade (a compatibilidade de incentivos e as restrições de revelação vão ficando mais fracas com a competição), mas não resolve os problemas de entrega, pois há neles deadweight losses irrecuperáveis. Pense nas rendas informacionais que têm de ser pagas em problemas do tipo Principal-Agent. Os problemas típicos de privacidade são aqueles estudados em mechanism design aplicados, por exemplo, à alocação de bens públicos em que a valoração do bem a ser ofertado é informação privada. A área de estudos aqui são os chamados mecanismos de Vickrey-Clarke-Groves.

Quem já estudou a demonstração do primeiro teorema do bem-estar deve ter notado a sua simplicidade e deve ter visto nela apenas uma aplicação matemática do teorema de separação de Minkovski. É uma demonstração essencialmente geométrica. O que há por trás dele? Onde está a economia? Por que, afinal de contas, o equilíbrio é eficiente?

Outro artigo seminal de Ostroy & Makowski foi publicado em 1995 na American Economic Review: "Appropriation and efficiency: a revision of the first theorem of welfare economics". Nesse artigo, eles revelam toda a economia que está por trás do Primeiro Teorema do Bem-Estar, segundo o qual o equilíbrio walrasiano é Pareto-eficiente.

Se os agentes são competitivos, eles não vão se satisfazer com as imposições teóricas de price-taking, market-taking, mercados completos e ausência de comportamento estratégico. Eles vão agir e procurar alternativas que lhes deem mais excedentes do que aqueles que receberam em equilíbrio. É como se o caminho para o equilíbrio ainda não tivesse chegado ao fim apenas porque as hipóteses tradicionais do modelo walrasiano simplesmente fecham as portas pro comportamento competitivo dos agentes. Logo, qual o sentido da eficiência de Pareto de um equilíbrio walrasiano? Se os agentes podem aumentar o excedente total com racionalidade individual, então a alocação walrasiana não é necessariamente eficiente, a não ser que o conceito de Pareto-eficiência seja o tradicional, supondo que os agentes são tomadores de preços e de mercados e estes sejam completos. Os autores então mantêm o conceito tradicional de eficiência para economias walrasianas, mas criam o conceito de eficiência global para o modelo de equilíbrio geral reformulado com a ajuda dos dois pilares do marginalismo: o tradicional pilar do marginalismo da mercadoria e (eis a novidade!) o pilar do marginalismo no indivíduo. Esse modelo reformulado do equilíbrio geral foi denominado por eles de modelo de escolha ocupacional, pois os agentes são livres pra escolher que tipo de agente eles serão, quais serão seus conjuntos de trocas líquidas, suas capacidades tecnológicas etc. Esse processo competitivo é modelado por um jogo de Nash. Cada perfil de ações estabelece uma economia com equilíbrio walrasiano. Porém, as hipóteses do jogo asseguram que o jogo tenha um equilibrio de Nash em estratégias puras e que esse equilíbrio seja Pareto-eficiente. Assim, o equilíbrio de Nash inclui a escolha ocupacional dos agentes e o equilíbrio walrasiano determinado pelas escolhas ocupacionais. As escolhas ocupacionais basicamente definem a economia que os agentes querem ter.

Desde Edgeworth, um dos ingredientes da competição perfeita são os grandes números. Essa ideia só foi formalizada por Lionel McKenzie e Robert Aumann nos anos 60. O grande número de agentes permite que compradores e vendedores tenham outside options, isto é, alternativas com as quais os compradores joguem vendedores contra vendedores e os vendedores compradores contra compradores. Aumann modelou isso com um continuum de agentes.

Nos anos 70 havia, na teoria de equilíbrio geral, uma distinção entre duas estruturas de mercado distintas: thick markets e thin markets, ou seja, mercados espessos e mercados finos. A literatura de equilíbrio geral, em suas hipóteses tradicionais de grande número de agentes, mercados completos, número finito de bens homogêneos e comportamento price-taking e market-taking tem essencialmente sido, muitas vezes sem saber, apenas uma versão simplificada da estrutura de mercados espessos. Ostroy retoma essas estruturas e mostra como a eficiência de Pareto sob competição perfeita é alcançada nas duas.

Os mercados são espessos se há um grande número de compradores e vendedores, um número fixo de mercadorias homogêneas, mercados completos e se os agentes enfrentam ofertas e demandas perfeitamente elásticas. Essa é a ideia de Joan Robinson, exposta em seu famoso artigo dos anos 30 sobre competição perfeita. Joan Robinson foi aluna de Marshall, by the way. Note como, mais ou menos, a estrutura de mercados espessos se aproxima das hipóteses tradicionais. A diferença é que o comportamento price-taking é substituído pela condição de ofertas e demandas perfeitamente elásticas. A primeira condição acima será destacada:

Padronização: grande número de compradores e vendedores, um número fixo de mercadorias homogêneas, mercados completos.

Vamos fugir agora dessas condições. Imagine que os bens sejam heterogêneos, isto é, pode haver diferenciação de mercadorias, competição monopolística etc. Imagine também que os preços são fragmentados e dispersos, que é a crítica de Hayek, ou melhor, um de seus grandes insights sobre a economia de mercado. O que existe na verdade são informações locais de preços. Um padeiro só tem que saber e levar em conta alguns preços. Não tem sentido exigir que ele tenha informação perfeita sobre todos os preços da economia. Hayek mostrou que o sistema de preços, ainda que fragmentado e disperso, transmite as informações de escassez, necessidade e abundância praticamente sem custos. As informações locais de preços são consistentes no sentido de elas serem como peças de um quebra-cabeça que se juntam para formar um vetor de preços completo e de tal maneira que, se esse vetor de preços fosse conhecido de todos, tal conhecimento não alteraria as decisões de troca. A ideia de Hayek da eficiência informacional dos preços é descrita na seguinte condição:

Consistência: as informações locais de preços são consistentes.

Precisamos de mais algumas condições, que servem pra qualquer das estruturas:

Regularidade: os indivíduos podem restringir suas possibilidades de troca e estabelecer limites à capacidade de oferta e podem escolher não participar da economia. NC (non-complementarity): a função de ganhos socais é sub-aditiva, isto é, mudanças coordenadas de ocupação não são melhores que mudanças individuais.

A condição NC tem a ver com a eliminação de convexidades que se manifestam em falhas de coordenação, como no famoso problema de inovação do software-hardware. Os dois são inventados por agentes distintos e um não funciona sem o outro, mas existe falha de coordenação e nenhum dos dois produtos é lançado. A condição NC diz que tal estado de coisas não pode ocorrer. Se ocorrer, haverá ineficiências.

Finalmente posso agora apresentar um diagrama que resume a caixa-preta por trás da eficiência global do equilíbrio ocupacional de Ostroy, que é o equilíbrio de Nash resultante da competição perfeita e que embute o equilíbrio walrasiano para a economia determinada pelo equilíbrio de Nash. Em primeiro lugar, as condições tradicionais de price-taking, market-taking e mercados completos são equivalentes a mercados espessos e regularidade:

O que leva à eficiência global do equilíbrio ocupacional são as condições de FA e NC. Agora, em cada estrutura diferente de mercados (espessos e finos), há condições próprias que implicam FA e NC. Em mercados espessos, as condições de regularidade, padronização e PEDS implicam FA e NC. Em mercados finos, precisamos regularidade, consistência e PEDS.

Mercados espessos: [regularidade, padronização, PEDS] leva a FA e NC, que, por sua vez, leva à eficiência global. Mercados finos: [regularidade, consistência, PEDS] leva a FA e NC leva à eficiência global.

Um aspecto matemático importante da abordagem pigoviana de Ostroy é que o problema alocativo do equilíbrio geral pode ser modelado por uma programação linear em dimensão infinita. O famoso marriage problem formulado por Shapley e Shubik nos anos 60 e que deu origem à moderna literatura de matching (que deu o Nobel a Alvin Roth, mas cujos resultados mais pesados foram demonstrados por Marilda Sotomayor) é tão-somente um caso particular, ou melhor, é da mesma família de problemas.

A CMS de um agente mede o impacto de sua presença sobre os ganhos sociais de troca. Se o problema alocativo é o de maximizar os ganhos sociais de troca, então é evidente que o CMS de um indivíduo é o seu *shadow-price*. O *shadow-price* ou multiplicador de Lagrange é obtido pelo problema dual associado ao problema primal da alocação ótima. A dualidade é um arcabouço matemático com imensa riqueza econômica, técnica infelizmente pouco conhecida em nossas lides e de difícil aprendizado, uma vez que é em dimensão infinita e requer uma forte base de Análise Funcional, Teoria da Medida e Teoria dos Operadores. Se um estudante quiser adquirir essa base, terá que passar por um curso completo de Análise no Rn, um curso de Teoria da Medida e um curso de Topologia Geral, além de ter suficiente maturidade matemática. É uma formação que vai depender exclusivamente da disciplina do aluno. Não é uma trajetória trivial.

É pela dualidade que se deram as ramificações da abordagem de Ostroy em equilíbrio geral, pois a condição de FA é caracterizada precisamente pela solução dual. O uso da programação linear em dimensão infinita e da dualidade em equilíbrio com bens públicos apareceu no artigo, de 1992, "Vickrey-Clarke-Groves mechanisms in continuum economies: characterization and existence", no Journal of Mathematical Economics, mas uma versão mais light aparece em um artigo anterior, de Makowski & Ostroy, de 1987, "Vickrey-Clarke-Groves mechanisms and perfect competition", no Journal of Economic Theory.

Para desenho de mecanismos, o artigo de 1999 de Makowski, Ostroy e Uzi Segal, "Efficient incentive compatible economies are perfectly competitive", no Journal of Economic Theory, e o artigo de Belén Jerez, de 2003, "A dual characterization of incentive efficiency", no Journal of Economic Theory.

O modelo de assignment, um caso mais geral que os modelos de matching, pode ser usado para estudar a competição perfeita em equilíbrio geral. O artigo para isso é o de Gretsky, Ostroy e Zame, de 1999, "Perfect Competition in the Continuous Assignment Model", no Journal of Economic Theory.

No sistema de pagamentos, a abordagem dual em dimensão infinita para determinação dos *shadow-prices* de bancos e a determinação de políticas monetárias ótimas de intradia com base na condição de FA pode ser encontrada no artigo de Peñaloza de 2009, "A duality theory of payment systems", no Journal of Mathematical Economics. A formulação, nesse paper, é uma versão variacional do modelo de assignment em dimensão infinita e mostra como políticas monetárias de intradia personalizadas podem fazer com que os bancos internalizem seus preços-sombra para o fluxo ótimo dos pagamentos interbancários, eliminando o custo de liquidez que existe no sistema de pagamentos em razão da retenção das reservas e outras restrições.

A dualidade na caracterização do equilíbrio competitivo com fricções em modelos de search pode ser encontrada no artigo de Belén Jerez, publicado em 2014 no Journal of Economic Theory, "Competitive equilibrium with search frictions: a general equilibrium approach".

Nos últimos dez anos as ideias de Ostroy sobre competição perfeita, FA e dualidade têm sido aplicadas à teoria de leilões de pacotes e leilões ascendentes, embora o primeiro artigo nessa área seja o de Friedman & Ostroy, de 1995, "Competitivity in auction markets: an experimental and theoretical investigation", no Economic Journal. Há ainda os artigos de Bickshandani & Ostroy, de 2006, "Ascending price Vickrey auction", no Games and Economic Behavior, e um artigo de 2002, também de Bickshandani & Ostroy, "The package assignment model", no Journal of Economic Theory.