

**Introdução à**  
**Economia Industrial e ao Antitruste**  
*Rascunho*

Victor Gomes  
*Universidade de Brasília*

©2020, 2024 Victor Gomes



# Sumário

<b>1</b>	<b>Introdução</b>	<b>9</b>
1.1	Questões Centrais . . . . .	9
1.2	Organização e Metodologia . . . . .	14
<b>2</b>	<b>Consumidores</b>	<b>17</b>
2.1	Preferências e Demanda . . . . .	17
2.2	Elasticidade . . . . .	22
2.2.1	Elasticidade-Preço da Demanda . . . . .	22
2.2.2	Elasticidade-Preço Cruzada . . . . .	25
2.2.3	Elasticidade da Renda . . . . .	27
2.3	Racionalidade . . . . .	28
<b>3</b>	<b>Firmas e Produção</b>	<b>29</b>
3.1	Firmas . . . . .	29
3.1.1	Produtividade e Mudança Técnica . . . . .	34
3.2	Minimização de Custo . . . . .	35
3.2.1	Minimização de Custos . . . . .	37
3.3	Heterogeneidade . . . . .	40
3.4	Medidas de Concentração de Mercado . . . . .	41
<b>4</b>	<b>Monopólio, Concorrência Perfeita e Falhas de Mercado</b>	<b>45</b>
4.1	Monopólio . . . . .	45
4.1.1	Maximização de Lucros . . . . .	46
4.1.2	Monopolista Multi-Produto . . . . .	50
4.2	Concorrência Perfeita . . . . .	52
4.2.1	Maximização de Lucros . . . . .	52
4.2.2	Equilíbrio de Mercado . . . . .	53
4.3	Eficiência . . . . .	54
4.4	Antitruste . . . . .	58
4.5	Informação Imperfeita . . . . .	58
<b>5</b>	<b>Determinação de Preços: Produto Homogêneo</b>	<b>61</b>
5.1	Preliminares . . . . .	61

5.2	Escolha Estratégica de Quantidades: Cournot . . . . .	61
5.2.1	Derivação Geométrica . . . . .	62
5.2.2	Modelo . . . . .	63
5.2.3	Exemplo: Firms Simétricas . . . . .	65
5.2.4	Exemplo: Indústria Naval . . . . .	66
5.3	Escolha Estratégica de Preços: Bertrand . . . . .	67
5.3.1	Duopólio Bertrand . . . . .	67
5.3.2	Paradoxo de Bertrand . . . . .	70
5.4	Modelo de Stackelberg . . . . .	73
<b>6</b>	<b>Estimação da Demanda: Produto Homogêneo</b>	<b>75</b>
6.1	Exemplos . . . . .	76
6.2	Variáveis Instrumentais . . . . .	78
6.2.1	O Problema . . . . .	78
6.2.2	Validade de Instrumentos e Identificação . . . . .	79
6.2.3	Estimador 2SLS . . . . .	80
6.2.4	Uma Formulação com Único Instrumento . . . . .	87
6.2.5	Validade dos Instrumentos . . . . .	88
6.2.6	Limites . . . . .	94
6.3	Exemplos . . . . .	95
6.3.1	Demanda por Galeto . . . . .	95
6.3.2	Demanda por Cigarros . . . . .	96
6.4	GMM . . . . .	100
6.4.1	Estimador do Método dos Momentos . . . . .	100
6.4.2	Equações de Momentos Sobre-identificados . . . . .	101
6.4.3	GMM Eficiente . . . . .	103
6.5	Estimação do Modelo de Cournot Linear . . . . .	104
6.6	Bootstrap . . . . .	105
<b>7</b>	<b>Determinação de Preços: Produto Diferenciado</b>	<b>109</b>
7.1	Demanda por Produtos Diferenciados I . . . . .	109
7.1.1	Introdução ao BLP . . . . .	110
7.2	Sistema de Demanda Multi-Produto: Espaço do Produto . . . . .	114
7.2.1	Modelo AIDS . . . . .	114
7.2.2	Exemplo Mercado de Cerveja . . . . .	117
7.2.3	Problemas com Sistemas Multi-Produto no Espaço dos Produtos	118
7.3	Agentes Heterogêneos: Agregação e Simulação em Estimação . . . . .	119
7.4	Demanda por Produtos Diferenciados II . . . . .	120
7.4.1	História . . . . .	120
7.4.2	Arcabouço Básico: Escolha do Consumidor . . . . .	121
7.4.3	Demanda Logit . . . . .	124
7.4.4	Estimação do Logit . . . . .	127
7.4.5	Modelo Vertical Puro . . . . .	128

7.4.6	Problemas de Estimação com Modelos Simples: Logit e Vertical Puro . . . . .	130
7.4.7	B-Logit: Solução de Berry . . . . .	131
7.4.8	Complementares Estratégicos e Substitutos . . . . .	134
7.4.9	BLP . . . . .	136
7.4.10	Markups . . . . .	137
7.5	Concorrência Monopolística . . . . .	138
7.6	Problemas de Índice de Preços . . . . .	139
7.7	Anexo: Detalhes do Logit . . . . .	145
<b>8</b>	<b>Tecnologia, Entrada e Estrutura de Mercado</b>	<b>149</b>
8.1	Função Custo: Tecnologias . . . . .	150
8.1.1	Economias de Escala . . . . .	150
8.1.2	Economias de Escopo . . . . .	151
8.1.3	Economias de Experiência . . . . .	151
8.2	Tamanho do Mercado e Custo de Entrada . . . . .	151
8.2.1	Barreiras Endógenas . . . . .	153
8.3	Entrada e Dinâmica de Mercado . . . . .	154
8.4	Dinâmica da Estrutura de Mercado: Destruição Criadora e Alocação Eficiente . . . . .	156
8.4.1	Estrutura de Mercado e Produtividade . . . . .	157
<b>9</b>	<b>Discriminação de Preços e Outras Estruturas de Demanda</b>	<b>163</b>
9.1	Introdução . . . . .	163
9.2	Seleção por Indicadores . . . . .	163
9.3	Auto-Seleção . . . . .	166
9.3.1	Discriminação Intertemporal . . . . .	166
9.4	Preço Não-Linear . . . . .	166
9.5	Publicidade . . . . .	167
9.6	Leilões . . . . .	167
9.7	Inteligência Artificial . . . . .	167
<b>10</b>	<b>Fusões Horizontais</b>	<b>169</b>
10.1	História . . . . .	169
10.1.1	Fusões e Aquisições . . . . .	170
10.2	Análise de Fusão . . . . .	170
10.2.1	Definição de Mercado Relevante . . . . .	172
10.2.2	Análise da Concentração . . . . .	173
10.2.3	Efeito Adverso da Fusão sobre Preços . . . . .	174
10.2.4	Entrada . . . . .	174
10.2.5	Eficiências e Falência da Firma . . . . .	175
10.2.6	Questões e Problemas na Análise de Guias . . . . .	177
10.3	UPP . . . . .	179
10.3.1	Metodologia UPP . . . . .	179

10.3.2	Problemas Conceituais e de Medida . . . . .	185
10.4	Reposicionamento de Produtos . . . . .	186
10.5	Efeitos Coordenados . . . . .	188
10.6	Fusão com Produtos Complementares . . . . .	188
<b>11</b>	<b>Mercados Verticais</b>	<b>191</b>
11.1	Relações Verticais . . . . .	191
11.2	Dupla Margem e a Integração Vertical . . . . .	192
11.2.1	Mercado Vertical: Externalidades . . . . .	197
11.2.2	Hold-Up . . . . .	197
11.3	Resenha da Evidência Empírica . . . . .	202
11.3.1	Integração Vertical e Preços de Gasolina . . . . .	204
11.3.2	IV e Ganhos de Eficiência . . . . .	206
11.4	Taxa de Passagem e Poder de Mercado . . . . .	207
<b>12</b>	<b>Produtividade e Função de Produção (*)</b>	<b>209</b>
12.1	Função de Produção . . . . .	209
12.2	Estimação de Função de Produção . . . . .	212
12.2.1	NLLS . . . . .	212
12.2.2	GMM . . . . .	216
12.3	Momentos na Estimação . . . . .	220
12.3.1	ACF . . . . .	220
<b>13</b>	<b>Conluio</b>	<b>223</b>
13.1	Introdução . . . . .	223
13.2	Jogos Repetidos . . . . .	224
13.2.1	Firmas diferentes . . . . .	228
13.3	Demanda Flutuante . . . . .	230
13.3.1	Demanda Cíclica . . . . .	232
13.3.2	Preço Dinâmico no Varejo de Gasolina . . . . .	233
13.4	Informação Assimétrica em Jogos Repetidos . . . . .	236
13.4.1	Conluio e Guerra de Preços . . . . .	239
13.5	Deteção de Conluio . . . . .	241
13.5.1	Licitações . . . . .	241
13.5.2	Quebra do Cartel . . . . .	246
<b>14</b>	<b>Bibliografia</b>	<b>249</b>
14.1	Notas sobre Referências . . . . .	249

*Nota:* O símbolo (\*) indica que a seção é avançada, para alunos de graduação mais avançados ou para alunos de pós-graduação.

# Prefácio

Este texto foi baseado em notas de aulas produzidas ao longo dos anos, mas grande parte do material é apoiado nas notas de aula de Ariel Pakes e Robin S. Lee e nos livros e material de apoio de Luis Cabral (1995, 2017), que foram gentilmente cedidos. Agradeço muito a estes autores pelo material produzido.





# Capítulo 1

## Introdução

Organização industrial é o estudo de comportamento de uma indústria. Indústria segue a definição de um conjunto de empresas em um atividade ou que produzem certo produto. Neste sentido, a análise da organização industrial está preocupada em analisar o funcionamento dos mercados e das indústrias, particularmente na forma como está estabelecida a concorrência entre firmas.

A organização industrial ou economia industrial enfatiza o estudo da estratégia das empresas que são as características de interação entre elas no mercado: concorrência por preço, posicionamento de produtos, propaganda, pesquisa e desenvolvimento de produtos e processos, comportamento de cartéis, entre outros.

### 1.1 Questões Centrais

Cabral (2017, cap. 1) resumiu bem quais são as questões centrais do estudo de economia industrial. Elas são:

1. as firmas possuem poder de mercado?
2. como as firmas adquirem e mantêm poder de mercado?
3. quais são as implicações do poder de mercado?
4. qual o papel da política pública a respeito do poder de mercado?

Todas as questões tratam de *poder de mercado*. Em economia *poder de mercado* significa que uma empresa tem capacidade de determinar preços acima dos custos de forma unilateral. Uma empresa sem poder de mercado tem a habilidade de escolher a relação preço-custo do seu produto ou serviço. Como consequência o poder de mercado pode ser aplicado a qualidade do produto. Isto é, ao invés de subir o preço de um produto a empresa pode reduzir a qualidade do bem. Apesar de poder de mercado não um bom nome para este conceito, ele é amplamente utilizado para significar a capacidade da firma afetar preço ou qualidade de um bem no mercado.

**Firmas possuem poder de mercado?** Esta é uma questão central no estudo dos mercados. Como se sabe diversas empresas são líderes em seus setores e podem possuir a capacidade de mudar o conjunto de escolha dos consumidores por meio de preços ou inovações tecnológicas.

Ao longo do tempo diversos estudos tem tentado mensurar a dimensão do poder de mercado. Assumindo que os custos são proporcionais ao produto, uma boa aproximação do tamanho do poder de mercado poderia ser obtida com dados sobre preços, produção e taxas de lucro. Um estudo aplicando este tipo de metodologia encontrou que o poder de mercado nos EUA seria muito baixo dado que as taxas de lucro observadas seriam igualmente pequenas (Harberger, 1954). Esta evidência é consistente com os preceitos da escola de Chicago: enquanto existir livre entrada em cada indústria, o tamanho do poder de mercado nunca será significativo. Se uma firma determina seus preços consistentemente acima do custo marginal, uma nova firma poderá entrar no mercado e cobrar um preço mais baixo em relação a incubente e tomar um pouco de mercado.

Nem todo economista concorda com a visão da escola de Chicago. Esta discordância pode ser teórica ou empírica. Do ponto de vista empírico, uma aproximação alternativa do valor do custo marginal poderia ser aproximado dividindo o aumento do custo do ano  $t$  para o ano  $t + 1$  pelo aumento do produto no mesmo período. Baseado nesta abordagem um estudo estimou que os preços podem ser até três vezes maior do que o custo marginal (Hall, 1988). Expandindo a metodologia de Hall com a utilização de estimativas de uso de insumos intermediários, alguns estudiosos mostraram que além do lucro e das margens preço-custo serem positivas, elas crescem ao longo do tempo nos EUA (ver De Loecker, Eeckhout, and Unger, 2020). A questão do crescimento diferença preço-custo ela é importante pois afeta o sistema de incentivo das firmas e da alocação de recursos na economia.

As margens preço-custo muitas vezes são mensuradas como *markups*, i.e. margem fixa sobre o custo médio. *Markups* são geralmente considerados ruins ou “um mal.” A questão de bem-estar não é tanto se existem markups ou o quanto os markups estão crescendo. A questão real é o quanto os markups possam ser justificados por meio dos seus impactos sobre bem-estar. Para responder esta questão devemos saber algo sobre como os markups se desenvolvem e são sustentados.<sup>1</sup>

A evidência para indústrias particulares também sugerem que o poder de mercado é significativo. Um exemplo é a indústria de transporte aéreo de passageiros nos EUA. Um relatório do governo americano de 1996 analisou tarifa média em 43 grandes aeroportos. Em 10 destes aeroportos, uma ou poucas empresas mantinham forte controle sobre os slots.<sup>2</sup> O relatório mostrou que na média os consumidores pagavam 31% a mais para voar nestes aeroportos do que nos demais (Cabral, cap. 1, p.4). Ao longo do livro são apresentados outros exemplos de poder de mercado.

<sup>1</sup>Na verdade existem poucos estudos sobre isto. Veja Pakes (2018).

<sup>2</sup>Slots são o acesso ao tempo de pouso e decolagem em determinado aeroporto.

**Como as firmas adquirem e mantêm poder de mercado?** O poder de mercado se traduz em lucros elevados. Portanto, criar e manter o poder de mercado é vital na estratégia de maximização de valor de uma empresa.

Uma possibilidade das firmas adquirirem poder de mercado é pela proteção legal de concorrência. Neste caso, a firma pode determinar preços acima dos custos que não irá atrair um novo competidor para o mercado. Exemplo: nos anos 40 e 50, a Xerox desenvolveu uma tecnologia de fotocopia de folhas de papel e a patenteou. Dada a proteção legal das patentes, a Xerox pode aumentar os preços acima dos custos sem atrair novos competidores.

A estratégia adotada por uma firma pode ter papel significativo no estabelecimento do poder de mercado. Um exemplo foi a ação tomada pela Samsung Eletronics na tentativa de entrar no lucrativo mercado de smartphones. Para entrar a Samsung vendia o seu aparelho a preços significativamente mais baixos do que a Apple, além de apostar em estar entre as primeiras empresas a adotar o novo sistema operacional Android (da Google). Em 2012 a Samsung já havia capturado um terço do mercado de smartphones.

Criar poder de mercado é apenas uma parte da história. Uma firma de sucesso deve também ter a habilidade de manter o poder de mercado.

Patentes expiram e imitação surge cedo ou tarde no mercado. Indústrias protegidas são desreguladas. O que as firmas incumbentes poderiam fazer para manter a sua posição? Uma possibilidade para uma empresa com poder de mercado, seria defender seu mercado com cortes drásticos de preços, criando assim reputação contra a investidas futuras.

Outra alternativa seria investir em conjunto de atrativos para o consumidor. Um exemplo é a cesta de serviços ofertadas pelas empresas de alta tecnologia do Vale do Silício. A Apple oferece hardware e software para os seus consumidores de forma integrada, tornando a utilidade dos bens e serviços inseparáveis.

**Quais são as implicações do poder de mercado?** Do ponto de vista da firma, poder de mercado implica em maiores margens, maiores lucros e maior valor das empresas. Do ponto de vista social, as implicações são mais complicadas.

O efeito de primeira ordem do poder de mercado é a transferência de renda dos consumidores para as empresas. Os consumidores precisam pagar mais para comprar um bem devido ao poder de mercado da empresa em determinar o preço acima do custo. Se a política pública coloca um peso maior sobre o bem-estar do consumidor em relação aos lucros das empresas, então esta transferência de poder de compra pode ser vista como um problema. Este é um dos principais balisadores da **política de defesa da concorrência (antitruste)**, que tem como foco defender os consumidores de mudanças no ambiente competitivo que visem apenas aumentar o poder de mercado das firmas.

Além do efeito de transferência de renda, preços elevados também implicam em alocação ineficiente de recursos na economia. Quando os preços são elevados acima do padrão

competitivo uma parte da demanda não será suprida. Esta é a **ineficiência alocativa** gerada pelo poder de mercado.

“O melhor de todos os lucros de um monopólio é a vida tranquila.” Esta frase é creditada ao economista John Hicks (1935). Um monopólio é a estrutura de mercado em que apenas uma firma oferta um bem ou um serviço, neste caso não existem rivais para competir por mercado. A implicação é a de que empresas com maior poder de mercado tem menos incentivos para reduzir custos. Por exemplo, antes dos anos noventa as empresas europeias de transporte aéreo eram conhecidas por serem menos eficientes do que as empresas americanas. Grande parte desta ineficiência pode ser explicada pela existência de mercado mais competitivo na América do Norte do que na Europa. Em outras palavras, o poder de mercado implica em **ineficiência produtiva**.

Quando o poder de mercado é artificialmente mantido pelo governo surge um terceiro tipo de ineficiência. Os economistas chamam esta ineficiência de **rent seeking** ou extração de renda. Rent seeking significa recursos que não produtivos mas são usados pelas empresas para tentar influenciar *policymakers*.

Considere por exemplo um artigo que explica o esforço da AT&T para manter sua posição no mercado de televisão a cabo. “Este verão, AT&T Corp. se deparou com um espectro de cidades em que será necessário abrir o seu mercado de TV a cabo para companhias de internet rivais... A ameaça nunca se materializou realmente. Por que não? Isto depende a quem você perguntar. A AT&T atribui o seu sucesso a sua capacidade de explicar a questão para oficiais locais... [Outros possuem opinião diferente] – Eles vem para oferecer suborno e ameaças, diz Greg Simon, co-diretor do Opennet Coalition, um grupo de empresas que lançou o seu próprio esforço de fazer lobby pela abertura do mercado.”<sup>3</sup>

Um resumo estilizado destas questões pode ser interpretado que o poder de mercado é bom para as empresas mas ruim para a sociedade. Primeiro, o poder de mercado em excesso torna os donos das empresas mais ricos às custas dos consumidores. Segundo, ele diminui a eficiência econômica. Terceiro, ele induz firmas a desperdiçar recursos para atingir e manter o seu poder de mercado.

Por outro lado, o poder de mercado em si não é um problema. Em geral o excesso de poder de mercado é custoso para a sociedade ou quando ele está sendo mantido para acordos não concorrenciais. Na visão da escola de Chicago, a existência de poder de mercado é transitória, pois se existem margens mais elevadas elas vão atrair novas empresas para o mercado.

**Existe função para a política pública em relação ao poder de mercado?** No contexto da economia industrial, o papel primário da política pública é evitar as consequências negativas do poder de mercado. A política pública nesta área pode ser dividida em duas categorias: regulação e defesa da concorrência (antitruste).

<sup>3</sup>The Wall Street Journal, 24 de Novembro de 1999. Cabral, 2017, p. 6.

1. Regulação econômica: em geral é o caso em que um empresa detém monopólio ou poder de mercado similar ao monopólio. Neste caso suas ações são continuamente acompanhadas pelo governo.
2. Defesa da concorrência (antitruste): defesa da concorrência é um campo mais amplo do que regulação. A ideia é prevenir que as firmas aumentem o poder de mercado por meio da redução da concorrência.

Alguns exemplos ajudam a explicar a defesa da concorrência.

- Pelas últimas duas décadas, Mars e Unilever entraram em uma série de casos legais nas cortes europeias. A questão era a legalidade das políticas de exclusividades nas vendas de sorvete. Em vários países da Europa, a Unilever impunha exclusividade no uso de refrigeradores. Se um loja aceita refrigerador da Unilever, então ela somente usar a unidade para produtos da Unilever. Mars alegava que a exclusividade tornava impossível a venda do sorvete Snickers e produtos correlatos, pois as lojas em geral não possuíam espaço para outro freezer. A Unilever respondia que como elas investiam em refrigeradores seria justo elas obterem retorno pelo seu investimento. Diversos casos parecidos foram julgados nas cortes europeias. Enquanto algumas decisões foram favoráveis à Unilever outras foram melhores para a Mars.<sup>4</sup>
- Em março de 2011, AT&T anunciou seus planos de aquisição da T-Mobile USA, uma operadora menor de telefonia móvel. Cinco meses depois, a Divisão de Antitruste do Departamento de Justiça (DOJ, sigla para *Department of Justice*) dos EUA formalmente anunciou que pediria o bloqueio da operação de aquisição da empresa, argumentando que isto aumentaria substancialmente o poder de mercado da AT&T. Em um primeiro momento a AT&T deu sinais de que poderia desafiar a decisão do DOJ no sistema de justiça, mas eventualmente a operação foi cancelada. Embora a fusão poderia trazer algum benefício para os consumidores, a concorrência entre as duas empresas são uma fonte muito importante de bem-estar para a sociedade.<sup>5</sup>

Os exemplos acima fornecem uma idéia da variedade de situações que podem fazer parte do escopo da política pública. A abordagem geral da política de defesa da concorrência é prevenir e apresentar remédios para situações onde o poder de mercado pode atingir níveis elevados.

A abordagem tradicional da escola de Chicago toma uma postura diferente frente a estes problemas. A afirmação clássica é a de que em um mundo de livre concorrência, o poder de mercado nunca será significativo. Em outras palavras, a escola de Chicago reverte a ordem de causalidade. Não é o poder de mercado que leva a intervenção governamental

---

<sup>4</sup>Luis M. Cabral. Case Study: Mars and Unilever. Leonard N. Stern School of Business, New York University.

<sup>5</sup>USA, DOJ, Antitrust Division v. AT&T Inc., T-Mobile, and Deutsche Telekom AG. Case 1:11-cv-01560. 8/31/2011. <https://www.justice.gov/sites/default/files/opa/legacy/2011/08/31/Justice-ATT-T-Mobile-Complaint.pdf>

mas o oposto. A intervenção do governo cria poder de mercado, protegendo os interesses das firmas e não o consumidor.

**Política industrial** Além da regulação de mercado e antitruste, alguns países tem perseguido políticas com a intenção de promover firmas em particular ou grupo de empresas. Esta é a chamada **política industrial**. O objetivo da política industrial é muito diferente da regulação e do antitruste. Enquanto esta última política visa a defesa da concorrência, a política industrial foca na proteção e fortalecimento da posição de mercado de uma firma ou indústria. Não raramente esta proteção ocorre em relação a empresas do exterior.

Um exemplo de sucesso é o caso da Airbus Industrie, um consórcio de empresas de quatro países europeus. Parte do sucesso da Airbus é resultado do suporte recebido dos respectivos governos ao longo de três a quatro décadas. Começando com participação de mercado inferior a 10% no começo dos anos 70, a Airbus agora lidera o mercado juntamente com a Boeing.<sup>6</sup>

Na prática, a política industrial funciona com os governos escolhendo os vencedores entre um número pontencial de firmas e indústrias. Por que os governos teriam informação mais qualificada do que o mercado sobre quais são as firmas ou indústrias mais promissoras? Um argumento frequente para defender a política industrial é o caso do MITI (Japanese Ministry of Industry and Foreign Trade). Por exemplo, o apoio do MITI foi um fator importante na emergência do Japão como um líder na exportação de semicondutores. Entretanto, junto com essa história de sucesso também existem algumas de fracasso: um exemplo é o projeto de desenvolver a “quinta geração” de computadores, que não teve sucesso frente aos concorrentes dos EUA.

Mesmo nos EUA existem exemplos de falhas na política industrial. Por exemplo, o Estado de Rhode Island aprovou 75 milhões de dólares em empréstimos para uma empresa iniciante de video games fundada por uma estrela da liga americana de Baseball. A empresa faliu em dois anos, deixando uma dívida de mais de 150 milhões de dólares no processo.

Em resumo, embora existam histórias de sucesso, os registros indicam que a política industrial gera mais fracassos do que sucessos. Por estas razões que geralmente quando se fala de políticas públicas o foco costuma ser em defesa da concorrência e regulação de mercados.

## 1.2 Organização e Metodologia

Muitos economistas analisam indústrias usando como referência a abordagem conhecida como o paradigma de **estrutura-conduta-desempenho** (veja Mason, 1939 e 1949,

---

<sup>6</sup>Luis M. Cabral. Case Study: Airbus and Boeing. Leonard N. Stern School of Business, New York University.

e Bain, 1956). A ideia do paradigma é primeiro olhar para os aspectos que caracterizam a estrutura de mercado (número e características de firmas, compradores e produtos/serviços). Em segundo lugar se presta atenção na conduta típica das firmas na indústria: determinação de preços, posicionamento dos produtos, publicidade, etc. Finalmente, se procede para estimar quão competitiva e eficiente a indústria é.

Este paradigma utiliza uma causalidade que é escolhida previamente. A estrutura de mercado determina a conduta das firmas, que por sua vez determina a performance da indústria e das firmas. Por exemplo, em uma indústria com poucos competidores, cada empresa tem maior possibilidade de manter preços mais elevados ou realizar conluio com outras empresas.

O problema é que a causalidade pode funcionar de forma reversa. Por exemplo, uma firma que não tem uma boa performance pode sair do mercado. Neste caso a performance afeta a estrutura do mercado. Da mesma forma uma firma pode manter um preço muito baixo para expulsar um rival do mercado. Novamente a conduta afeta a estrutura neste exemplo.

Este material é organizando partindo de fundamentos como demanda, produção e noção de equilíbrio.





## Capítulo 2

# Consumidores

Neste capítulo se trata do comportamento dos consumidores e da demanda por bens e serviços.

### 2.1 Preferências e Demanda

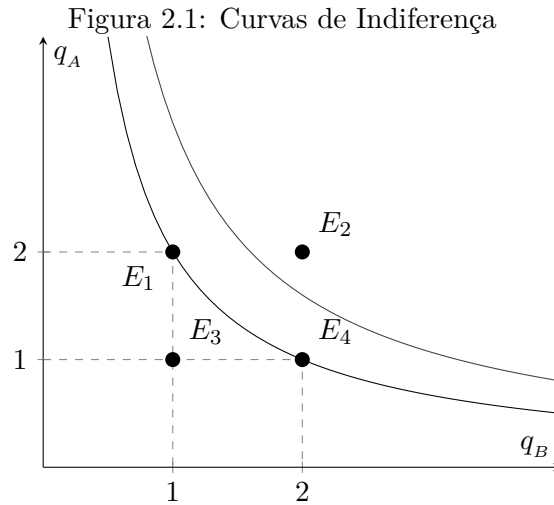
Preferências dos consumidores. O setup tradicional para pensar sobre demanda com gostos ou preferências para cada consumidor individualmente.

Começamos com um exemplo de dois produtos,  $A$  e  $B$ . Podemos expressar diferentes combinações de compra no diagrama com quantidades do produto  $A$  no eixo vertical e as quantidades do produto  $B$  no eixo horizontal, como descrito na Figura 2.1. Por exemplo, um consumidor pode querer a cesta  $E_2$  entre todas as 4 possibilidades. Este consumidor seria indiferente entre as cestas  $E_1$  e  $E_4$ , pois elas possuem a mesma utilidade. Esta linha que conecta os pontos  $E_1$  e  $E_4$  é uma curva de indiferença. Como explicado, o consumidor seria indiferente entre quaisquer combinações ao longo desta curva, pois sempre obtém a mesma utilidade. Em geral, as curvas de indiferença são negativamente inclinadas, pois necessitamos mais de um produto para compensar a perda de outro. Também se assume que as curvas de indiferença se posicionam a partir da origem, uma vez que o consumidor sempre deseja mais bens ou serviços.

Se empilharmos diversas curvas de indiferença em nosso gráfico, podemos obter uma descrição completa dos gostos do consumidor. Como mais “é sempre melhor”, as curvas de indiferença que estão à nordeste da origem são as mais desejadas pelos consumidores.

Fundamental na análise do consumidor esta no fato de que a escolha pelas cestas é feita sujeita à restrição de compra. Isto é, o consumidor deseja consumir mais, mas ele é restrito pelo sua renda. Chamamos esta decisão de escolha sujeita à restrição orçamentária.

Voltando ao nosso exemplo de dois produtos, uma restrição orçamentária de compra



entre os bens  $A$  e  $B$  é expressa pela seguinte desigualdade:

$$p_A q_A + p_B q_B \leq Y \quad (2.1)$$

tal que  $p_i$  é o preço do produto  $i = \{1, 2\}$ ,  $q_i$  é a quantidade comprada de  $i$ , e  $Y$  é a renda total disponível pelo consumidor.

Se solucionarmos para  $q_A$ , podemos observar que o **conjunto orçamentário** corresponde a área abaixo da seguinte reta (negativamente inclinada):

$$q_A = \frac{Y}{p_A} - \frac{p_B}{p_A} q_B \quad (2.2)$$

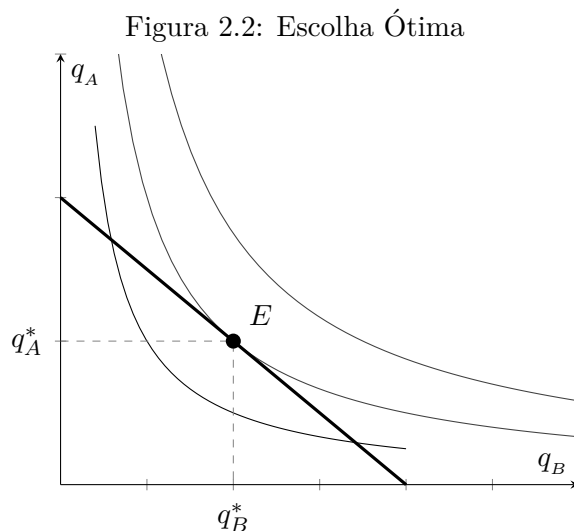
Como descrita aqui a restrição orçamentária é uma linha reta. Se o orçamento aumenta, então esta reta se desloca para a direita (nordeste). Se o preço do bem  $B$  se eleva, então a linha rotaciona no sentido horário (em cima do intercepto vertical  $Y/p_A$ ).

**Demanda.** Colocando junto as preferências e as possibilidades de consumo como função da renda podemos entender o processo de escolha do consumidor. Suponha que o objetivo do consumidor seja **maximizar a utilidade**. Neste sentido, ele deseja atingir a curva de indiferença mais alta que sua reta de orçamento permitir. Em outras palavras, a melhor escolha possível é aquela em que a curva de indiferença mais alta possui um ponto em comum com a restrição orçamentária. Isto é ilustrado na Figura 2.2. A melhor escolha possível é representada pelo ponto  $E$ , que corresponde as quantidades  $q_A^*$  e  $q_B^*$ . Implicitamente, estas demandas dependem do que se gosta, também dependem da renda  $Y$  e dos preços dos produtos. No ponto  $E$ , o consumidor está resolvendo o seguinte problema de maximização:

$$\max_{\{q_A, q_B\}} u(q_A, q_B) \quad (2.3)$$

sujeito à:  $p_A q_A + p_B q_B \leq Y$

tal que  $u(\cdot, \cdot)$  é a função utilidade ao se consumir os bens  $A$  e  $B$  e a restrição orçamentária é dada pela restrição (2.1). A solução do problema (2.3) é a escolha ótima  $(q_A^*, q_B^*)$ .



O que apresentamos são fundamentos do processo de escolha de produtos e serviços. A partir destes fundamentos podemos escrever funções de demanda por bens, que são as soluções destas escolhas. Uma *função demanda* é representada por

$$q_i(p_i, \mathbf{z}) \quad (2.4)$$

Esta função nos diz que a quantidade demanda  $q$  do bem  $i$ , depende ou dependeu do seu preço  $p$  e de valores de outras variáveis  $\mathbf{z}$ . Dentro do vetor  $\mathbf{z}$  podemos incluir a renda do consumidor, o produto de outros bens que afetem a compra, e quaisquer outros fatores relevantes do ponto de vista econômico para a compra do produto  $i$ .

Por exemplo, se  $i$  é o consumo de um tipo de refrigerante, dizemos que a sua demanda é função do preço do refrigerante, dos refrigerantes concorrentes, do preço de um sanduíche, bem como de sua renda. Que outros fatores poderíamos pensar que possam afetar a sua escolha?

Ao invés da função de demanda (6.6), é comum os economistas escreverem uma forma equivalente que é denominada de *função de demanda inversa*,  $p_i(q_i, \mathbf{z})$ . Esta função diz que o preço depende do volume vendido do produto  $i$  e dos demais fatores contidos em  $\mathbf{z}$ . Comumente, quando se desenha um gráfico da demanda se utiliza a função de demanda inversa. Na Figura (2.3) é representada a relação inversa entre preço e quantidade, mantendo fixo as demais influências sobre a demanda  $\mathbf{z}$ .

É importante distinguir de mudanças ao longo da curva de demanda e mudanças na posição da curva. Na Figura 2.3, mudanças no vetor  $\mathbf{z}$  podem fazer a curva se deslocar para cima ou para baixo.

Por exemplo, suponha a demanda por Coca-Cola em uma cidade seja  $q_c = q_c(p_c, \mathbf{z})$ , sendo representada na Figura 2.4. Suponha que o consumo atual seja descrito no ponto  $E_1$  utilizando uma representação de demanda inversa. Vamos imaginar alguns cenários. Se ocorrer um aumento do preço da Coca-Cola vamos observar queda no volume vendido na cidade. A nova relação preço-quantidade passa de  $E_1$  para  $E_2$ , sendo que volume de vendas será menor,  $q_2 < q_1$ , e o preço maior,  $p_2 > p_1$ .

Suponha que o preço da concorrente Pepsi faz parte do vetor  $\mathbf{z}$ . Portanto, após o aumento de preço da Coca-Cola o revendedor da Pepsi decide reduzir o preço do rival. Neste caso, teremos um deslocamento da curva de demanda em direção a origem. Como a Pepsi é concorrente, uma redução em seu preço provoca um *desvio de demanda* da Coca-Cola para a Pepsi. Neste caso, no ponto  $E_3$ , a quantidade vendida de Coca-Cola cai  $q_3 < q_2$ , mas o preço continua a mesmo,  $p_3 = p_2$ .

**Excedente do Consumidor.** Para analisar se os consumidores ficam em uma situação melhor quando um preço cai e a quantidade sobe é comum o uso do conceito de excedente do consumidor. Apesar do ganho do consumidor parecer direto quando ocorre uma queda de preço, ela não é tão simples devido a forma e curvatura da função demanda.

Um conceito auxiliar é importante: *disposição à pagar*. Disposição à pagar é o preço máximo que um indivíduo ainda estaria disposto a pagar para consumir.

O excedente do consumidor que é derivado de uma unidade de consumo é a *diferença entre o preço que o consumidor teria disposição à pagar e o que ele pagou efetivamente (a pagamento efetivo é o preço de mercado)*. Como o excedente do consumidor mede a diferença entre o preço pago e a disposição a pagar, temos um cálculo de área entre o máximo e o piso (dado pelo preço atual). Quantitativamente, o excedente do consumidor para o preço  $p_1$  e quantidade  $q_1$ , para a curva de demanda inversa  $p(q, \cdot)$ , será:

$$CS_1 = \int_0^{q_1} (p(q, \cdot) - p_1) dq = \int_0^{q_1} p(q, \cdot) dq - p_1 q_1 \quad (2.5)$$

tal que  $dq$  é a variação diferencial de  $q$ .

Na Figura 2.5 mostramos o excedente do consumidor. A área  $CS$  é da parte embaixo da curva de demanda e acima do dispêndio. O dispêndio é o gasto com a quantidade  $q_1$ , i.e.  $\text{dispêndio} = p_1 \times q_1$ .

As medidas de excedente do consumidor, que é uma medida de bem-estar, depende da função demanda escolhida. Como a função demanda é uma escolha do economista, então a medida de  $CS_1$  depende da escolha de  $p(q, \cdot)$ . Um aspecto prático de medida de  $CS_1$  indicaria o seu cálculo para algumas formas funcionais da demanda inversa (função log-linear, elasticidade-constante, etc.).

Particularmente, o comportamento da função de demanda inversa  $p(q, \cdot)$  perto da origem,  $q = 0$ , pode ter um grande impacto sobre o cálculo do excedente do consumidor. Como a função demanda é construída teoricamente e um pouco difícil de estimar a disposição a pagar para quando não existem vendas, o cálculo do  $CS$  pode ser muito sensível.

Figura 2.3: Função de Demanda Inversa

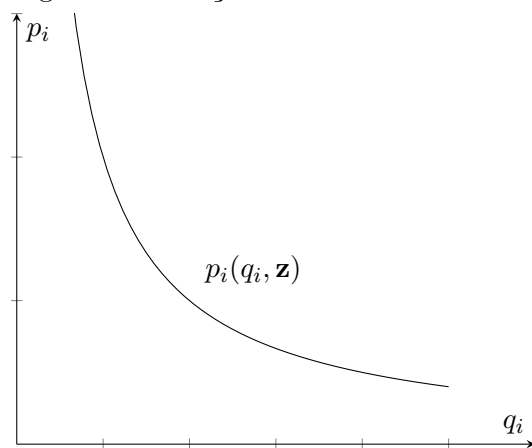
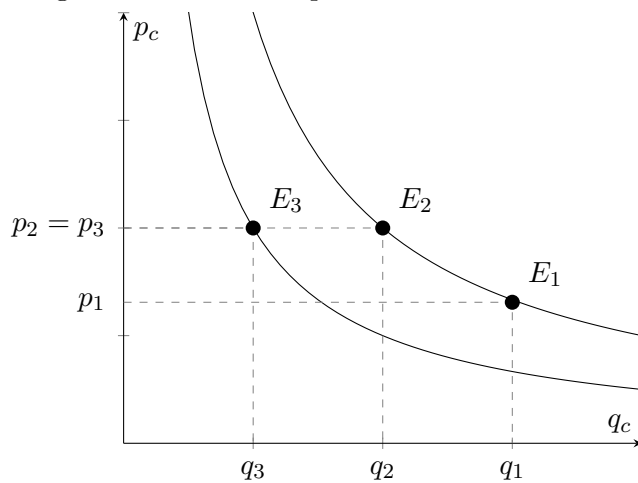


Figura 2.4: Demanda por Coca-Cola: Cenários



Devido a este tipo de dificuldade é comum que economistas estimem a mudança no excedente do consumidor e não o  $CS$ . Assim, a mudança no excedente do consumidor, representada por  $\Delta CS$ , depende menos das hipóteses sobre a curva de demanda. A mudança de bem-estar devida a uma mudança de preços de  $p_1$  para  $p_2$  é calculada por

$$\Delta CS = CS_2 - CS_1 = \int_0^{q_2} p(q, \cdot) dq - p_2 q_2 - \int_0^{q_1} p(q, \cdot) dq + p_1 q_1 \quad (2.6)$$

É importante lembrar que uma curva de demanda trata de um mercado. Portanto, existem consumidores que pagam mais o precisa mais de certo bem. Além da distribuição dos consumidores existe “saciedade” no consumo. Consumidores não compram demais um bom, pois ele vai perdendo utilidade a medida que se “compra mais.”

Excedente do consumidor é uma medida de um sistema econômico. Por exemplo, compare concorrência com monopólio, o impacto de tarifas e impostos, etc.

## 2.2 Elasticidade

Um conceito chave para a análise de mercado e antitruste é a elasticidade-preço da demanda de um produto. A elasticidade da demanda nos informa sobre a sua forma e posição, dizendo como o consumo depende dos diversos fatores: renda dos consumidores, preços dos produtos rivais, etc.

### 2.2.1 Elasticidade-Preço da Demanda

Baseado em evidências, assumimos que a demanda por um produto aumenta quando o seu preço cai. A questão é quão sensível é a demanda ao preço do produto. Pense no mercado de petróleo bruto, a demanda de petróleo diminui em um dia em 1,3 milhões de barris quando o preço aumento de 50 para 60 dólares por barril. Você consideraria a demanda por petróleo muito ou pouco sensível ao preço? Como bem explicado por Cabral (2017, p. 19-20), isso é difícil de dizer se você não tiver uma boa ideia do tamanho do mercado mundial de petróleo.

O problema aqui é que medir a inclinação da curva de demanda é dependente da unidade em que o produto é comercializado. A solução é medir em termos relativos, i.e. em termos percentuais. Portanto, a medida de sensibilidade de demanda é a elasticidade-preço da demanda ou elasticidade-própria ao preço,<sup>1</sup> que é definida pelos economistas como

$$\epsilon = \frac{\frac{dq_i}{q_i}}{\frac{dp_i}{p_i}} = \frac{dq_i}{dp_i} \frac{p_i}{q_i} \quad (2.7)$$

tal que  $\frac{dq}{dp}$  é a derivada parcial de  $q$  com respeito a  $p$ . A elasticidade é definida para o produto  $i$ . Definindo: *a elasticidade-preço da demanda,  $\epsilon$ , é a razão entre a mudança percentual na quantidade e a mudança percentual (pequena) do preço.*

<sup>1</sup>Também chamamos apenas de elasticidade da demanda.

Importante é que a elasticidade não é a mesma coisa que inclinação da curva de demanda. A inclinação da curva é  $dp/dq$ , que é um componente da elasticidade (2.7).

Em termos práticos, podemos escrever a elasticidade da demanda (2.7) como uma aproximação da variação, dado que a mudança nos preços, que é infinitesimal  $\Delta p$ , é muito pequena:

$$\epsilon \approx \frac{\Delta q_i p_i}{\Delta p_i q_i} \quad (2.8)$$

Geralmente as elasticidades são negativas, uma vez que para bens normais a quantidade demanda diminui com aumento no preço. A questão aqui então é quão negativa é a elasticidade da demanda. Geralmente para analisar esta sensibilidade pensamos no módulo da elasticidade da demanda:  $|\epsilon|$ . Portanto, as elasticidades podem ser classificadas como:

**Elástica:**  $|\epsilon| > 1$  quando a quantidade demanda é muito sensível ao preço. Quanto mais elevada for  $|\epsilon|$  mais sensível é a quantidade ao preço.

**Inelástica:**  $0 < |\epsilon| < 1$  neste caso a elasticidade é pequena. Significa que a quantidade é pouco sensível ao preço. Quanto mais próxima for  $|\epsilon|$  de zero menos sensível ela é ao preço.

**Unitária:**  $|\epsilon| = 1$  Elasticidade unitária é um caso teórico e prático.

A função de demanda linear permite interpretação “interessante” da elasticidade da demanda. Na Figura 2.6 a elasticidade vai de  $-\infty$  a 0, quando a quantidade é 0 até o preço atingir 0. No ponto intermediário, limítrofe, a elasticidade assume valor unitário,  $|\epsilon| = 1$ , quando a elasticidade é linear. No exemplo, da demanda inversa linear a elasticidade varia ao longo da curva, mas também existem casos em que a elasticidade da demanda é constante ao longo da função de demanda. Observe que a inclinação da curva ( $dq/dp$ ) é constante, mas a elasticidade não (dado que muda a relação  $p/q$ ). O caso de elasticidade constante é a demanda da Figura 2.3.

Aa elasticidade da demanda (2.7) também pode ser expressa em logaritmo. A forma log pode ser útil para realizar cálculos rápidos. Dado que  $dq/q = d(\ln(q))$ , a fórmula da elasticidade pode ser reescrita como:

$$\epsilon = \frac{\frac{dq}{q}}{\frac{dp}{p}} = \frac{d \ln(q)}{d \ln(p)} \quad (2.9)$$

neste caso a elasticidade da demanda é expressa pela derivada do log natural de  $q$  e  $p$ . Como na fórmula (2.8) aproximamos a derivada pela variação bem pequena, também podemos representar usando logs:

$$\epsilon \approx \frac{\Delta \ln(q)}{\Delta \ln(p)} \quad (2.10)$$

A aproximação (2.10) é exata se a elasticidade é constante ao longo da curva de demanda.

Figura 2.5: Excedente do Consumidor

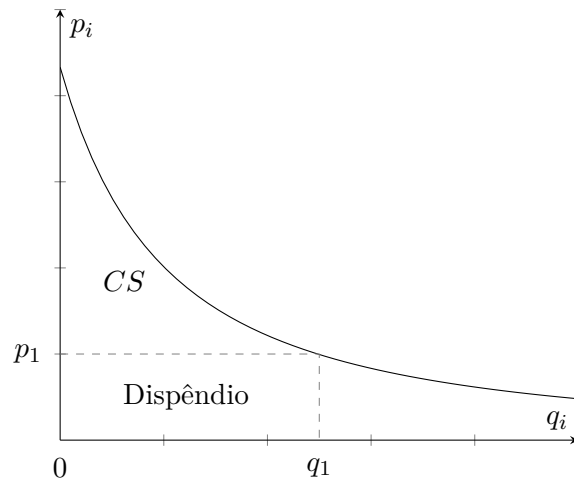


Figura 2.6: Elasticidade na Demanda Inversa Linear

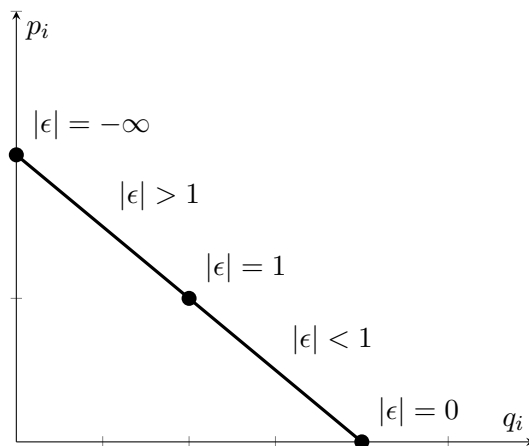




Tabela 2.1: Elasticidade-Preço da Demanda: Exemplos

Café nos Países Baixos	-0,2
Gás natural, Europa (curto-prazo)	-0,2
Gás natural, Europa (longo-prazo)	-1,5
Carros de luxo, EUA	-1,9
Carros de luxo importados, EUA	-2,8
Serviço transporte marítimo	-4,4

Fonte: Cabral, 2017, p. 21.

Na Tabela 2.1 apresentamos alguns exemplos da elasticidade da demanda. Observe que a demanda depende da definição do mercado e do produto. Além disso o tipo de preço também importa. Por exemplo, no mercado de petróleo existem preços à vista e também futuro (com diversas datas de vencimento). Outro caso interessante é que a demanda de longo prazo do gás natural é maior do que a de curto prazo.

A estimação da elasticidade da demanda é muito importante e é objeto de muitos estudos de mercado. No capítulo 6 vamos estudar alguns métodos de estimação da elasticidade. A elasticidade é importante entre outras coisas, pois sabendo o seu valor se pode projetar o quanto a quantidade vendida responde a variações no preço:  $\Delta q/q \approx \epsilon(\Delta p/p)$ . Da mesma forma que relacionamos a variação da quantidade com a variação no preço podemos projetar o impacto da variação no preço na variação na receita total. Relembre que receita total é:  $q \times p$ . Então

$$\frac{d(qp)}{pq} = \frac{qdp + pdq}{pq} = \frac{dp}{p} + \frac{dq}{q} \frac{p}{p} = (1 + \epsilon) \frac{dp}{p}$$

Como a elasticidade da demanda é negativa, então o efeito direto do preço (o termo 1) e o efeito indireto (via quantidade: termo da elasticidade) possuem efeito oposto. Se a demanda é elástica,  $|\epsilon| > 1$ , então o efeito da demanda é maior e um aumento no preço leva à queda da receita.

### 2.2.2 Elasticidade-Preço Cruzada

Vimos na análise da função demanda  $p_i(q_i, \mathbf{z})$  que ela depende do preço do próprio produto ou serviço, mas depende de outros fatores. Estes demais fatores são representados no vetor  $\mathbf{z}$ . Mudanças neste vetor causam deslocamentos da curva de demanda. Entre possíveis “componentes” deste vetor estão os preços dos produtos concorrentes. No exemplo da demanda por Coca-Cola (ver Figura 2.4), usamos o preço da Pepsi para mostrar deslocamentos da quantidade vendida do primeiro. Um outro exemplo poderia ser a demanda por viagem de metrô: se o preço da gasolina sobe é possível que mais pessoas usem o metrô para ir ao trabalho.

A elasticidade-preço cruzada da demanda (este é o nome completo, mas às vezes nos

Tabela 2.2: Elasticidades Próprias e Cruzadas: Carros nos EUA

Modelo	323	Cavalier	Accord	Taurus	Century	BMW 735i
323	-6,4	0,6	0,2	0,1	0,0	0,0
Cavalier	0,0	-6,4	0,2	0,1	0,1	0,0
Accord	0,0	0,1	-4,8	0,1	0,0	0,0
Taurus	0,0	0,1	0,2	-4,2	0,0	0,0
Century	0,0	0,1	0,2	0,1	-6,8	0,0
BMW 735i	0,0	0,0	0,0	0,0	0,0	-3,5

Fonte: Berry, Levinsohn, Pakes, 1995.

referimos apenas por elasticidade cruzada) entre os produtos  $i$  e  $j$  é definida como:

$$\epsilon_{i,j} = \frac{\frac{dq_i}{q_i}}{\frac{dp_j}{p_j}} \quad (2.11)$$

isto é, a elasticidade cruzada do produto  $i$  com respeito ao produto  $j$  é dado pela razão da mudança percentual na quantidade demandada do produto  $i$  em relação a mudança percentual do preço do produto  $j$ .

A elasticidade cruzada depende fundamentalmente se os produtos  $i$  e  $j$  são *complementares* ou *substitutos*. Se elasticidade cruzada é positiva,  $\epsilon_{i,j} > 0$ , então os produtos são substitutos. No exemplo relacionado à Coca-Cola, a Pepsi é um produto substituto. Neste caso, quando a Pepsi sob seu preço a demanda por Coca-Cola sobe.

Por outro lado, se os produtos são complementares a elasticidade cruzada será negativa. Se o preço da gasolina ficar mais elevado, as pessoas dirigem menos, levando a queda da demanda por carros. Neste caso, gasolina e carros seriam produtos complementares. Resumindo as elasticidades, temos:

- $\epsilon_{i,j} > 0$ , então  $i$  e  $j$  são *substitutos*
- $\epsilon_{i,j} < 0$ , então  $i$  e  $j$  são *complementares*
- $\epsilon_{i,j} = 0$ , então  $i$  e  $j$  são *independentes*

Como exemplo de elasticidade cruzada podemos mostrar os números estimados para a demanda por carros nos EUA. Estas elasticidades foram estimadas para os anos setenta e oitenta nos EUA e faz parte de um modelo inovador proposto por Berry, Levinsohn e Pakes (1995). O modelo proposto é muito avançado, mas podemos ter uma visão clara das elasticidades cruzadas. Por exemplo, o número 0,2 na terceira coluna, segunda linha, significa que um aumento de 1% no preço do Accord leva a  $0,2 \times 1\%$  de aumento nas vendas do Cavalier. Também observe que na diagonal da tabela são as elasticidades próprias, sempre assumindo valores negativos.

### 2.2.3 Elasticidade da Renda

Como estamos falando de demanda por bens e serviços, a renda disponível é fundamental para entender o processo de escolha. Em geral, a renda mais elevada significa mais demanda por produtos e menor sensibilidade ao preço. A elasticidade da renda é definida como:

$$\eta = \frac{\frac{dq_i}{q_i}}{\frac{dy}{y}} \quad (2.12)$$

Isso significa que a elasticidade-renda da demanda é dada pela mudança percentual na quantidade demandada induzida por uma mudança de 1% na renda ( $y$ ).

Os bens podem ser classificados em relação à sensibilidade da renda. *Bens normais*, que comumente se consomem, possuem elasticidade-renda positiva. Entre os bens normais, aqueles com elasticidade entre 0 e 1 são denominados de bens como *necessidades*, e aqueles com elasticidades superior a 1 são bens de *luxo*. Alguns bens possuem elasticidades-renda negativa. Estes são denominados de *bens inferiores*, pois quando a renda sobe as pessoas deixam de consumir o produto ou serviço.

#### Demanda por Gasolina

Vamos considerar um simples exemplo para trabalhar os conceitos de elasticidade. Usando dados históricos do preço de gasolina e consumo nos EUA de 2001 a 2006, Hughes, Knittel e Sperling (2008) estimaram a seguinte relação:

$$\ln q = -1,697 - 0,042 \ln p + 0,530 \ln y \quad (2.13)$$

tal que  $q$  representa o consumo de gasolina,  $p$  o preço médio da gasolina e  $y$  a renda. Aqui não vamos entrar em detalhes a respeito do modelo de estimação, mas observe que o mercado de gasolina aqui é amplo e não estamos considerando produtos substitutos ou alternativas no uso da gasolina.

A elasticidade da demanda da gasolina é:

$$\epsilon = \frac{d \ln q}{d \ln p} = -0,042$$

Qual a elasticidade da demanda por gasolina? Como  $|\epsilon| < 1$ , concluímos que ela é inelástica.<sup>2</sup> Qual a elasticidade-renda da gasolina?

$$\eta = \frac{d \ln q}{d \ln y} = 0,530$$

A gasolina portanto é um bem normal, dado que seu consumo sobe com a renda. Como esta elasticidade é menor do que 1, interpretamos a gasolina como uma *necessidade*. Observe que o corte de bem necessário para luxo não é puramente arbitrário. Na verdade,

<sup>2</sup>Observe que quando consideramos mercados amplos, agregados e sem substitutos maior a tendência da demanda ser inelástica.

se a elasticidade é maior do que 1 então um aumento médio na renda implica que a fração de renda gasta no bem aumenta. Então, na média, pessoas mais ricas gastam mais em gasolina mas gastam proporção menor da sua renda em gasolina.

De 1953 a 2004,  $p$  e  $y$  aumentaram nas seguintes taxas anuais de crescimento: 3.9 e 3.4, respectivamente. Quanto poderíamos esperar de crescimento na demanda? Usando os conceitos de elasticidade podemos calcular o crescimento previsto. Relembre que  $\frac{dz}{z} = \epsilon_{zx} \frac{dx}{x}$ , para qualquer  $z$  e  $x$ . Assim:  $\frac{\Delta q}{q} = -0.042 \times 3.9\% + 0.530 \times 3.4\% = -0.164\% + 1.802\% = 1.638\%$ . Então 1,638 % seria a previsão de crescimento (note que o observado foi de 2.7%).

## 2.3 Racionalidade

O que foi apresentado até o momento está baseado na premissa de que os agentes (indivíduos, firmas, etc.) são racionais e otimizam suas decisões. O modelo de consumidor apresentado aqui assume que o indivíduo ou família maximiza sua utilidade de consumo sujeito à restrição de renda (equação (2.3)). O maior desafio ao paradigma econômico está no axioma de que consumidores/agentes são hábeis ao comparar e ordenar alternativas de uma forma consistente. Um desafio central a esta hipótese é dada pela teoria de prospecto (prospect theory), que a teoria de que as preferências dependem dos pontos de referência.

Por exemplo, faça  $u_1$  ser a utilidade de um consumidor ter uma maçã e  $u_0$  a utilidade de não ter este bem. Se o preço é menor do que  $u_1 - u_0$ , então eu espero que o consumidor compre uma maçã. Se o preço  $p$  é maior do que  $u_1 - u_0$ , então eu espero o oposto. Até aqui estamos na teoria padrão de comportamento do consumidor. Agora suponha que a utilidade também depende das expectativas. Se um consumidor estava esperando não ter uma maçã, então a utilidade de ter uma maçã é  $u_1 + \mu_G(u_1 - u_0)$ , tal que  $\mu_G(\cdot)$  mede a “surpresa” de ter uma maçã. Por outro lado, se o consumidor estava esperando ter uma maçã, então a utilidade de não ter uma maçã seria  $u_0 + \mu_L(u_1 - u_0)$ , onde  $\mu_L(\cdot)$  mede o impacto extra de não ter a maçã. Finalmente, se o consumidor tem que o espera, então a utilidade é simplesmente dada por  $u_i$  (para  $i = \{0, 1\}$ , dependendo se o consumidor tem ou não a maçã).

Evidência experimental sugere que os consumidores são avessos à perdas. Ou seja,  $\mu_L$  é negativo e mais importante do que  $\mu_G$ . Desta forma, pode ser mostrado que para alguns preços  $p$ , um consumidor cujo ponto de referência é comer uma maçã, ele irá comprar uma. O contrário é verdade para quem não iria comer uma maçã. Em outras palavras, dois consumidores com a mesma função utilidade fazem escolhas diferentes simplesmente porque eles possuem expectativas diferentes.

## Capítulo 3

# Firmas e Produção

Neste capítulo apresentamos fundamentos da análise dos determinantes da oferta.

### 3.1 Firmas

Produtos e serviços são ofertados por firmas. Vamos abstrair do controle das firmas, ou seja, se uma firma tem papéis negociados na bolsa ou é uma empresa familiar. Pensamos na firma como uma unidade que precisa tomar decisões para comercializar bens e/ou serviços. A principal decisão e objetivo da firma é produzir lucros para os seus proprietários.

Como enfatizado na introdução, a variável de maior interesse em estudos de mercados e antitruste é a lucratividade das firmas. O lucro de uma firma sempre será

$$\text{Lucro total} = \text{Receita total} - \text{Custo total}$$

por receita total se entende a soma de todas as receitas das firmas. Custo total também representa a soma dos custos.

Mais precisamente podemos definir uma função lucro para uma firma que produz um único bem (ou serviço) como

$$\pi = p \times q - C(q) \tag{3.1}$$

tal que  $\pi$  é o lucro total,  $p$  o preço do produto,  $q$  a quantidade total comercializada e produzida e  $C(\cdot)$  é o custo total, que é uma função da quantidade total produzida. A receita total neste caso é  $p \times q$ . Assumimos que a função custo cresce com a quantidade produzida. Isto é, quanto mais se produz mais custo se tem. A receita total cresce com o preço cobrado e com a quantidade vendida. Para entender o comportamento das firmas, i.e. a lucratividade delas, precisamos analisar o comportamento da função custo, da capacidade produzida e da determinação dos preços. Vamos começar com  $q$ .

A firma mais simples de se pensar em uma firma é como uma unidade que transforma insumos em produtos ou serviços. Esta é a forma mais simples de pensar, mesmo

correndo o risco de super-simplificar a realidade. Por exemplo, uma refinaria produz diversos produtos derivados do petróleo, mas para produzir ela precisa de muito insumo (petróleo) e de trabalhadores operando a estrutura de uma refinaria. Portanto a união do mix produto gera produtos como gasolina e óleo diesel.

Podemos representar esse mix de insumos gerando um produto ou serviço por meio de uma **função de produção**. Portanto, a função de produção representa um mapeamento do quanto de produto uma firma gera para um dada combinação de insumos. Naturalmente, a capacidade produtiva varia de firma para firma, pois algumas serão mais eficientes para combinar insumos. A capacidade produtiva também depende da qualidade dos insumos. Produtos construídos com insumos melhores podem possuir um valor maior do que os que foram construídos com insumos mais baratos.

Formalmente a função de produção e a transformação de insumos em produto e é dada pela função  $f(\cdot)$ . Inicialmente podemos supor dois insumos para a função de produção: capital e trabalho. Portanto representamos a transformação de insumos em produto pela seguinte função:

$$q = f(k, l) \tag{3.2}$$

aqui  $k$  representa o estoque de capital da firma,  $l$  é a quantidade de trabalho utilizada pra gerar a quantidade de produto  $q$ . Quando os insumos são capital e trabalho, naturalmente o produto gerado é medido como valor-adicionado. Valor-adicionado é uma medida de produção que contempla apenas o que a firma adiciona de valor ao total de insumos utilizados na produção. Por exemplo, todo bem possui um preço e neste valor final contém o preço pago pelos insumos (incluindo impostos, mas vamos abstrair deles por enquanto). Se um bem tem o preço final subtraído do preço (proporcional!) de todos os insumos então temos uma medida de valor adicionado. Em suma, valor-adicionado é o valor da produção final menos o valor dos insumos. Esta é a medida natural de  $q$  quando se considera uma medida de produção sem insumos intermediários.<sup>1</sup>

Uma função de produção em que o produto  $q$  seja uma medida bruta de produção deve incluir todos os seus insumos. Naturalmente, quais insumos incluir depende do produto, da firma e do setor analisado. Uma função de produção que tem insumos foi apelidada por economistas como uma função KLEM, pois além de capital e trabalho é utilizado dois insumos intermediários: energia,  $e$ , e materiais,  $m$ . Materiais é tudo que é comprado de outra firma e que entra na composição do produto final. A função de produção KLEM é representada como

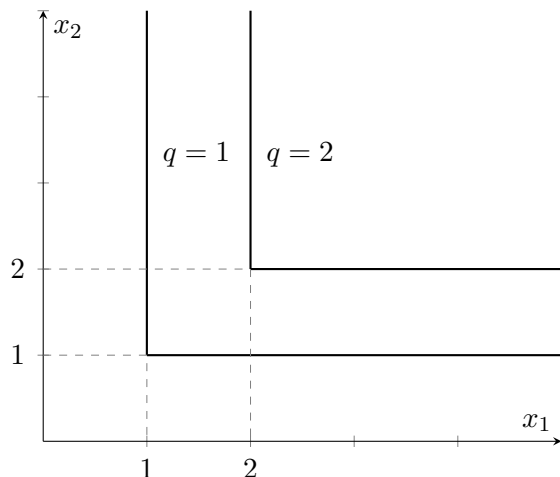
$$q = f(k, l, e, m) \tag{3.3}$$

Como a função de produção  $f(\cdot)$  deveria se parecer? Uma forma de responder esta questão é responder uma questão relacionada: qual combinação de insumos leva a um

<sup>1</sup>As vezes pode ser usado um conceito de função de produção com único insumo. É uma ideia menos frequente mas não é estranha nas aplicações em economia. Veja por exemplo o trabalho de Chandra, Finkelstein, Sacarny, e Syverson (2016).

dado nível de produto ( $q$ )? A resposta é dada pela *curvas isoquantas*, que se parecem com as curvas de indiferença.

Figura 3.1: Proporções Fixas



Na Figura 3.1 é apresentado a função de produção de proporções fixas. Por exemplo, Delta Airlines utiliza aviões ( $x_1$ ) e pilotos ( $x_2$ ) para produzir serviços de transporte. Cada avião requer dois pilotos. Então ter mais aviões do que pilotos não permite a empresa transportar mais passageiros. Como podemos ver, a função de produção de proporções fixas leva a curvas com ângulos retos. A função de produção com proporções fixas assume a seguinte forma:  $f(x_1, x_2) = \min\{x_1, x_2\}$ . Dizemos que nesse caso, os insumos são complementos perfeitos.

Na Figura 3.2 é o caso oposto que é de substitutos perfeitos. Nesse caso, se pode substituir perfeitamente entre os dois insumos. Ou seja, o quanto que se produz independe da escolha de insumo. Substitutos perfeitos leva a isoquantas retas.

Finalmente, mostramos o caso intermediário, isto é, um caso quando os insumos não são nem substitutos perfeitos nem complementos perfeitos. As isoquantas deste caso são derivadas de uma função de produção particular. Ela é a chamada função de produção Cobb-Douglas. Se o capital,  $k$ , e o trabalho,  $l$ , são os dois fatores produtos, ela assume a forma:

$$q = k^\alpha l^\beta \quad (3.4)$$

Os parâmetros  $\alpha$  e  $\beta$  capturam a elasticidade de substituição entre os fatores produtivos.

Na Figura 3.3 apresentamos um exemplo paramétrico de Cabral (2017, p. 37-38) para a função (3.4). Neste exemplo, o valor dos parâmetros são  $\alpha = \beta = 0,5$  e são descritas três possibilidades de produção, i.e. quando a quantidade for 1, 2 e 3. Uma característica interessante destas isoquantas é que se usarmos menos capital, podemos compensar com maior uso de trabalho. Além disso, o quanto mais  $l$  for diminuído mais é preciso aumentar  $k$  para compensar a queda no fator trabalho. Esta última propriedade reflete a **lei**

**dos retornos marginais decrescentes.** Isto também funciona no sentido contrário: quando temos aumento do fator trabalho. Considere a isoquanta em que  $q = 2$ . Uma forma possível de atingir este nível de produção é empregar 2 unidades de  $k$  e 2 de  $l$ . Pelo aumento de  $l$  para 3, podemos reduzir  $k$  de 2 para 1.33. Isto é uma queda de 0.66 no estoque de capital. Se novamente aumentarmos  $l$  de 3 para 4, podemos reduzir  $k$  de 1.33 para 1. Isto gera uma nova queda de 0.33 no estoque de capital. Este padrão de redução é um caso onde está funcionando a lei de rendimentos decrescentes.

Devido a **lei dos retornos marginais decrescentes**, as isoquantas são curvas convexas. Além disso, as isoquantas são mais convexas quanto mais próximos em complementariedade os dois insumos são.

A escolha das formas funcionais tem como base o uso da função de produção para análise econômica. Os objetivos foram bem resumidos por Fuss, McFadden e Mundlak (1978, pp. 220-222):

**Distribuição:** distribuição trata da estimação da participação dos fatores na produção.

No caso da função de produção Cobb-Douglas seriam os parâmetros  $\alpha$  e  $\beta$ . Historicamente grande parte da atenção em estimação dos parâmetros foi relativa a questões macroeconômicas, questões como crescimento econômico e ciclo econômico. Pelo menos nas últimas três décadas os estudos microeconômicos ganharam proeminência. Análise de impacto de reformas, estudos tributários e eficiência alocativa fizeram uso deste arcabouço.

**Escala:** se preocupa com questões relativa a existência ou não de retornos constantes de escala, ou a presença de retornos crescentes ou decrescentes. A escala de produção tem implicações agregadas para o crescimento econômico e para a estrutura da indústria. Questões microeconômicas sobre a oferta e financiamento de serviços públicos se baseia na questão tecnológica da existência de retornos crescentes de escala.

**Substituição:** é o grau de substitubilidade dos fatores de produção. Trata da análise da substituição dos fatores produtivos.

**Separabilidade:** decomposição das relações de produção em componentes aditivos ou “casados.” Separabilidade é muito importante na escolha da função de produção. Por exemplo, capital e trabalho podem ser separáveis na função de produção?

**Mudança Técnica:** modificação da estrutura tecnológica ao longo do tempo. A forma como o progresso técnico incide na função de produção pode variar. Existem inovações que não são específicas ao capital utilizado, chamada de *mudança técnica desincorporada* (do capital). Outras inovações precisam de um tipo de capital para ocorrer, i.e. elas ocorrem incorporadas a um tipo de capital ou incorporadas à mão-de-obra qualificada. Este tipo de setup é chamado de *mudança técnica incorporada* (em algum fator produtivo).



Figura 3.2: Substitutos Perfeitos

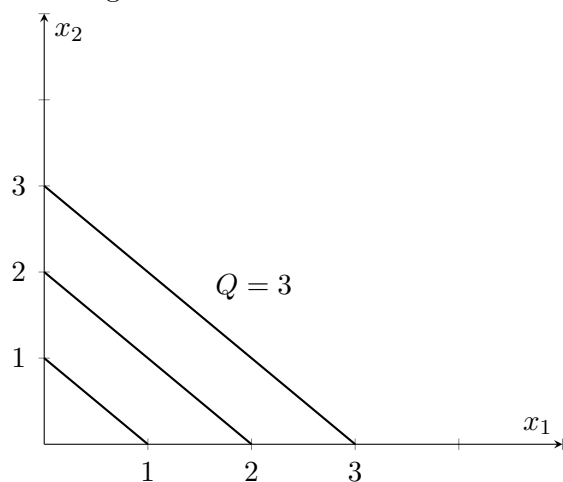
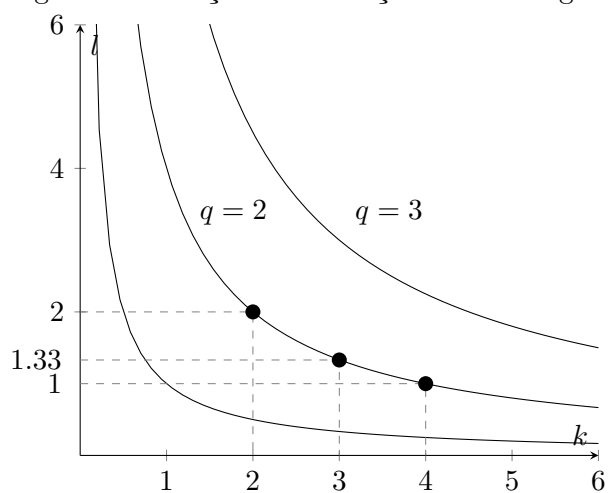


Figura 3.3: Função de Produção Cobb-Douglas



### Função Geradora de Vendas

Se substituirmos a variável de resultado da função de produção de quantidade produzida para receita total, i.e. de  $q$  por  $p \times q$  – sendo  $p$  o preço do produto, a denominamos de função geradora de vendas:

$$pq = f(k, l) \quad (3.5)$$

Neste caso, a combinação do mix de insumos entrega um resultado composto pelo preço do produto e pela quantidade produzida.

#### 3.1.1 Produtividade e Mudança Técnica

O termo produtividade é utilizado com muita frequência para descrever a performance de uma firma ou mesmo da economia. Existem algumas medidas de produtividade e cada uma pode ser bem utilizada dependendo do objetivo.

Se uma firma produz apenas um produto a sua receita total é o preço do produto multiplicado pela quantidade:  $pq$ . Uma medida de produtividade pode ser simplesmente a receita por trabalhador:  $(pq)/l$ . Esta medida é a produtividade do trabalho. A produtividade do trabalho também pode ser calculada como  $q/l$ , neste caso seria o produto por trabalhador. Como em algumas indústrias a firma pode determinar o preço em um processo distinto da produção, a produtividade  $q/l$  costuma ser mais apropriada. A produtividade do trabalho tem a sua utilidade, entretanto como estamos analisando a firma a partir da função de produção, outras medidas são mais adequadas.

Considere o seguinte exemplo: suponha duas firmas com a mesma função de produção. Se uma firma tem mais capital do que a outra, então ela terá maior produtividade do trabalho para a mesma quantidade produzida. Entretanto, não significa que no uso geral dos insumos uma firma seja mais produtiva do que outra, pois apenas existe uma combinação diferente de insumos.

Uma medida que corrige para o uso de todos os insumos é a **produtividade total dos fatores** (PTF). A medida da PTF é similar a produtividade do trabalho, pois dividimos o produto pelos insumos. A combinação dos insumos depende da forma funcional da função de produção. Por exemplo, a combinação é exatamente correta quando utilizamos a função Cobb-Douglas. A medida comumente utilizada de produtividade total dos fatores assume três hipóteses:

1. o preço de mercado é dado (é o mesmo para todas as firmas na indústria);
2. cada firma tem uma função de produção Cobb-Douglas com insumos  $k$  e  $l$  (no caso de valor-agregado);
3. todas as firmas tem os mesmos coeficientes da função de produção,  $\alpha$  e  $\beta$  em nosso exemplo, mas podem utilizar diferentes combinações dos insumos.

Hipóteses 2 e 3, implicam que as firmas operam a mesma tecnologia. A hipótese 1 é importante quando a produtividade é medida utilizando a receita e não a quantidade produzida ( $q$ ). Quando for disponível a quantidade produzida é geralmente recomendável

medir a produtividade pela quantidade. Quando se mede a produtividade usando receita, podemos dizer que a PTF foi medida a partir de uma função geradora de vendas.

**Produtividade Hicks-Neutra** é o caso em que a produtividade  $\omega$  pode ser separável dos fatores produtivos.<sup>2</sup> No caso da função de produção Cobb-Douglas temos:

$$q = k^\alpha l^\beta \omega \quad (3.6)$$

tal que  $\omega$  é a produtividade total dos fatores. Medidas de produtividade são usualmente calculadas como resíduo. Neste caso a PTF é calculada como:  $\omega = q/(k^\alpha l^\beta)$ . As elasticidades de produção ( $\alpha$ ,  $\beta$ ) não são diretamente observadas, elas são propriedades da função de produção.

No caso da função geradora de vendas:

$$pq = k^\alpha l^\beta a. \quad (3.7)$$

aqui  $a$  é a PTF da função geradora de vendas. Sob as hipóteses 1 a 3, as duas medidas de produtividade resultam na mesma medida.<sup>3</sup> A análise da produtividade é central em estudos de mercado e de performance econômica e grande da análise faz uso destes conceitos.

**Produtividade Fator-Expandida** é quando a produtividade pode expandir a capacidade produtiva de algum, ou mesmo todos, os fatores de produção. Dito de outra forma, nesta formulação geral é permitido aumentos na tecnologia para expandir a produtividade marginal de cada insumo separadamente:

$$q = (\gamma_k k)^\alpha (\gamma_l l)^\beta \quad (3.8)$$

tal que  $\gamma_k$  e  $\gamma_l$  são os fatores expansores da tecnologia para  $k$  e  $l$ , respectivamente.<sup>4</sup>

## 3.2 Minimização de Custo

Na seção anterior relacionamos o mix de insumos com a capacidade produtiva de uma firma e definimos algumas medidas de produtividade. Agora vamos entender a função custo  $C(q)$ .

A pergunta que podemos fazer é: dado o nível de produto desejado e o preço dos insumos qual o mix de insumos que minimiza o custo de produção. Este é um processo de escolha similar a maximização de utilidade do consumidor. Anteriormente vimos que o consumidor escolhe sua cesta de bens/serviços sujeito a restrição de renda e aos preços. No lado da firma, analisamos a escolha dos insumos dado seus preços e qual quantidade

<sup>2</sup>Este conceito é denominado Hicks-Neutra ao tipo de concepção desenvolvida por John Hicks (Hicks, 1946).

<sup>3</sup>Questões que tocam as medidas em mais detalhe são apresentadas no Capítulo 12.

<sup>4</sup>Sobre outras medidas de função de produtividade, veja por exemplo Hulten (2001).

se deseja produzir. Colocando de outro forma, assim como os consumidores possuem curvas de demanda por bens, as firmas possuem curvas de demanda por insumos.

A função custo, anteriormente definida por  $C(q)$ , possui alguns componentes:

**Custo fixo:** é o custo produtividade que não depende da quantidade produzida em períodos curtos de tempo. Por exemplo, o custo do prédio onde se produz não depende do quanto de  $q$  está sendo produzido nele. O custo do de prédio terá que ser pago de qualquer forma. Vamos representar esse custo por  $CF$ .<sup>5</sup>

**Custo variável:** é o custo que é função direta do volume de produção. Neste caso, o custo variável é zero quando não se produz bens. O custo variável é  $CV$ .

**Custo total:** o custo total ( $CT$  ou  $C(q)$ ) é a soma do custo variável e do custo fixo. Isto é:  $C(q) = CV + CF$ .

A função custo total,  $C(q)$ , depende do quanto se produz. Logo podemos calcular o custo médio (ou custo unitário) por produto, isto é:

$$\text{Custo médio (CM)} = \frac{C(q)}{q}$$

Outro conceito de custo importante é o *custo marginal*. O custo marginal pode ser definido como o custo de se produzir uma unidade a mais de produto. Esta medida que também é chamada de custo incremental é custo de produzir  $q + 1$  unidades de produto menos o custo de se produzir  $q$  unidades. Formalmente o custo marginal é a derivada do custo total com respeito ao nível de produção:

$$\text{Custo marginal (CMg)} = \frac{\partial C(q)}{\partial q}$$

Um exemplo simples de relação dos componentes do custo seria por exemplo:

$$C(q) = 30 \times q + 3000$$

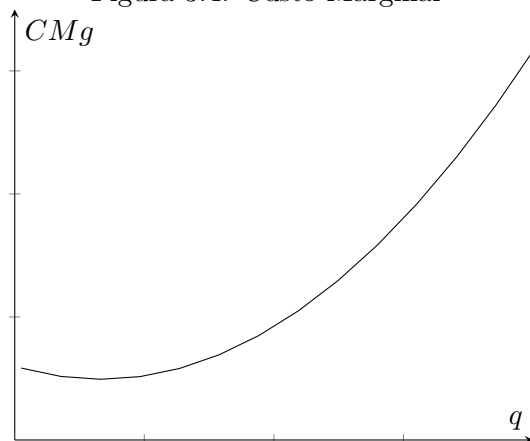
Se a firma produzir 300 unidades, então o custo total será 12000, o custo variável será 9000, o custo unitário é  $12000/300 = 40$ . O custo marginal é a derivada da função custo com respeito a quantidade:  $\partial C(q)/\partial q = 30$ .

O custo marginal é o conceito apropriado para se decidir o quanto produzir, enquanto que o custo médio é o conceito de custo apropriado para decidir o quanto se produz no total. Uma curva de custo marginal típica é apresentada na Figura 3.4.

---

<sup>5</sup>Observação: o custo do prédio não muda em períodos curtos. Se a firma tiver sucesso ela pode querer construir mais prédios. Por hora, o conceito importante do custo fixo é que ele independe do quanto se produz de  $q$ .

Figura 3.4: Custo Marginal

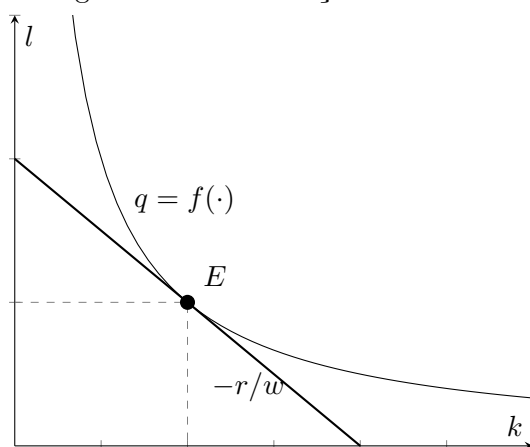


### 3.2.1 Minimização de Custos

O processo de minimizar custo trata da escolha da produção dada a restrição de contratação de insumos. Por exemplo, se os insumos forem capital e trabalho, a restrição é  $rk + wl$ , sendo  $r$  o preço do capital, que é uma taxa de juros, e  $w$  uma medida de salarial (preço do trabalho). Na Figura 3.5 a restrição, similar a restrição orçamentária do consumidor, é uma reta relacionando  $k$  com  $l$ . A restrição tem inclinação  $-r/w$ , dado que  $l = -(r/w)k$ . A escolha de intensidade dos insumos depende do respectivo preço relativo.

As firmas buscam custo mais baixo que seja consistente com um certo nível de produto. Esta escolha de custo mínimo é o ponto tangente da função de produção com a reta de preços relativos, que representado pelo ponto  $E$  na Figura 3.5.

Figura 3.5: Minimização de Custo



A demanda da firma por insumos pode ser mais ou menos dependente do preço, bem como dependente do preço dos outros insumos. Deste modo, a elasticidade preço-cruzada da demanda por insumos depende da forma da função de produção (Cobb-Douglas, CES, etc.).

Se a função de produção é próxima do caso extremos de complementares perfeitos (Figura 3.1), então um aumento no custo do capital, por exemplo, leva a queda na demanda por capital e também na demanda por trabalho. Neste caso a elasticidade-preço cruzada é negativa. Por outro lado, se a função de produção é próxima de substitutos perfeitos, dependendo do preço relativo dos insumos, a firma usa apenas capital ou trabalho. Neste caso, a elasticidade-preço cruzada é positiva: um aumento no preço do capital leva ao aumento da demanda por trabalho, por exemplo.

Vamos apresentar o setup do problema de minimização de custos. Suponha uma função de produção KLEM:

$$q = f(k, l, e, m) \quad (3.9)$$

como antes:  $q$  é produção total,  $k$  o estoque de capital,  $e$  energia e  $m$  materiais. Precisamos assumir que a função  $f(\cdot)$  é contínua e duas vezes diferenciável em respeito aos seus argumentos  $(k, l, e, m)$ .

O produtor deseja produzir  $q$  com o menor custo possível. A função custo total é

$$C(q) = wl + p_e e + p_m m + rk$$

tal que  $w$  é a taxa salarial,  $p_e$  é o preço da energia,  $p_m$  é o preço dos materiais,  $r$  é a taxa de juros (preço do capital). Portanto o problema de minimização é:

$$\min\{wl + p_e e + p_m m + rk\} \quad (3.10)$$

sujeito à:

$$q \leq f(k, l, e, m)$$

O problema minimização com restrição pode ser solucionado usando o lagrangiano  $L(\cdot)$ :

$$L(l, e, m, k, \lambda) = \{wl + p_e e + p_m m + rk\} + \lambda(q - f(k, l, e, m)) \quad (3.11)$$

$\lambda$  é o multiplicador de Lagrange.

A condição de primeira ordem com respeito ao salário  $w$  é:

$$\left[ \frac{\partial L(\cdot)}{\partial w} \right] l - \lambda \frac{\partial f(\cdot)}{\partial w} = 0 \quad (3.12)$$

A condição de primeira em relação a outros fatores é similar.

$$\frac{\partial f(\cdot)}{\partial w} = \frac{l}{\lambda}$$

Multiplicando ambos os lados por  $w/f(\cdot)$  e lembrando que  $f(\cdot) = q$ , temos

$$\frac{\partial f(\cdot)}{\partial w} \frac{w}{q} = \frac{l}{\lambda} \frac{w}{q}$$

organizando a equação temos:

$$\frac{\partial f(\cdot)}{\partial w} \frac{w}{q} = S_l \frac{w}{\lambda} \quad (3.13)$$

tal que  $S_l = \frac{l}{q}$  é a participação do insumo na produção final, e  $\lambda$  pode ser interpretado como o custo marginal da produção para um dado nível de produção,  $\lambda = \frac{\partial L}{\partial q}$ . Portanto, o quanto se usa do insumo  $l$  depende da elasticidade do insumo com respeito ao preço e da razão entre o custo marginal da produção e o preço do insumo.

$$\lambda = l \frac{w}{q} \frac{1}{\frac{\partial f(\cdot)}{\partial w} \frac{w}{q}}$$

$$\frac{k}{l} = \frac{w}{r} \frac{\frac{\partial f(\cdot)}{\partial r} \frac{r}{q}}{\frac{\partial f(\cdot)}{\partial w} \frac{w}{q}}$$

A intensidade no uso do capital depende do preço relativo ( $w/r$ ) e da elasticidade do produto em relação ao preço de cada insumo – isto é, depende da curvatura da função de produção.

**Demanda ótima por insumos.** Para o preço dos insumos constantes também podemos calcular a demanda ótima por insumos.<sup>6</sup> Suponha novamente o Lagrangiano (3.11), para calcular o uso ótimo de insumos derivamos (3.11) com respeito ao uso dos insumos, no caso em relação a materiais:

$$\left[ \frac{\partial L(\cdot)}{\partial m} \right] m - \lambda \frac{\partial f(\cdot)}{\partial m} = 0 \quad (3.14)$$

rearranjando os termos e multiplicando ambos os lados por  $\frac{m}{q}$ , temos a seguinte expressão:

$$\frac{\partial f(\cdot)}{\partial m} \frac{m}{q} = \frac{p_m m}{q} \frac{1}{\lambda} \quad (3.15)$$

esta expressão implica que a demanda ótima por materiais é satisfeita quando uma firma iguala a elasticidade do produto em relação ao insumo com o participação do insumo (medido em valor) no produto multiplicando o inverso do custo marginal de produção (i.e.,  $\frac{p_m m}{q} \frac{1}{\lambda}$ ).<sup>7</sup>

<sup>6</sup>Uma abordagem moderna sobre esta questão é dada por De Loecker e Warzynski (2012).

<sup>7</sup>Neste ponto, cabe ressaltar que utilizamos capital e trabalho como fatores típicos de uma função de produção. Uma classificação geral organiza os fatores entre variáveis e quase-fixos. Capital é o fator quase-fixo na especificação KLEM, enquanto os demais são variáveis.

### 3.3 Heterogeneidade

A análise dos mercados mostra que as firmas são diferentes. Elas não são diferentes apenas em tamanho, mas também em tipos de produtos ofertados, relação com clientes, entre tantos outros. Uma questão importante é a de que se observa diferenças significantes de performance entre as empresas.

Os trabalhos empíricos mostram que firmas na mesma indústria de tamanho similar possuem performance distintas. Apenas 20% de variância na lucratividade das firmas pode ser explicada por variáveis relacionadas a tamanho, tipo da indústria, etc.<sup>8</sup> Além disso estas diferenças parecem ser bem persistentes.

Uma pergunta que pode ser feita é qual a fonte de 80% da variabilidade na performance das empresas que não é explicada? Por que algumas firmas mantêm sustentada vantagem competitiva?

Explicações tradicionais são impedimentos à imitação, ambiguidade causal e estratégia. Impedimento à imitação pode ocorrer devido à impedimento legais (como patentes) ou a detenção de algum ativo muito específico. Ambiguidade causal está relacionado à dificuldade de se separar fatores que expliquem o sucesso das firmas. Por exemplo, algumas empresas rivais tentam contratar executivos das concorrentes para melhorar a performance empresarial do seu negócio, entretanto em alguns casos essa estratégia não funciona. O insucesso desta estratégia em geral é baseado na impossibilidade de se replicar casos de sucesso pela simples união de outros fatores produtivos em outra empresa. Estratégia está relacionado com o ‘timing’ correto de entrada em um mercado, expansão de capacidade, fusões e aquisições, propaganda, estratégias ilegais, entre outros. *Estratégia* merece destaque especial pois é um dos principais tópicos de estudo em organização de mercados. Empiricamente as fontes de heterogeneidade entre empresas podem ser explicadas por diferenças nas práticas de gestão, qualidade de insumos, capital intangível, estrutura da firma e por diferenças nos produtos e serviços ofertados.

Uma fonte de heterogeneidade entre empresas é a qualidade na gestão. A evidência sugere que a qualidade gerencial é positivamente relacionada com a performance da firma. Além disso, a qualidade gerencial varia consideravelmente entre empresas. Esses resultados foram obtidos em um estudo que se baseou em questionários aplicados a centenas de firmas localizadas em diferentes países.<sup>9</sup>

Na Figura 3.6 mostramos a média da distribuição das práticas gerenciais entre países. O índice de qualidade de práticas gerenciais foca em como os produtores coordenam as atividades produtivas: gerenciamento alinhado aos objetivos da firma; gerenciamento de estoque; gerenciamento da alocação do capital humano; relacionamento com consumidores. Os Estados Unidos possuem a maior pontuação, seguido por Alemanha, Suécia e Japão. Países africanos são os que possuem pontuação mais baixa, enquanto países da América Latina estão na metade inferior do ranking. O Brasil está entre Argentina e

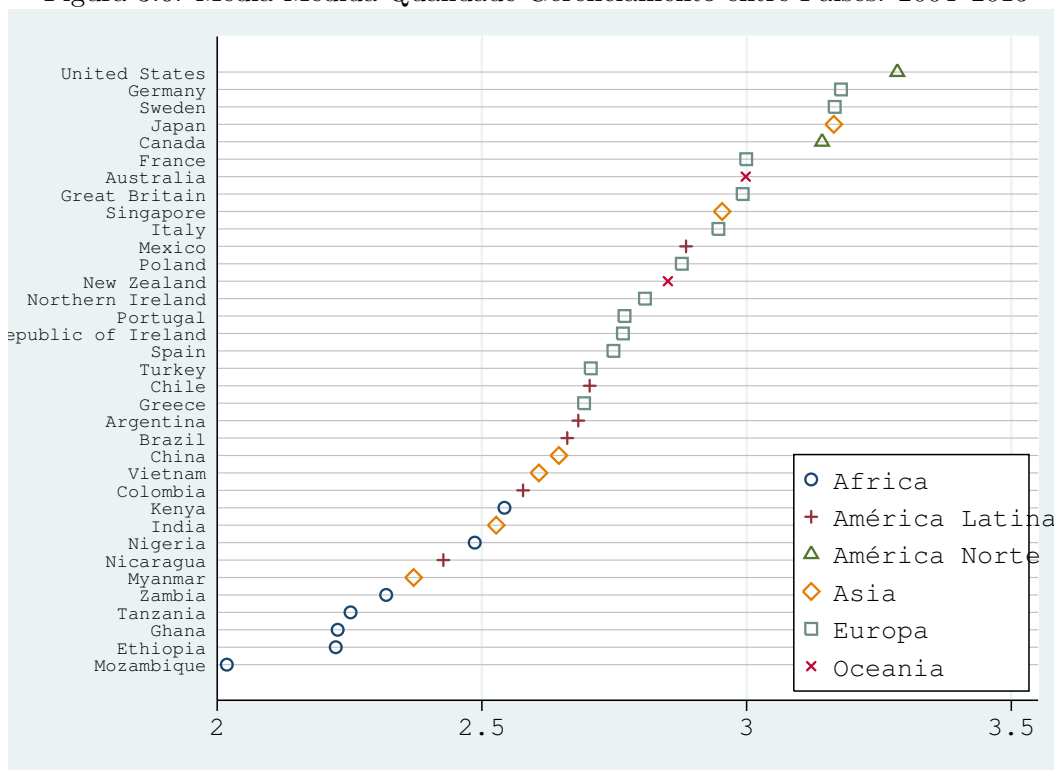
<sup>8</sup>Veja Cabral (2017, p. 56), e Schmalensee (1989).

<sup>9</sup>Veja Bloom e Van Reenen (2007).



China, que apresentam práticas com baixa qualidade, na média.

Figura 3.6: Média Medida Qualidade Gerenciamento entre Países: 2004–2015



Fonte: Bloom, Lemos, Sadun, Scur, e Van Reenen (2014) e World Management Survey.

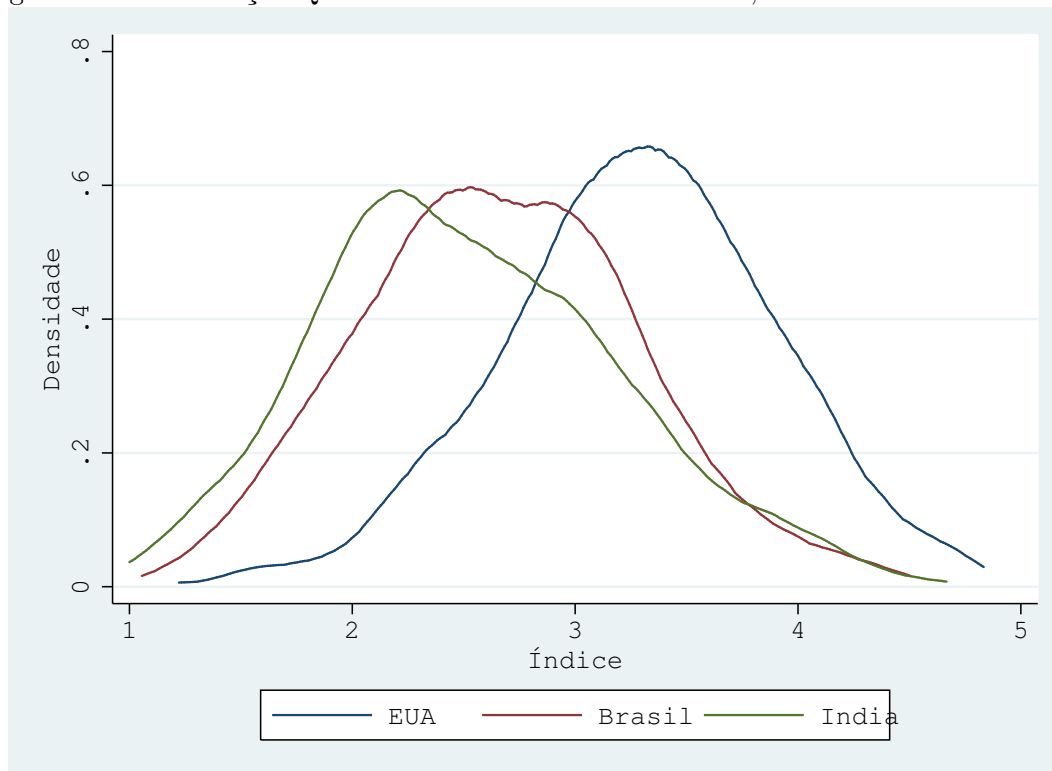
Na Figura 3.7 são apresentados a distribuição dos valores do índice de gerenciamento para EUA, Brasil e Índia. EUA é o líder no ranking, enquanto Brasil e Índia estão no terço inferior. Uma característica marcante da comparação da distribuição é que grande parte das empresas de Brasil e Índia possuem qualidade de gerenciamento similar às piores empresas dos EUA.

Estes dados de qualidade mostram a grande heterogeneidade entre empresas. Entretanto estes dados ainda não são suficientes para mostrar se a qualidade do gerenciamento é importante para a lucratividade e produtividade.

### 3.4 Medidas de Concentração de Mercado

A grande maioria dos mercados encontra-se entre os extremos de concorrência perfeita e monopólio. Como dizer se estão mais próximos de um extremo ou outro? Se as empresas tivessem o mesmo tamanho seria fácil responder a esta questão, pois bastaria medir o número de empresas no mercado. Todavia as empresas possuem tamanhos diferentes. Portanto, para saber se o mercado se aproxima mais da concorrência perfeita ou do

Figura 3.7: Distribuição Qualidade Gerenciamento nos EUA, Brasil e Índia: 2004–2015



Fonte: Bloom, Lemos, Sadun, Scur, e Van Reenen (2014) e World Management Survey.

monopólio podemos utilizar as medidas de concentração descritas aqui.

### Índice Herfindahl-Hirshman

Uma medida muito utilizada é o índice de Herfindahl-Hirshman (HHI) que pode ser definido como

$$HHI = \sum_{j=1}^N s_j^2 \quad (3.16)$$

tal que  $s_j$  é a participação no total do mercado da empresa  $j$  e  $n$  é o número total de empresas. Por exemplo, a participação da empresa no total do mercado é dada simplesmente pelo total do faturamento da empresa  $j$  sobre a soma do faturamento de todas as empresas que operam no mercado. O valor do índice de Herfindahl-Hirshman,  $HHI$ , varia entre  $1/n$  (concentração mínima) e 1 (concentração máxima).

As agências antitruste geralmente utilizam o índice HHI utilizando as participações de mercado como percentual. Neste caso o índice varia entre  $100/n$  e 10000. Agências antitruste consideram mercados com HHI acima de 1800 altamente concentrados. Por exemplo, HHI para um mercado com cinco empresas com participação de mercado idênticas é 2000 ( $5 \times 20^2 = 2000$ ) e para seis firmas é 1667 ( $6 \times 16,67^2 = 1667$ ).

### C4

Outra medida muito comum na literatura econômica é a medida de concentração das quatro maiores firmas em determinado mercado. Este índice é apelidado apenas como C4. Ele é a soma da participação de mercado das quatro maiores firmas:

$$C4 = \sum_{j=1}^4 s_j. \quad (3.17)$$

### Índice de Concentração Geográfica(\*)

Aqui apresentamos um índice de concentração geográfica proposto por Ellison e Glaeser (1997). Suponha que os dados contendo da participação do emprego de uma indústria em cada área geográfica  $M$ . Vamos denotar esses dados de  $s_1, s_2, \dots, s_M$ . Suponha também a existência de dados de emprego total para todas estas  $M$  áreas. Vamos denotar estes dados por  $x_1, x_2, \dots, x_M$ . Além disso, tome um índice de Herfindahl calculado sobre o tamanho da planta de cada empresa, dado por  $H = \sum_{j=1}^N z_j^2$ , tal que  $z$  é a participação da planta de cada empresa sobre o total de plantas no mercado.

Como um índice conveniente do grau de concentração geográfica da indústria, Ellison e Glaeser propõem a medida  $\gamma$ :

$$\gamma \equiv \frac{G - H}{1 - H} \quad (3.18)$$

**Construção da medida  $G$ .** A medida  $G$  é construída a partir de um modelo de concentração geográfica.



## Capítulo 4

# Monopólio, Concorrência Perfeita e Falhas de Mercado

Neste capítulo veremos os modelos básicos de concorrência perfeita e de monopólio.

Concorrência perfeita e monopólio são casos polares, i.e. são importantes como ferramentas de análise. São casos utilizados em muitas situações, mas geralmente os estudos e análises de organização industrial e antitruste lidam com casos de concorrência imperfeita, que são configurações de mercado entre concorrência perfeita e monopólio.

### 4.1 Monopólio

Vamos considerar um mercado com uma única empresa que vende um bem homogêneo. Os monopólios podem existir por diversas razões.

Um monopólio pode ser conquistado (ou recebido) por uma decisão do governo. Por exemplo, o governo francês instituiu um monopólio por decreto para a Electricité de France ser a única geradora de energia do país até 1999 (Cabral, 2017, p. 102). A Petrobras teve monopólio na exploração de petróleo e gás até 1994 no Brasil instituído pela Constituição.

Outra forma do governo criar monopólio é por meio de patentes. Patentes são direitos fornecidos por um conjunto de leis a uma empresa inovadora a ter o monopólio de um produto que ela criou por um certo período de tempo. Um exemplo é a empresa farmacêutica Novo Nordisk que construiu posição dominante no mercado europeu com patentes de drogas para diabetes baseado em tecnologia de produzir insulina a partir de fonte animal.<sup>1</sup> Direitos de cópia e licenciamentos também pode fornecer as empresas monopólio sobre suas marcas e ideias – esse é o caso da Disney, por exemplo.

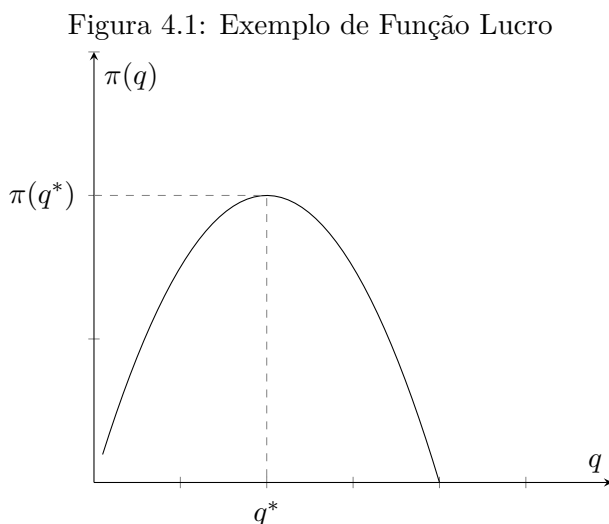
---

<sup>1</sup>Veja Reitzig (2004).

Uma segunda fonte de monopólio está na natureza da estrutura de custo ou na demanda de um produto ou serviço. Por exemplo, a determinação de um padrão tecnológico pode conferir a uma empresa o monopólio sobre um tipo de produto. Isto está associado ao conceito de *efeito de rede*. Um exemplo clássico foi o monopólio construído pela AT&T no mercado de longa distância nos EUA.<sup>2</sup>

#### 4.1.1 Maximização de Lucros

Uma empresa consegue maximizar os lucros ao comparar o benefício marginal de produzir uma unidade a mais com o custo de oferta desta unidade. Nos exemplos introdutórios, geralmente pensamos em uma função lucro côncava em relação a preços e quantidades (certamente, a forma da função lucro depende do mercado, do produto e da estrutura de custos) como apresentada na Figura 4.1. O ponto ótimo de produção é aquele que maximiza os lucros, que no caso seria dado pela quantidade  $q^*$ .



Conceitualmente o ponto de máximo é atingido quando a receita marginal é igual ao custo marginal. A receita marginal é o valor incremental das vendas quando se muda a quantidade produzida em uma pequena fração. Por sua vez o custo marginal é a mudança incremental no custo quando se oferta uma unidade a mais. A firma produz até quando não for mais marginalmente lucrativa a produção.

A seguir apresentamos o processo de maximização de lucros do monopolista. Considere uma firma que lida com a curva de demanda  $D(p)$ . Aqui o preço é determinado pela firma que opera como monopolista. Entretanto, vamos considerar a curva de demanda inversa  $P(q)$ , tal que  $P(\cdot)$  é a função inversa de  $D(\cdot)$ . Neste caso, a firma determina o preço que a permite vender a quantidade  $q$ . Quando produzir  $q$ , a firma incorre no custo  $C(q)$ .

<sup>2</sup>Veja Shapiro e Varian (1999), capítulo 7.

O problema de maximização de lucros é encontrar o preço ótimo em face a curva de demanda e os custos da firma. Uma vez que cada preço está relacionado com uma certa quantidade vendida, é equivalente calcular o preço ótimo ou a quantidade ótima. Vamos apresentar a solução do problema para a quantidade ótima  $q^*$ , e então calcular o preço como função da quantidade:  $p^* = P(q^*)$ .

No caso da função receita escrita para a demanda inversa temos:

$$R(q) = pq = P(q)q$$

Neste caso, o problema da firma é maximizar a função lucro, receita total - custo total:

$$\pi = P(q)q - C(q) \quad (4.1)$$

Encontramos o máximo da função lucro pela sua derivada em relação à quantidade (veja a Figura 4.1). Em outras palavras, a condição de máximo é quando a curvatura da função lucro atinge o maior valor possível e isto ocorre quando  $\frac{\partial \pi(q)}{\partial q} = 0$ . A condição de primeira ordem do problema do monopolista é

$$\left[ \frac{\partial \pi(q)}{\partial q} \right] : P(q) + P'(q)q - C'(q) = 0 \quad (4.2)$$

tal que  $P'(q)$  é a derivada de  $p$  com respeito a quantidade. Vale observar que para calcular o produto marginal é necessário utilizar a regra da cadeia e a do produto. O termo  $C'(q)$  é o custo marginal (*CMg*) do produtor, enquanto que  $P(q) + P'(q)q$  é a receita marginal (*RMg*). Como o valor ótimo de  $q^*$  resulta em nível de preço ótimo,  $p^* = P(q^*)$ , então rescrevemos (4.2) como:

$$p + P'(q)q = C'(q) \quad (4.3)$$

A condição de lucro máximo é simplesmente:

$$RMg = CMg \quad (4.4)$$

ou seja, *o nível de produto que maximiza lucros é tal que a receita marginal é igual ao custo marginal*. Esta condição é satisfeita para a quantidade ótima  $q^*$ .

**Regra da elasticidade.** Rearranjando e dividindo ambos os lados de (4.5) por  $p$ , temos

$$\frac{p - C'(q)}{p} = -P'(q) \frac{q}{p} \quad (4.5)$$

relembre que a elasticidade-preço da demanda é  $\epsilon = \frac{dq}{dp} \frac{p}{q}$ , então  $P'(q) \frac{q}{p} = \frac{1}{\epsilon}$ , então temos a seguinte relação

$$\frac{p - C'(q)}{p} = -\frac{1}{\epsilon} \quad (4.6)$$

A razão da diferença preço-custo marginal sobre o preço é uma forma de expressar a margem de lucro do produto. Esta margem é denominada *índice de Lerner*. Como dividimos a diferença preço-custo sobre o preço, ela é uma medida proporcional de retorno. Assim como escrevemos a margem como função da elasticidade, podemos escrever o preço como função da margem e do custo marginal:

$$p = \frac{CMg}{1 + \frac{1}{\epsilon}} \quad (4.7)$$

Ambas as equações relacionam preços, custo marginal e elasticidade-preço da demanda. Esta “regra de elasticidade” diz que quanto mais baixo for a elasticidade-preço da demanda (em valor absoluto), maior será a margem preço-custo.

Considere uma pequena variação no preço de  $p'$  para  $p''$ , como o lucro muda com o preço? A margem de lucro aumenta de  $p' - CMg$  para  $p'' - CMg$ , este ganho representamos por:  $\Delta p \cdot q''$ . O custo dessa mudança de preços é a perda de vendas, de  $q'$  para  $q''$ . A perda total é de  $-\Delta q(p' - CMg)$ . A pergunta é: o aumento de preços foi vantajosa para o monopolista? Para pequenas variações em preços e quantidades, a mudança de preço é boa para firma se a mudança na receita é maior do que a perda de vendas. Isto pode ser escrito na seguinte condição:

$$-\Delta q(p' - CMg) < \Delta p \cdot q''$$

reorganizando e aproximando os valores de  $q$  e  $q$ , se tem

$$\frac{p - CMg}{p} < -\frac{\Delta p}{\Delta q} \frac{q}{p} \quad (4.8)$$

Se a mudança no preço for infinitesimal, então esta condição se torna

$$m < \frac{1}{-\epsilon} \quad (4.9)$$

tal que  $m \equiv \frac{p - CMg}{p}$ . Isso significa que se a margem,  $m$ , é mais baixa do que o inverso da elasticidade-preço da demanda (em valor absoluto), então um aumento no preço leva a aumento nos lucros. A conclusão tradicional diz que aumentos de preços apenas não são lucrativos quando a margem for igual ao inverso da elasticidade da demanda.<sup>3</sup>

### Definições de Margem e Markup

Além das definições anteriores é comum utilizar a definição de markup, que é definido como:

$$k \equiv \frac{p - CMg}{CMg}$$

Portanto as definições de diferença entre preço e custo são:

<sup>3</sup>Observe que a regra da elasticidade de determinação do markup funciona quando a elasticidade da demanda é constante em todos os pontos da curva de demanda.



**Margem:**  $p - CMg$

**Índice de Lerner:**  $m \equiv \frac{p - CMg}{p}$

**Markup:**  $k \equiv \frac{p - CMg}{CMg}$

A relação do markup com a regra da elasticidade pode ser escrito primeiro a partir da relação entre margem e markup

$$\frac{k}{1 + k} = \frac{\frac{p - CMg}{CMg}}{1 + \frac{p - CMg}{CMg}} = \frac{p - CMg}{p} = m$$

então a relação markup elasticidade pode ser

$$\frac{k}{1 + k} = \frac{1}{-\epsilon} \text{ ou } k = \frac{1}{-\epsilon - 1}$$

como antes esta fórmula torna claro que quando a demanda é inelástica,  $|\epsilon| < 1$ , a regra da elasticidade não se aplica. Neste caso, sempre será ótimo para a firma aumentar seu preço.

### Exemplo

Um tradicional exemplo do problema de maximização do monopolista é quando assumimos função de demanda linear na forma,  $P(q) = a - b \cdot q$ , e custo marginal constante  $c$ . A função receita total neste exemplo é

$$p \times q = (a - b \cdot q)q.$$

Logo o problema de maximização do monopolista análogo a (4.1) e (4.2) é

$$\max_q \{(a - b \cdot q)q - c \cdot q\} \quad (4.10)$$

derivando a função lucro em relação a quantidade temos a condição de primeira ordem

$$a - 2bq - c = 0$$

utilizando o fato de que receita marginal igual a custo marginal no ponto de produção ótima  $q^*$  teríamos:

$$a - 2bq^* = c \quad (4.11)$$

neste caso a quantidade ótima é:

$$q^* = \frac{a - c}{2b} \quad (4.12)$$

retornando a função demanda podemos calcular o preço como função da quantidade ótima (4.12):<sup>4</sup>

$$p^* = a - b \left( \frac{a - c}{2b} \right) = \frac{a + c}{2}. \quad (4.13)$$

Solucionando para o lucro máximo do monopolista:

$$\pi(q^*) = p^* \left( \frac{a - c}{2} \left[ a - c - \frac{c}{b} \right] \right) \quad (4.14)$$

<sup>4</sup>Escrevendo na forma margem preço-custo temos  $p^* = a - \frac{a}{2b} + \frac{1}{2b}c$ .

**Demanda log-linear:** considere um problema similar, mas agora com demanda log-linear. Este tipo de função é muito conveniente para levar para a análise dos dados (veja o capítulo 6). Vamos assumir a seguinte curva de demanda inversa log-linear:

$$\ln p = \ln \alpha + \beta \ln q. \quad (4.15)$$

Retirando o log natural se tem

$$p = \alpha q^\beta$$

Assumindo custo marginal constante, o lucro do monopolista neste caso é

$$\pi(q) = (\alpha q^\beta) \cdot q - c \cdot q.$$

Por sua vez, a condição de primeira ordem para  $\frac{\partial \pi(q)}{\partial q}$  é

$$(\beta + 1)\alpha q^\beta - c = 0$$

Usando a função demanda,  $p = \alpha q^\beta$ , solucionamos para  $p$ . Então temos:

$$p = \left( \frac{\epsilon}{\epsilon + 1} \right) c$$

que é novamente a margem preço-custo marginal escrita na forma de elasticidade-preço da demanda.

### 4.1.2 Monopolista Multi-Produto

A maioria das firmas vendem mais de um produto. Grandes conglomerados vendem muito mais do que um produto. Para uma firma que vende muitos produtos, a análise desenvolvida acima vale se a demanda ou custo não for interrelacionado. Se não há relação entre as demandas, então se pode fazer a análise de mercado isoladamente para cada produto. Por outro lado, se as demandas forem relacionadas então a análise é um pouco mais complexa, pois temos que lidar com a relação de demanda. Isto significa que o preço ótimo de um produto deve levar em conta a determinação do outro produto e vice-versa.

Por exemplo, diversas montadoras vendem veículos grandes, médios e pequenos. Dependendo do preço de cada tipo de veículo, ele pode tirar mercado de outra classe dentro da mesma montadora. Isto é o que a indústria apelida de *canibalismo*. Uma empresa que produz presunto também pode ofertar mortadela e peito de peru. A escolha do preço do presunto pode afetar a demanda pelos outros produtos da firma. Mas nem todos os exemplos são de produtos concorrentes – Gillete vende creme de barbear e lâminas, neste caso estes dois produtos são complementares.

Vamos aproveitar estes exemplos e apresentar a escolha de preço ótimo para um monopolista que vende dois produtos. A princípio não precisamos definir se são complementares ou concorrentes. Este tipo de setup também é muito importante para a análise de fusões horizontais e verticais.

Considere um monopolista que produz os bens  $A$  e  $B$ . Neste exemplo, a firma determina ao mesmo tempo o preço dos dois bens e obtém receita independente para cada um deles (ou seja, eles são vendidos separadamente em um mercado à vista).

Vamos representar  $(p_i, q_i)$  como preço e a quantidade de cada produto  $i = \{A, B\}$ . Como vimos no capítulo 2, os consumidores que compram o bem  $A$  também são influenciados pelo preço do bem  $B$  além do preço do próprio bem  $A$ . Para manter a intuição de efeito do preço, vamos expressar a função de demanda inversa – quantidade como função do preço – para o bem  $A$  e para o bem  $B$  como:

$$q_A(p_A, p_B) = \alpha_A - \beta_A p_A + \phi_{A,B} p_B \quad (4.16)$$

$$q_B(p_A, p_B) = \alpha_B - \beta_B p_B + \phi_{B,A} p_A \quad (4.17)$$

nestas equações  $\beta$  é o principal componente da elasticidade-própria da demanda, e  $\phi_{i,j}$  captura a sensibilidade com que o aumento do preço do outro produto,  $j$ , aumenta a demanda pelo produto  $i$  (tal que  $i = \{A, B\}$  e  $j \neq i$ ).

Assumindo custo marginal constante e idenpedente entre os dois produtos, a expressão para o lucro do monopolista que produz os bens  $A$  e  $B$  é:

$$\pi(p_A, p_B) = \{p_A \cdot q_A(p_A, p_B) + p_B \cdot q_B(p_A, p_B)\} - (c_A \cdot q_A(p_A, p_B) + c_B \cdot q_B(p_A, p_B)) \quad (4.18)$$

Para maximizar a função lucro a firma deve escolher qual o preço ótimo para  $A$  e  $B$ . A CPO (condição de primeira ordem) para o produto  $A$ ,  $(\partial\pi(\cdot)/\partial p_A) = 0$ , é:

$$\alpha_A - 2\beta_A p_A + \phi_{A,B} p_B + \phi_{B,A}(p_B - c_B) + c_A \beta_A = 0.$$

Solucionando para  $p_A$ :

$$p_A = \frac{\alpha_A + \phi_{A,B} p_B + \phi_{B,A}(p_B - c_B)}{2\beta_A} + \frac{1}{2} c_A, \quad (4.19)$$

observe que a margem depende dos parâmetros de demanda pelo produto  $A$ , mas também da demanda e do custo do produto  $B$ . Fazendo  $MKP_B = p_B - c_B$ , a margem do produto  $B$ , e assumindo simetria entre os dois produtos, i.e.  $\phi_{A,B} = \phi_{B,A}$ , então o preço do produto  $A$  é:

$$p_A = \frac{\alpha_A + \phi_{A,B}(p_B + MKP_B)}{2\beta_A} + \frac{1}{2} c_A, \quad (4.20)$$

Portanto, a margem do produto  $A$  depende do preço e da margem do produto  $B$ , além dos parâmetros da demanda por  $A$ . Um parâmetro importante é  $\phi_{A,B}$ , que mede o grau substituição entre os dois produtos. Se este parâmetro for elevado, maior a margem sobre o custo marginal  $c_A$ . A equação de preço para o produto  $B$  é similar.

Observem que este caso é um caso particular, pois a margem também depende da relação com os custos e depende se os bens são complementares ou substitutos. Por outro lado, o resultado apresentando em (4.20) é importante para a análise de fusões em produtos diferenciados.

Uma análise relacionada com firma multiproduto é a de preços dinâmicos. A demanda por filmes, livros, entre outros, é baseada na dinâmica da reputação do produto. Esta análise de firma multi-produto também é a base para a análise do caso do *monopolista de bens duráveis*, que enfrenta concorrência do seu próprio produto ao longo do tempo.

## 4.2 Concorrência Perfeita

Concorrência perfeita é uma definição clássica de indústrias que não possuem barreiras à concorrência. Por ser uma estrutura teórica simples é largamente utilizado como um caso padrão para se comparar estruturas teóricas e resultados empíricos. As características da concorrência perfeita são aplicáveis para alguns mercados, como setores da agricultura e do mercado de trabalho, bem como para contratos longos de alguns produtos.

As principais hipóteses ou características que definem concorrência perfeita são:

**Firmas atomísticas:** no mercado existem muitos competidores, todos possuem tamanho pequeno em relação ao mercado. As firmas não conseguem afetar a formação de preços por meio de suas ações.

**Produto homogêneo:** Competidores produzem os mesmos produtos. Produto homogêneo implica que o preço governa a demanda.

**Informação perfeita:** Firms e consumidores possuem informação perfeita e livre sobre as variáveis de mercado. Entre as variáveis estão preços, características dos produtos, número de firmas, custos, etc.

**Livre entrada e saída:** As firmas podem entrar e sair dos mercados sem barreiras. Isto é, o custo para entrar e sair do mercado não é relevante. Além disso as empresas possuem acesso livre a tecnologia existente, sendo possível uma firma que entrar copiar as “melhores práticas” das firmas que operam nos mercados.

No mercado que opera em *concorrência perfeita* existe intensa competição entre as firmas, tal que os lucros das empresas são os baixos possíveis.

### 4.2.1 Maximização de Lucros

Na concorrência perfeita firmas lidam com o mesmo problema enfrentado por um monopolista, qual seja, maximizar os lucros. Todavia, na concorrência perfeita a empresa oferta produtos ao preço que é praticado no mercado. Como por hipótese a firma não afeta o preço, resta a ela assumir o preço de mercado e ofertar a quantidade de produto faz o lucro ser máximo.

Tecnicamente, a firma concorrencial maximiza a função lucro

$$\pi = P \times q - C(q)$$

com a *observação* de que  $P$  é dado no mercado. Novamente, o ponto ótimo de produção

é dada pela equação (4.4), i.e.

$$RMg = CMg$$

como na concorrência perfeita o preço de mercado é dado,  $P = RMg$ , então a condição de máximo é:

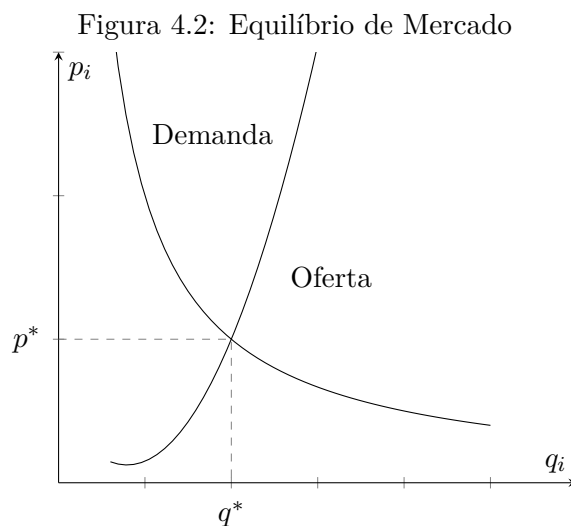
$$P = CMg \tag{4.21}$$

Portanto, o preço de equilíbrio no mercado concorrencial é quando o preço é igual ao custo marginal. Neste, caso essa condição é determinante da quantidade que a firma oferta. Para cada valor de  $p$ , a equação  $CMg = P$  fornece um valor  $q_i$  que é a oferta da firma  $i$ . Colocando todas as firmas junto, somamos as quantidades a cada preço. Assim derivamos a curva de oferta da indústria, que é dada por  $S(p) = q_1 + \dots + q_n$ , para  $i = \{1, \dots, n\}$ .

### 4.2.2 Equilíbrio de Mercado

O equilíbrio de mercado é resultado da interação entre compradores e vendedores. A interseção entre as curvas de demanda e oferta é denominada de *equilíbrio de mercado*.

Na Figura 4.2 é apresentado um exemplo de equilíbrio de mercado. No ponto  $(q^*, p^*)$  as curvas de demanda e oferta se cruzam. Este resultado é equilíbrio no sentido de que nenhum participante no mercado tem incentivo para mudar o seu comportamento.



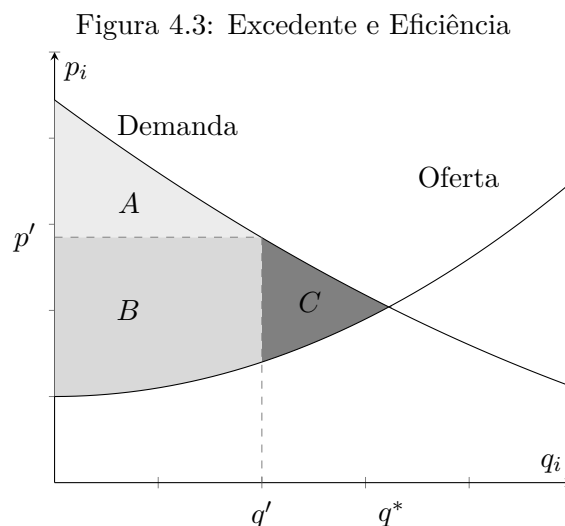
Se o preço de mercado for maior do que o preço de equilíbrio  $p^*$ , menos pessoas desejaria comprar à vender. O excesso de vendedores levaria o preço de mercado para baixo. Por outro lado, se o preço é mais baixo do que o valor de equilíbrio  $p^*$ , menos pessoas estariam dispostas de vender à comprar. O excesso de compradores faria o preço de equilíbrio subir em direção ao preço de equilíbrio. A tendência do preço se mover em direção ao preço de equilíbrio é frequentemente referido como a *lei da oferta e da demanda*.

No mundo real, existem muitos fatores exógenos deslocando curvas de oferta e demanda a todo instante. Por esta razão este ponto de equilíbrio pode estar sempre mudando. Portanto, o ponto de equilíbrio pode ocorrer durante algum tempo, para em seguida mudar para outra vizinhança. Neste sentido, poderia se chamar o ponto de equilíbrio de um ponto de repouso (veja Doraszelski, Lewis, e Pakes, 2018).

### 4.3 Eficiência

Comércio cria excedentes. Se isso não fosse verdade não ocorreria comércio como conhecemos. Neste sentido, **eficiência** é um conceito-chave em economia.

Para compradores a curva de demanda inversa representa a sua disposição à pagar. A diferença entre a curva de demanda e o preço de mercado, i.e. a área *A* na Figura 4.3, é o excedente dos compradores. Ele é chamado de *excedente do consumidor*.



De forma similar, a curva de oferta inversa mede o preço ao qual os vendedores estão dispostos à vender. A diferença entre o preço e a curva de oferta, a área *B* na Figura 4.3 é o excedente dos vendedores, que por sua vez é chamado de *excedente do produtor*. A soma das áreas *A* e *B* é o excedente total – que é o excedente total gerado pelo comércio e pela produção.

Existem tipos diferentes de eficiência. A primeira que podemos definir é a *eficiência alocativa* – ela requer que os recursos sejam alocados ao seu uso mais eficiente. Vamos pensar que a Figura 4.3 descreve um mercado de automóveis. Ao preço  $p'$ ,  $q'$  unidades são vendidas. Isto implica que existe um número de consumidores que desejaria comprar um veículo mas não o faz pelo preço cobrado. Estes consumidores possuem disposição à pagar menor do que  $p'$ . Uma parte dos consumidores estariam dispostos a pagar preço superior ao custo de produção. O grupo destes consumidores está entre os preços  $q'$  e

$q^*$ , tal que  $q^*$  é o nível de produção tal que o custo marginal é igual a disposição à pagar. Agregando entre estes consumidores, o custo de *ineficiência alocativa* quando o produto é  $q'$  é dado pela área  $C$ .

No capítulo 3 analisamos a produtividade das firmas. Por exemplo, a produtividade média de uma firma americana é maior do que a média das firmas na Índia operando na mesma indústria. As chances são que as firmas indianas operem com custos maiores do que as firmas americanas – quando valoramos os insumos ao mesmo preço. As chances são que as firmas indianas sejam menos produtivas do que as americanas (veja capítulo 3). A baixa produtividade é resultante do uso equivocado do mix de insumos ou de fazer uso sub-ótimo do mix de insumos. A baixa produtividade implicaria em uma curva de custo marginal mais elevada no setup da Figura 4.3. Em uma definição geral, a *eficiência produtiva* se refere a quão perto está o custo de produção corrente do mais baixo que pode ser alcançado.

Estes dois conceitos são de eficiência estática. Em alguns casos a eficiência dinâmica é mais importante. Isto é verdade quando a taxa de introdução de novos produtos, bem como aumento na eficiência produtiva, é a base eficiência dinâmica de uma indústria. Dito de outra forma, na eficiência dinâmica estamos interessados na evolução da eficiência produtiva (veja a análise de produtividade no capítulo 3).

**O Teorema Fundamental:** *em um equilíbrio de mercado competitivo, os níveis de produto e preços correspondem ao excedente total máximo.*

O Teorema Fundamental trata de eficiência estática, tanto a eficiência alocativa e produtiva. A eficiência produtiva é alcançada porque as empresas ineficientes deixam o mercado. A força da concorrência encaminha as empresas ineficientes para fora do mercado.

Na Figura 4.3 nosso exemplo mostra uma alocação  $(q', p')$  sub-ótima, nos termos do Teorema Fundamental. A alocação  $(q', p')$  tem como excedente total a área  $A + B$ .

A alocação ótima é atingida pela concorrência perfeita, pois os lucros são zero para todas as empresas. Isto é, quando o preço for igual ao custo marginal. Na Figura 4.3 o equilíbrio competitivo será aquele em que a quantidade ótima for  $q^*$ , com o preço correspondente (não representado na figura). Neste caso, o excedente total do equilíbrio competitivo é  $A + B + C$ , e claramente  $A + B + C > A + B$ .

O excedente total  $A + B + C$  é máximo que pode ser atingido. Em outras palavras, o equilíbrio máximo é atingido porque qualquer outra configuração leva a excedente total menor, bem como a preços maiores e quantidades menores. Por outro lado, se  $q > q^*$ , então existem consumidores que possuem disposição à pagar menor do que o custo marginal de oferta do produto. Portanto, o equilíbrio acima do ótimo não é factível.

Alguns comentários adicionais sobre o Teorema Fundamental. O Teorema fala de

eficiência e não equidade. Isto é, o foco é no excedente total e não como ele é distribuído entre consumidores e empresas.

O outro ponto importante é o de que o Teorema trata de concorrência perfeita. Em grande parte dos mercados existem produtores que afetam o preço, produtos são diferenciados, entrada pode ser restrita, entre outros. Neste caso, temos casos em que a configuração de mercado não é de concorrência perfeita. Apesar de útil, o resultado do Teorema Fundamental não é atingido nestas condições de *concorrência imperfeita*.

Por último, o Teorema Fundamental não trata de eficiência dinâmica – o Teorema trata de eficiência estática. É difícil mensurar a eficiência dinâmica, ainda mais difícil é comparar a eficiência dinâmica com a estática.

**Eficiência e Monopólio:** Na introdução definimos poder de mercado e descrevemos como chave para o entendimento do funcionamento dos mercados. O poder de mercado em análise de economia industrial é a capacidade da firma determinar margem acima do custo marginal. Vimos estas definições na seção 4.1.

Suponha dois mercados com a mesma demanda e a mesma curva de custo marginal.<sup>5</sup> A diferença entre os dois mercados é que no primeiro existe uma configuração de concorrência perfeita, com muitas pequenas empresas, enquanto que no segundo existe apenas uma firma.

Em concorrência perfeita o equilíbrio é determinado pela interseção entre demanda e oferta. Como a curva de demanda inversa mede a disposição à pagar do consumidor e a curva de oferta mede o custo marginal, o ponto de equilíbrio é aquele em que a disposição à pagar é igual ao custo marginal. Toda a atividade comercial em um mercado competitivo são tais que a disposição à pagar é maior do que o custo marginal. Esta é a essência do Teorema Fundamental.<sup>6</sup>

Na Figura 4.4 o equilíbrio de concorrência perfeita é descrito por  $q^*$  e  $p^*$ , onde a curva de oferta intercepta a curva de demanda. O equilíbrio de monopólio são os pontos  $q^M$  e  $p^M$ , onde a receita marginal é igual ao custo marginal. Observe nesta figura que a quantidade ofertada pelo monopólio é menor do que a quantidade do mercado concorrencial e o preço do monopólio é mais elevado do que na configuração de concorrência perfeita. Além disso, todas as outras possibilidades de configuração de mercado,  $q^M < q < q^*$ , não podem ser atingidas por um monopólio.

Portanto, dizemos que ocorre perda de eficiência quando comparamos os resultados de equilíbrio de mercado do monopólio em relação à concorrência perfeita. Esta perda de eficiência é a área  $C$  na Figura 4.4, que a soma de todos os excedente  $p - CMg$  que se perde com transações mais eficientes do que o monopólio.<sup>7</sup> Esta área é conhecida em

<sup>5</sup>Não é difícil de imaginar uma situação como esta. Pense no mercado de gasolina em cidades similares.

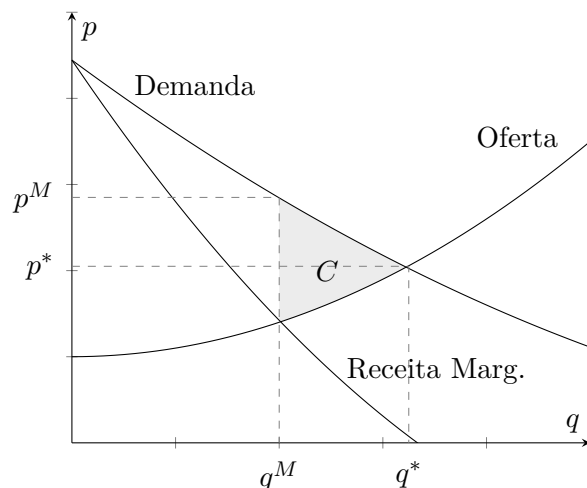
<sup>6</sup>Veja Cabral, 2017, capítulo 5 para mais detalhes.

<sup>7</sup>Esta área é frequentemente denominada de triângulo de Harberger, embora somente seja um triângulo quando as curvas de demanda e oferta são lineares.



economia como *perda do peso-morto*.

Figura 4.4: Concorrência Perfeita e Monopólio



Um ponto importante na comparação destas duas situações de mercado é a transferência de renda dos consumidores para os produtores. Como no monopólio a margem de lucro ou *markup* é o maior possível, além da perda de eficiência também ocorre a transferência de riqueza dos consumidores para o produtor (no caso de monopólio).

Como um monopolista não precisa competir, estas empresas tendem a fazer menos esforço para redução dos custos de produção. Como dito por John Hicks (1935, p. 8), “o melhor do monopólio é a vida tranquila.” Este problema, ou resultado da falta de competição, é chamado de eficiência-X. Este termo foi cunhado por Leibenstein (1966), e diz basicamente que um monopólio tem baixa *produtividade total dos fatores*.<sup>8</sup> A evidência sugere que firmas em mercados mais competitivos são também mais produtivas; e grande parte desta variação é contabilizada pela eficiência-X.

Estimar o tamanho da diferença entre preço e custo marginal é geralmente o lugar em que economistas de organização industrial começam para analisar o poder de mercado. A diferença entre preço e custo marginal é o *markup* e ele mede a habilidade de um produtor praticar preços acima do custo marginal de produção.

Então, o poder de mercado do monopólio implica em perda de eficiência produtiva, isto é, que o produto total não é produzido ao custo mais baixo que seria possível. Por outro lado, na concorrência perfeita observamos firmas ineficientes serem obrigadas a deixar o mercado.

Um outro problema associado ao monopólio é o *rent seeking*. O conceito de renda aqui é o de extração de excedente da sociedade. Neste sentido, o monopólio cria renda extra extraindo excedente do consumidor.

<sup>8</sup>Veja também Holmes e Schmidt (2010) para uma resenha, bem como o trabalho de Backus (2020).

## 4.4 Antitruste

A forma mais antiga de ação empresarial que foca em obter vantagem competitiva é algum tipo de comportamento de conluio, e entre a ação mais comum do conluio é o acordo de preços. Considere o exemplo do mercado de leilão de obras de artes. Christie's e Sotheby's controlam quase a totalidade deste mercado. Embora essa indústria não seja um monopólio, existe a tentadora possibilidade destas duas empresas se comportarem como um monopólio. De fato elas podem fazer isto se decidirem determinar preços como se fossem um monopolista. Isto não é um simples exercício de análise, de fato as empresas foram condenadas por este tipo de prática.<sup>9</sup> No capítulo 13 vamos analisar com mais detalhes comportamentos colusivos.

Muitos monopólios foram criados por meio de fusões e aquisições (M&A). Portanto, uma parte significativa do antitruste é evitar excessiva concentração nos mercados – principalmente se estes movimentos de M&A são feitos para redistribuir excedente como no caso do monopólio. Nos capítulos 10 e 11 vamos analisar os efeitos das fusões nas variáveis de mercado.

Em último lugar, a firma ser monopolista ou ter posição dominante não é ilegal. Entretanto, utilizar da sua posição de mercado para afetar preços, vendas ou contratos é ilegal. Grandes empresas, tais como IBM, Google, Microsoft, são constantemente monitoradas por autoridades antitruste em diferentes jurisdições.

Em resumo, as principais áreas do antitruste ou defesa da concorrência são:

- comportamento colusivo;
- análise de fusões e aquisições; e
- abuso de posição dominante.

## 4.5 Informação Imperfeita

A existência de estruturas de mercado com poder de estabelecer margens sobre o custo representa um desvio dos resultados do Teorema Fundamental. Uma interpretação desta configuração é quando se muda a hipótese da concorrência perfeita de atomicidade das firmas. Outra hipótese importante para o modelo de competição pura é a existência de informação perfeita. Relaxando a hipótese de informação perfeita também desviamos do resultado da alocação ótima de concorrência perfeita.

Em resumo, existem três configurações básicas para informação assimétrica:

1. seleção adversa;
2. perigo moral;

---

<sup>9</sup>Veja “Christie's, Sotheby's Agree to Pay \$512 Million Collusion Settlement.” Wall Street Journal, matéria de Vanessa O'Connell, 25 sept 2000.

3. .

O Teorema Fundamental assume informação perfeita, portanto, compradores e vendedores possuem informação plena sobre custos, preços, e mesmo qualidade do produto.



## Capítulo 5

# Determinação de Preços: Produto Homogêneo

Neste capítulo vamos apresentar os modelos clássicos de oligopólio no espaço das quantidades, ou seja, de produtos homogêneos. Embora ainda limitados para análise empírica, os conceitos de equilíbrio de Nash em quantidades e Nash em preços são a ferramenta padrão na análise estática de mercados.

### 5.1 Preliminares

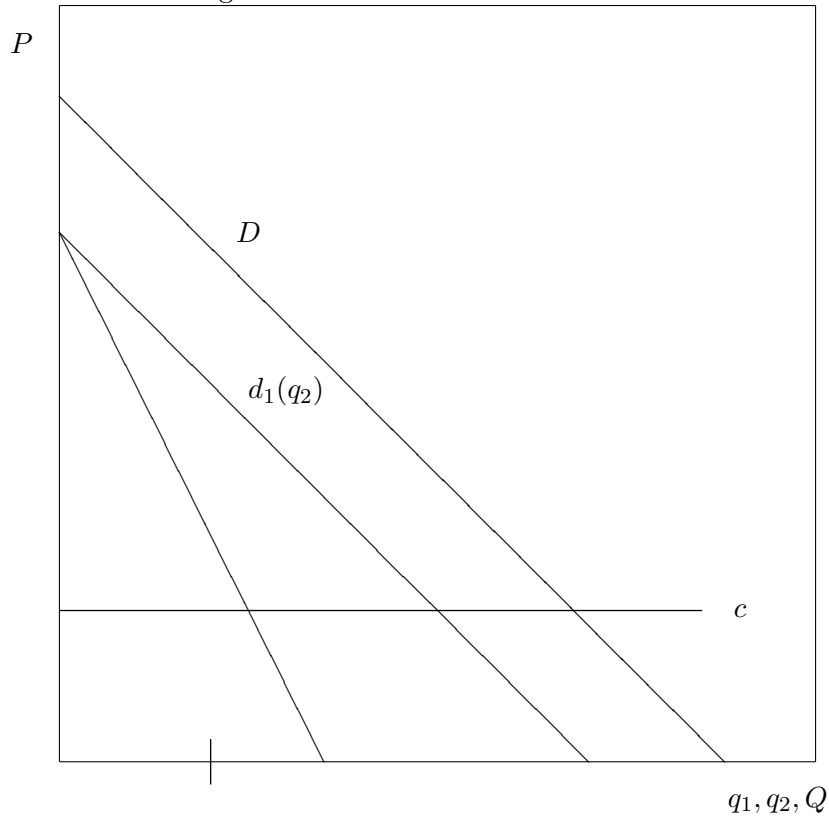
Equilíbrio de Nash de jogo não-cooperativo. Um par (conjunto) de estratégias constitui um equilíbrio de Nash se nenhum jogador pode unilateralmente mudar a sua estratégia de forma que melhore o resultado (*payoff*).

### 5.2 Escolha Estratégica de Quantidades: Cournot

Vamos apresentar primeiro o modelo de Cournot. Como de praxe vamos começar com o caso de um duopólio e em seguida generalizar para o modelo com um número qualquer de empresas.

As hipóteses fundamentais subjacentes ao modelo básico de Cournot são que o preço de mercado resulta da oferta agregada das firmas e a decisão das empresas é simultânea. A estrutura teórica do jogo é baseada na escolha de quantidade por cada firma e estas quantidades são escolhidas simultaneamente. Assim, o lucro de cada firma é função desta escolha de quantidade e do preço de mercado. Como adotado por Cabral (1994), apresentamos primeiro a derivação geométrica e depois a algébrica do modelo de Cournot.

Figura 5.1: Escolha Ótima da Firma 1



### 5.2.1 Derivação Geométrica

Vamos considerar o problema da firma 1 e da firma 2 de decisão simultânea da produção. Vamos supor que a firma 1 tem a expectativa de algum nível de produção da firma 2. Vamos representar este nível de produção da firma 2 como  $q_2$ . Então o problema da firma 1 é a escolha da produção face uma demanda residual. Portanto, o problema de maximização de lucros da firma 1 é similar ao problema de um monopólio com uma demanda residual,  $D_1(q_2) = A - q_2$ , veja Figura 5.1. Dada uma curva de custo marginal, basta derivar a receita marginal e resolver  $Rmg = Cmg$  e determinar a quantidade de produto ótimo da firma 1,  $q_1^*(q_2)$ .

Note que esta escolha de quantidade ótima é condicional a um valor determinado de  $q_2$ . Isto significa que a firma 1 muda a sua escolha ótima para cada escolha ótima da firma 2. A função  $q_1^*(q_2)$ , que relaciona estas escolhas ótimas entre as diferentes expectativas entre as empresas é denominada de *função de reação* ou *função melhor-resposta* da firma 1 em relação a firma 2.

Para derivar a função de reação é interessante considerar dois casos extremos em relação a  $q_2$ . Suponha que  $q_2 = 0$ , então a procura residual da firma 1 é igual a demanda total do mercado. A reação ótima desta firma é produzir a quantidade de monopólio. Agora, suponha que a firma 2 produz ao nível de concorrência, isto é,  $q_2 = Q^C$ . Neste caso, a

demanda residual é sempre inferior ao custo marginal. Portanto, o ótimo para a firma 1 é não produzir (sair do mercado).

Para o caso de curvas de demanda e de custo lineares, podemos dizer que a função de reação também será linear. Assim, com base na discussão do parágrafo anterior podemos derivar a função de reação da firma 1 e da firma 2.

### 5.2.2 Modelo

Vamos considerar primeiro as definições para o modelo de oligopólio de Cournot. Definimos  $Q = \sum_j q_j$ , tal que  $q_j$  é o produto da firma  $j$ ; uma função demanda inversa de mercado  $p(Q)$ ; e  $C_j(\cdot)$  e  $CMg_j(\cdot)$  são as funções custo e marginal, respectivamente.

Como a curva de custo marginal é indexada por  $j$ , podemos permitir que o custo marginal seja diferente entre empresas. Fazendo isso podemos fazer análise de alocação eficiente. Por outro lado, estamos utilizando uma hipótese de curva de demanda agregada,  $p(Q)$ . Esta não é uma aproximação ruim para avaliar eficiência, embora não nos permita olhar para a distribuição dos consumidores.

A função lucro,  $\pi(\cdot)$ , de uma firma  $j$  que opera neste oligopólio é:

$$\pi_j(q_j) = p(Q)q_j - C_j(q_j) \quad (5.1)$$

Assumindo diferenciabilidade e que todas as firmas produzem quantidades positivas, a condição de primeira ordem para o equilíbrio Nash-Cournot é um sistema de  $J$  equações com  $J$  incógnitas:

$$p(Q) + \frac{\partial p}{\partial Q} q_j - CMg_j(q_j) = 0 \quad (5.2)$$

**Condição para equilíbrio:**  $q > 0$  não é a única condição necessária para um equilíbrio. Para este sistema ser um equilíbrio, a escolha de  $q_j$  deve satisfazer a condição de segunda ordem:

$$2 \frac{\partial p}{\partial Q} + q_j \frac{\partial^2 p}{\partial^2 Q} - \frac{CMg_j}{\partial q_j} < 0 \quad (5.3)$$

As condições conjuntamente suficientes para isto são que a receita marginal seja negativamente inclinada e a curva de custo marginal seja positivamente inclinada.

É preciso observar que estas condições mudam na presença de custos fixos ou custos afundados – quando talvez não seja ótimo fechar uma planta quando não tiver produção. Quando existirem restrições de capacidade, que é o caso em que a derivada da função lucro pode ainda ser positiva a medida que aumentamos a restrição de capacidade.

Vamos comparar o resultado do equilíbrio de Cournot com monopólio e concorrência perfeita. Na Figura 4.4 mostramos o problema da perda de eficiência alocativa quando desviamos da concorrência perfeita. O pior cenário para um mercado é quando existe monopólio. Nesta configuração o sistema econômico é menos eficiente. O equilíbrio Nash-Cournot entrega um resultado intermediário. A eficiência do mercado é maior

quando mais empresas estão ativas. Importante chamar atenção que se a configuração de equilíbrio tiver um grande número de firmas (também dependendo das curvas de demanda e oferta) a eficiência alocativa se aproxima muito do resultado de concorrência perfeita.

Uma implicação deste resultado do modelo de Cournot é que a consolidação de um setor pode implicar em dois resultados. Uma onda de consolidação provavelmente significa em aumento de preços para o consumidor – preço e quantidade vão em direção à configuração de monopólio. Por outro lado, essa nova configuração se traduz em maior produtividade ou eficiência alocativa das firmas, que se traduz em aumento do excedente do produtor (pelo menos entre as firmas envolvidas na fusão/aquisição).

**Consolidação** é um termo usado para descrever quando a oferta um setor/mercado passa de uma configuração de mais empresas para menos empresas. A consolidação pode ser devido à saída de muitas empresas (menos eficientes, por exemplo) – efeito este que pode ser chamado de “shake-out.”<sup>1</sup> A consolidação também pode ser devido a uma onda de fusões/aquisições em um setor. Recentemente, tem se observado uma consolidação de petrolíferas que operam *shale gas* ou *oil sands* nos EUA e Canadá.<sup>2</sup>

O termo de preço é muitas vezes apresentado como o Índice de Lerner (relembre da seção sobre margens no capítulo 4):

$$p(Q) = CMg_j - q_j \frac{\partial p(\cdot)}{\partial Q}$$

multiplicando por  $Q/Q$  podemos rescrever a CPO na forma do Índice de Lerner, temos:

$$\frac{p - CMg_j}{p} = s_j \frac{1}{\eta} \tag{5.4}$$

tal que  $s_j$  é a participação de mercado da firma  $j$  ( $s_j = q_j/Q$ ), e  $\varepsilon$  é o valor absoluto da elasticidade da demanda. Observe que o tamanho  $s_j$  de cada empresa é função do custo marginal.

A partir da equação 5.4, podemos construir a margem preço-custo média do mercado:

$$\frac{\sum_j s_j (p - CMg_j)}{p} = \sum_j s_j^2 \frac{1}{\varepsilon} \tag{5.5}$$

Como  $\sum_j s_j^2$  é o índice Herfindahl-Hirshmann (HHI), veja seção 3.4, temos

$$\frac{\sum_j s_j (p - CMg_j)}{p} = \frac{HHI}{\eta} \tag{5.6}$$

que mostra que o Índice de Lerner é positivamente relacionado à concentração do mercado e a demanda menos elásticas. Dito de outra forma, um resultado que emerge do

<sup>1</sup>Sobre onda de saída de empresas veja a descrição em Sutton, 1997.

<sup>2</sup>Veja Denning (2017) e Gruley et al (2020).



modelo de Cournot é de que mercados concentrados e demanda inelásticas implicam que maior lucratividade média. Este é um típico resultado da antiga organização industrial de Harvard (Bain, Mason, etc.).

**Comportamento Estratégico.** No equilíbrio Cournot-Nash as empresas fazem o que é melhor para elas em termos de maximização de lucros. Se existe um equilíbrio único neste mercado, então cada empresa tem uma *função de reação* que define sua estratégia ótima de equilíbrio, i.e. em relação a se manter competitiva frente às decisões dos rivais.

A função de reação é

$$q_j = r_j(q_{-j}) \quad (5.7)$$

tal que  $r_j(\cdot)$  é a função de reação da firma  $j$  e  $-j$  indexa todas as demais firmas que não seja a  $j$ .

### 5.2.3 Exemplo: Firmas Simétricas

Um exemplo intuitivo das implicações do modelo de Cournot é quando assumimos firmas simétricas e demanda linear. Neste exemplo fica mais clara a afirmação anterior de que quanto mais firmas operam em um oligopólio mais próximos estamos do resultado da concorrência perfeita.

Assuma uma função de demanda inversa de mercado como  $p(Q) = a - bQ$ . A função custo total de cada firma é  $C(Q) = c.q_j$ , tal que  $c$  é o custo marginal. Observe que o custo marginal é o mesmo entre as  $n$  firmas neste mercado (o custo marginal não é indexado por  $j$ . A quantidade é indexada por  $j$ , mas será irrelevante pois as firmas tem tamanhos idênticos sob  $c$ ). O total de vendas no mercado é suprido pelas  $n$  firmas:  $Q = q_1 + q_2 + \dots + q_n$ .

O lucro da firma  $j$  é:

$$\pi_j() = P(Q)q_j - C(q_j) = (a - bQ)q_j - cq_j \quad (5.8)$$

Maximizando esta função lucro,  $(\partial\pi(\cdot)/\partial q_j) = 0$ , temos a CPO (observando que  $q_j \in Q$ )

$$-bq_j + a - bQ - c = 0 \quad (5.9)$$

como por construção as firmas tem o mesmo tamanho, pois o custo marginal é idêntico, podemos escrever  $Q = n.q_j = n.q$ . Substituindo na CPO, temos:

$$-bq + a - bnq - c = 0$$

solucionando para  $q$  temos:

$$q = \frac{a - c}{(n + 1)b}. \quad (5.10)$$

Substituindo (5.10) na função demanda obtemos o preço de mercado

$$p = \frac{a + nc}{n + 1}. \quad (5.11)$$

Observe nas equações de quantidade (5.10) e preço (5.11) o papel do número de empresas  $n$ . A medida que aumentamos o número de firmas ( $n \rightarrow \infty$ ), o preço converge para  $c$  ( $p \rightarrow c$ ) e a quantidade fica pequena para cada firma. Este resultado aproxima o modelo de concorrência perfeita.

### 5.2.4 Exemplo: Indústria Naval

A indústria de construção de navios é global e dominada por três países com mais de 90% da produção: Japão, Coreia e China. Europa é a quarta força, mas como participação declinante. Em 2006, a China lançou mão de diversos subsídios para expandir a sua produção de navios transportadores de carga (). Entre 2006 e 2013, a China subsidiou a entrada de novos estaleiros, linhas de produção, bem como investimentos, totalizando o total de 91 bilhões de USD. A análise de Kalouptsidi (2018, 2021) mostrou que com o suporte do governo Chinês, o market share de navios de carga aumentou em 42% (principalmente em *bulk carriers*). Em termos absolutos a política de subsídio aumentou a produção chinesa mundial em 30%, sendo que o restante do crescimento produtivo foi pela tomada de mercado de outros produtores. Durante este período, o market share da Coreia caiu de 48% para 39% e do Japão de 23% para 20%. A expansão de mercado Chinês foi impulsionada principalmente por estaleiros ineficientes, uma vez que o aumento de lucratividade foi muito modesto.

Para analisar o impacto da política de subsídios em um modelo teórico, Kalouptsidi (2018) desenvolveu um modelo de Cournot para este caso. Aqui adotamos uma versão modificada do modelo de Cournot – essencialmente ela analisa o problema dos estaleiros e aqui vamos assumir que o problema da firma é o problema do país: {China, Coreia, Japão}. Por simplicidade podemos utilizar a função de demanda inversa por navios novos como:  $p = a - bQ$ ,  $Q = \sum_j q_j$ . O problema de maximização de lucro de cada país é:<sup>3</sup>

$$\max_{q_j} \{(a - bQ)q_j - c_j q_j\}. \quad (5.12)$$

Calculando a condição de primeira ordem ( $\partial \text{lucro}_j / \partial q_j$ ) e solucionando para o custo marginal temos:

$$c_j = a - b(Q_{-j}) - 2bq_j, \quad (5.13)$$

tal que  $Q_{-j}$  é a soma das quantidades dos demais países exceto o  $j$ . A estimativa completa deste tipo de modelo é apresentada na seção 6.5, mas aqui podemos fazer cálculos simples para mostrar a aplicação do modelo. Para estimar o custo marginal por país, precisamos do volume para cada mercado ( $q_j$ ), e dos parâmetros  $a$  e  $b$  da função de demanda inversa. O parâmetro  $b$  tem sido utilizado como  $b = 0.42$ , com o preço médio por navio de 17,8 milhões de USD, e produção trimestral 19 mil unidades até 2006.<sup>4</sup> Por fim, utilizando os dados de market share antes e depois de 2006, podemos computar os volumes para cada mercado.

<sup>3</sup>Kalouptsidi (2018), seção 3.1, apresenta o problema para cada estaleiro  $j$  localizado em cada país. Seguindo esta ideia, nenhum estaleiro tem capacidade de afetar o preço de mercado, implicando que eles são tomadores de preço.

<sup>4</sup>Veja Cabral (2017, p. 213). Com estes valores podemos utilizar  $a = 17,8 + 0,42(19) = 25,78$ .

Dadas estas informações calculamos o custo marginal para Japão, Coréia e China. O custo marginal até 2006 dos dois primeiros era de 16,36 e 14,36 milhões de USD, implicando em margens (índice de Lerner) de 8% e 19%, respectivamente. No caso da China o custo marginal era de 16,28 milhões de USD e margem de 8,5%. Em relação ao subsídios aplicados pela China, Kalouptsidi (2018) calcula o custo marginal da China entre 2007 e 2013 (a preços constantes) e compara o valor com o custo marginal anteriormente estimado. Naturalmente, o artigo da autora é bem detalhado e com estimativas e contra-factuais extensivos, mas a essência é estimar o custo marginal antes da intervenção e comparar com o custo marginal após a política de subsídios. Supondo tudo o mais constante e mudando o market share da China de 0,24 para 0,5, calculamos o custo marginal após 2006 em 13,09. Comparando as duas estimativas de custo, obtemos a diferença de 19%, que é efetivamente o tamanho do subsídio aplicado pela China. Este número representa o subsídio, pois sem ele os estaleiros chineses teriam amargado prejuízos, isto é, somente foi possível aumentar a produção pelo canal do subsídio para compensar o eventual prejuízo.<sup>5</sup>

### 5.3 Escolha Estratégica de Preços: Bertrand

O uso de preços como variáveis estratégica provavelmente é mais comum do que a escolha de quantidades. A demanda dos produtos de uma firma depende da sua escolha de preço, bem como dos preços escolhidos pelos seus rivais. Para representar esta situação de mercado utilizamos o modelo de Bertrand, que é um modelo de concorrência em preços.<sup>6</sup>

Nesta seção apresentamos o modelo de Bertrand no espaço das quantidades. O modelo básico de Bertrand não é muito utilizado em trabalhos empíricos, mas com alterações o modelo é a base para se analisar produtos diferenciados e comportamento de conluio. Devido a suas especificidades é usual começar a exposição no caso de duopólio com produto homogêneo.

#### 5.3.1 Duopólio Bertrand

Assim, assumimos que existem duas firmas no mercado para um produto homogêneo, cujo preços são escolhas simultâneas. Por simplicidade assumimos custo marginal,  $c$ , idêntico entre as empresas. Como assumimos produtos homogêneos para as duas firmas, estes bens são substitutos perfeitos. Isto implica que a firma com preço menor permanece sozinha no mercado, i.e. se torna monopolista. Podemos descrever este argumento da seguinte forma: se  $p_1$ , o preço determinado pela firma 1, for menor do que  $p_2$ , o preço determinado pela firma 2, então a demanda da firma 1 é a demanda de mercado,  $D(p_1)$ , e a demanda da firma 2 é zero (inexistente). Se ambas as firmas determinam o mesmo

<sup>5</sup>Este número é consistente com o obtido por Kalouptsidi (2018), que foi entre 13-20%, dependendo do modelo. Além disso, ela mostra que a expansão de atividade foi baseado em empresas ineficientes, que é uma história relacionada com baixo lucro dos estaleiros chineses.

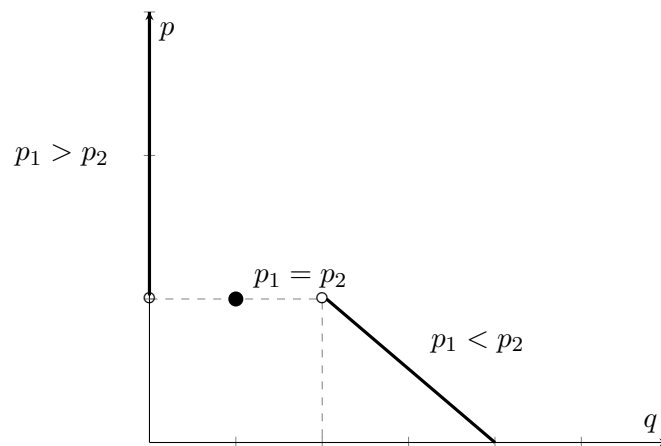
<sup>6</sup>Os fundamentos deste modelo estão na revisão crítica do livro de Cournot (1838) escrito por Joseph Bertrand em 1883.

preço  $p_1 = p_2 = p$ , então elas dividem a demanda do mercado. Se a firma cobrar mais do que sua rival então ela não possui demanda para seu produto. Resumindo, a demanda de mercado é  $D(p)$ , e se  $D_1(p_1, p_2)$  é a demanda para o primeiro produtor:

$$D_1(p_1, p_2) = \begin{cases} D(p_1) & \text{se } p_1 < p_2, \\ (1/2)D(p_1) & \text{se } p_1 = p_2, \text{ e} \\ 0 & \text{se } p_1 > p_2. \end{cases} \quad (5.14)$$

Na Figura 5.2 apresentamos a curva de demanda inversa (5.14). Observe que esta curva de demanda é discreta e apresenta as três configurações: (i) firma 1 como monopolista,  $p_1 < p_2$ ; (ii) duopólio - firmas dividem o mercado,  $p_1 = p_2$ ; (iii) firma 1 não vende,  $p_1 > p_2$ . A curva de demanda inversa para a firma 2 é simétrica.

Figura 5.2: Demanda Bertrand para Firma 1



A função lucro para a firma 1 com custo marginal simétrico é:

$$\pi_1(p_1, p_2) = D_1(p_1, p_2)(p_1 - c) \quad (5.15)$$

Qual é a melhor estratégia da firma 1 neste contexto? Se a firma 2 escolhe um preço maior do que o custo marginal e menor do que o preço de monopólio, a melhor escolha para firma 1 será determinar o preço menor do que o preços da firma 2,  $p_2$ . Se  $p_1 < p_2$  então a empresa 1 recebe lucros de monopólio,  $\pi_1 = (p_1 - c)D(\cdot)$ . Caso contrário se a firma 1 fizer seu preço igual a firma 2,  $p_1 = p_2$ , elas dividem o lucro do mercado, i.e.  $\pi_1 = (p_1 - c)D(\cdot)/2$ , e no caso extremo de preço maior do que o da firma 2, os lucros da firma 1 serão zero pois ela será expulsa do mercado. Como este procedimento é simétrico, ou seja, também é adotado pela firma 2, o melhor que as firmas podem fazer é

$$p_1 = p_2 = c.$$

Neste caso o lucro resultante será zero. Este é o equilíbrio Nash-Bertrand, que é o mesmo resultado na configuração com concorrência perfeita.

Observe que devido a “geometria” da função demanda não podemos usar a condição de primeira ordem tradicional (como em monopólio e Cournot). O equilíbrio tem que ser calculado como acima – para cada  $p_2$  escrevemos um comportamento para  $p_1$ . A função de reação ou *melhor-resposta* para cada firma é uma função  $p_j^*(p_{-j})$  que representa a melhor escolha de preço da firma  $j$  dada a escolha de preço do concorrente  $-j$ . O mapa de escolha para a firma 1 é

$$p_1^*(p_2) = \begin{cases} p_M & \text{se } p_2 > p_M, \\ p_2 - \epsilon & \text{se } c \leq p_2 \leq p_M, \\ c & \text{se } p_2 < c. \end{cases} \quad (5.16)$$

para  $\epsilon$  pequeno e  $p_M$  representando o preço de monopólio.

Como comentando anteriormente, o equilíbrio Bertrand-Nash corresponde a  $p_1 = p_2 = c$ , que é o mesmo resultado da concorrência perfeita. Se relaxarmos a hipótese de custos idênticos apenas uma firma produz e ela cobra preço igual ao custo marginal do segundo menor preço – no caso de dois produtores com custos marginais diferentes,  $c_1 < c_2$ , temos  $p_1 = c_2 - \epsilon$ . Neste caso a produção é eficiente, mas a produção é inferior à quantidade maximizadora do excedente.

i.e. o preço e a quantidade de equilíbrio no modelo de Bertrand (com empresas idênticas) são iguais ao resultado de concorrência perfeita.

Quando apresentamos o modelo de Cournot, vimos como o resultado de equilíbrio se aproxima do concorrencial, relativamente rápido, quando o número de firmas aumenta. No caso da escolha simultânea de preços entre firmas, encontramos este resultado de forma mais rápida. Bastam apenas duas firmas para se atingido o resultado de equilíbrio de concorrência perfeita.

O resultado do modelo de Bertrand é suspeito por algumas razões. Primeiro, e principalmente, se houver algum custo afundado ou fixo nunca ocorrerá entrada nesta indústria. Portanto, existirá uma firma para sempre. Sendo monopolista, a firma não fará preço igual ao custo marginal e sua habilidade para deter entrada de competidores depende apenas dos seus custos. Como não estamos assumindo informação imperfeita, a estrutura de custos é de conhecimento público. Por sua vez, sabendo da estrutura de custos, nunca ocorrerá entrada neste mercado. Em segundo lugar, não é comum observar a relação entre pequena variação nos preços e não produção de uma firma. Isto significa que a curva de demanda descontínua não parecer ser empiricamente plausível.

**Generalização.** Suponha que as firmas podem escolher preços entre 0 e  $+\infty$ , com o preço variando continuamente, e vamos derivar um mapeamento de estratégias que podem ser adotadas em um oligopólio, permitindo encontrar equilíbrio de Nash em preços.

A função melhor-reposta da firma  $j$  é um mapeamento que fornece, para cada preço de uma firma rival,  $-j$ , o preço ótimo  $p_j^*(p_{-j})$ . Suponha novamente duas firmas e que a

empresa 1 espera que a 2 escolherá o preço acima do preço de monopólio. Neste caso, a estratégia ótima para a firma 1 é fazer o preço igual ao nível de monopólio, pois assim ela obtém toda a demanda com lucro de monopólio. Se a firma 1 espera que a firma 2 cobre abaixo do preço de monopólio, então a estratégia ótima para a firma 1 é cobrar um preço um pouco abaixo ao praticado pela firma 2 (com um pequeno desconto de  $\epsilon$ ). Se a firma 2 planeja determinar o seu preço abaixo do custo marginal, então a firma 1 faz preço igual ao custo marginal.

No caso das duas firmas serem simétricas em termos de custo marginal, elas cobram preços iguais, não obtém lucro e dividem a demanda. Isto é:  $p_1^*(p_2) = p_2^*(p_1) = c$ . O resultado do modelo diz que se passamos da estrutura de mercado com uma empresa para duas, o lucro cai do valor de monopólio para a concorrência perfeita. Como esta situação não é considerada realista, isto ficou conhecido na literatura como o *Paradoxo de Bertrand*.

### 5.3.2 Paradoxo de Bertrand

Os problemas e questões que emergem do modelo de Bertrand foram analisados pela literatura econômica. Grande parte do problema pode ser resumido assim: como em uma configuração de duopólio o resultado é de concorrência perfeita? Algumas soluções foram sugeridas por diversos autores.

Considere o caso em que os custos marginais não são simétricos. Isto significa que uma firma é mais eficiente do que outra. Neste caso se a firma 2, que tem custo mais alto, faz  $p_2 = c_2$ , a firma 1 faz  $p_1 = c_2 - \epsilon$  se  $c_1 < c_2 - \epsilon$ . Nesta configuração a firma 1 leva o mercado inteiro, expulsando a rival. Neste caso específico o paradoxo de Bertrand não se sustenta pois apenas uma empresa continuaria ativa no mercado. Se a lógica for aplicada para a saída de algum número de empresas, mas se permanecem ao menos duas firmas com custos marginais iguais, continuamos a ter equilíbrio de Bertrand com preços iguais e lucro econômico igual a zero.<sup>7</sup>

A seguir apresentamos algumas configurações em que o resultado do paradoxo de Bertrand não é mantido.

#### Restrição de Capacidade

No modelo com custos simétricos, a firma que cobra menos leva todo o mercado, se tornando monopolista. Mas o que ocorre se a firma não tem capacidade instalada para suprir toda a demanda que ela dispõe na situação de monopólio?

Adicione capacidade ao modelo, digamos  $k_1$  e  $k_2$ . Por simplicidade o custo de produção será zero se produzir abaixo da capacidade:  $C(q) = 0$ . Por outro lado, não se pode produzir acima da capacidade. Esta é a restrição de capacidade:  $q_1 \leq k_1$  e  $q_2 \leq k_2$ . A demanda de mercado é  $P = 1 - Q$ .

<sup>7</sup>Todavia, como destacado por Cabral (2017), p. 192, liderança de custo dura pouco. O que significa que a configuração de exclusão de um rival pode ser temporária.

Assuma que a restrição de capacidade do mercado seja  $k_1 + k_2 \leq 1$ . Isto implica que as capacidades das firmas não são suficientes para tomar o mercado. Significando que não é possível expulsar o rival do mercado.

Considere a seguinte situação.<sup>8</sup> Se a firma 2 produz a capacidade  $q_2 = k_2$ , então a função melhor resposta da firma 1 é  $R_1(k_2) = (1 - k_1)/2$ . Vamos assumir que as capacidades são pequenas, limitando os ganhos. Neste caso,  $k_1 \leq R_1(k_2)$  e, similarmente,  $k_2 \leq R_2(k_1)$ . Neste exemplo as firmas nunca irão construir mais capacidade do que elas precisam. Como o jogo é com capacidade pequena, elas devem usar toda a sua capacidade, implicando em  $q_1 = k_1$  e  $q_2 = k_2$  – elas produzem no limite da capacidade.

Como a demanda de mercado é  $P = 1 - Q$ , neste caso o preço será  $P^* = 1 - k_1 - k_2$ . Este preço será acima do custo marginal, uma vez que as firmas tem capacidade pequena. Considere a seguinte racionalidade para uma firma. Se a firma 1 faz  $p_1 = p_2 < P^*$  ela não irá produzir mais. Se  $p_1 = p_2 > P^*$ , pelo menos uma firma não pode vender capacidade. Se a firma cortar o preço um pouco, ela pode vender a sua capacidade. A solução  $p_1 < p_2$ , ou vice-versa, não é factível pois a firma 1 possui restrição de capacidade e será bom aumentar os preços. Além disso, a outra situação seria a firma 1 como monopolista, mas isto somente seria possível se a firma 2 faça lucro-zero, e isto não ocorreria.

Considere agora um jogo de dois estágios: no primeiro estágio as firmas escolhem a sua capacidade e no segundo elas determinam preços condicional à escolha do primeiro estágio. No segundo estágio as firmas determinam os preços suficiente para produzir utilizando toda a capacidade instalada. Neste caso,  $q_1 = k_1$ ,  $q_2 = k_2$  e o preço ótimo será  $P^* = 1 - k_1 - k_2$ . No primeiro estágio a firma 1 escolhe a capacidade instalada:

$$\max_{k_1} k_1(1 - k_1 - k_2). \quad (5.17)$$

O problema para a firma 2 é simétrico. O resultado deste jogo é o mesmo do modelo de Cournot. Neste caso  $k_1 = k_2 = 1/3$  e  $P = 1/3$ .

### Produto Diferenciado

Neste caso os bens comercializados não são substitutos perfeitos. Neste caso, quando um produtor corta o preço do seu bem, ele não se torna monopolista, pois outros consumidores ainda preferem pagar mais para consumir o bem que ficou relativamente mais caro. Vamos ver esta solução no capítulo 7.

### Concorrência Dinâmica

O modelo de Bertrand assume concorrência estática – em apenas um período. A consequência da redução do preço em um percentual pequeno pode ser o início de guerra de preços entre os competidores. Mas por ser um modelo estático, retaliação não é previsto no modelo. Quando existem configurações dinâmicas permitindo comportamento estratégico entre os agentes é possível o preço ser maior do que o custo marginal. Outra

<sup>8</sup>Seguindo aqui exemplo proposto por Matthew Schum.

possibilidade é a existência de rigidez de preço, i.e. quando os preços não podem mudar livremente. Neste caso, a escolha do meu preço hoje determina a escolha do preço do meu rival. Neste setup pode existir mais de um equilíbrio. Alguns tipos de setup são guerras de atrito e ciclos de Edgeworth.

Exemplo: ciclo de Edgeworth. No exemplo de equilíbrio descrito na Tabela 5.1, cada firma reduz o preço em função da mudança do rival até o preço atingir o nível concorrencial (equilíbrio competitivo),  $p(0)$ , até o ponto em que uma firma eventualmente reverte para o preço elevado  $p(5)$ . A dinâmica de mercado consiste em guerra de preços seguido por período cedente. Este último período é uma “guerra de atrito” em torno de  $p(0)$  em que cada firma espera pela outra para aumentar os preços. Poderia ser perguntado por que uma empresa poderia manter o preço baixo fazendo lucro zero. A explicação é que o período cedente é um bom público a partir do ponto de vista da firma. Ambas as firmas querem que o rival aumente o preço para ela ter a liderança na redução dos preços e ficar em posição vantajosa.<sup>9</sup>

Tabela 5.1: Exemplo de Ciclo de Edgeworth

p	R(p)
$p(6)$	$p(5)$
$p(5)$	$p(4)$
$p(4)$	$p(3)$
$p(3)$	$p(2)$
$p(2)$	$p(1)$
$p(1)$	$p(0)$
$p(0)$	$p(5)$ com probabilidade positiva

Observe que durante a fase da guerra de preços, uma firma reduz o preço em relação ao rival não apenas para aumentar o seu market share, mas porque ela não confia no rival. Isto é, a firma antecipa que mantendo o preço o seu preço ela não irá prevenir que as outras firmas deixem de ser agressivas.<sup>10</sup>

**Exemplo: Precificação com Restrição de Capacidade.** Vamos considerar o exemplo de restrição de capacidade com duas firmas, mas agora com grandes capacidades. Por simplicidade assuma que  $k_1 = k_2 = k$ . A função de reação é relacionada com a capacidade como  $k > R(k)$ . Neste caso  $p = 1 - 2k$  é um equilíbrio? Ele não será um equilíbrio porque a firma 1, por exemplo, pode desviar, reduzindo a produção para  $q_1 = R(k) < k$ , resultando em preço mais elevado.

Vamos seguir diversos eventos para exemplificar o comportamento estratégico. Por conveniência vamos começar com o preço como  $p_1 = p_2 = 1 - 2k$  e com ambas as firmas

<sup>9</sup>Veja Maskin e Tirole, 1988, p. 577.

<sup>10</sup>Para uma aplicação veja Noel (2007).



produzindo na capacidade  $q_1 = q_2 = k$ . Assuma também que as firmas se revezam mudando os preços e o novo valor dura por dois períodos.

A firma 1 desvia e produz  $R(k) < k$  unidades, aumentando o preço para  $p^h = (1 - k) - R(k)$ . Como assumimos que o preço das firmas não pode mudar em dois dias, a firma 2 não responde imediatamente.

Após a decisão da firma 1, o preço  $1 - 2k$  não continua é ser decisão ótima. A firma pode tomar duas decisões. Ela pode reduzir um pouco o preço em  $\epsilon$  para fazer  $p_2 = p^h - \epsilon$  e ter volume de vendas próximo de  $k$ . A outra possibilidade para a firma 2 é ser “monopolista” para a demanda residual com preço  $p_2 = (1 - R(k)) - q_2$ . Vamos assumir que a primeira decisão seja mais lucrativa, portanto a firma faz  $p_2 = p^h - \epsilon$  e  $q_2 \approx k$ .

A firma 1 tem as mesmas escolhas que a firma 2 teve anteriormente. Vamos assumir que para a firma 1 a primeira alternativa é a melhor. Então ela faz  $p_1 = p_2 - \epsilon$  e  $q_1 \approx k$ . Quando a firma 2 toma a sua decisão, novamente ela reduz o preço em  $\epsilon$ . Este processo de concorrência continua até certo ponto quando a firma 2 domina o mercado. Isto ocorre quando o preço  $p_2^{**}$  satisfaz a condição de indiferença  $p^{**}k = p^h R(k)$ . Esta condição é atingida quando  $p_j = p^h$  e  $q_j = R(k)$  para  $j = \{1, 2\}$  e o ciclo de Edgeworth recomeça.

Este comportamento das firmas não é um fenômeno de equilíbrio. Na verdade, este modelo não tem equilíbrio de estratégia-pura, mas tem equilíbrio de estratégia-mista. Pode ser mostrado que ele é uma trajetória de equilíbrio de um jogo dinâmico. Observe que a restrição de capacidade é fundamental, caso contrário a firma 1 nunca aumentaria o preço pois a firma 2 capturaria todos os consumidores. Neste modelo, a firma 1 “paga o preço” de ser a líder de mercado ao mudar os preços e criar o benefício para todas as firmas.

## 5.4 Modelo de Stackelberg

O modelo de Stackelberg, também é conhecido como modelo firma líder-seguidora, é um modelo de equilíbrio sequencial. Este modelo é uma modificação do modelo de equilíbrio simultâneo de Cournot para permitir escolha sequencial de produção entre as firmas.

A versão mais simples do modelo assume duas empresas com custo e demanda linear.<sup>11</sup> Neste caso, suponha que a empresa 1 seja a líder, portanto ela escolhe primeiro. A firma 2 segue a empresa líder e escolhe em seguida a sua produção. Este problema pode ser escrito como um jogo de dois estágios: a firma líder escolhe no primeiro estágio e a seguidora no segundo.

A solução do jogo simultâneo é por *indução retroativa*: a firma líder escolhe a quantidade ótima sabendo que a escolha da seguidora é residual:  $q_2^*(q_1)$ . Neste caso a escolha ótima da firma 1 consiste na escolha da firma 2 que maximiza o lucro (da firma 1). O problema

<sup>11</sup>Assumindo credibilidade das ações das firmas.

da firma 1 usando indução retroativa é:

$$\pi_1(q_1, q_2^*(q_1)) = D(q_1, q_2^*(q_1))(p_1 - c_1) \quad (5.18)$$

Portanto, a firma 1 escolhe a quantidade ótima,  $q_1^*$ , utilizando a função de reação da firma 2. A firma 1, como é líder, escolhe a quantidade elevada para que a firma 2 opere com lucro mais baixo.

Em comparação ao resultado de duopólio de Cournot, a firma 1 produz mais e a firma 2 menos. Na oferta total de mercado o modelo de Stackelberg entrega maior produção do que o modelo de Cournot (com custo e demanda linear). Neste caso, o aumento na produção da firma 1 mais do que compensa a queda na produção da firma 2.

## Capítulo 6

# Estimação da Demanda: Produto Homogêneo

A aplicação de análise econômica dos mercados está intimamente ligada ao uso de dados. Uma introdução a métodos econométricos torna a análise empírica dos problemas completa e por vezes mais simples. De forma geral, a principal ferramenta empírica para analisar mercados é a aplicação de classes de modelos com variáveis instrumentais. A utilização deste método é fundamental, pois sem o seu uso o economista invariavelmente precisará evitar a análise completa de mercado e da dinâmica de firma. Por ser central na análise empírica de economia industrial, entendemos que este método é fundamental para o economista aplicado.

Os problemas econômicos relacionados a identificação da função demanda são clássicos em economia aplicada e remontam os trabalhos de Wright (1915), Working (1927) e Marschak e Andrews (1944). Para estimar uma especificação de função demanda é fundamental a utilização de variáveis instrumentais. Neste capítulo vamos analisar casos de demanda em mercados com produtos homogêneos. Vale ressaltar que o estudo empírico da demanda é a maior ferramenta para a análise de estática comparativa de qualquer mudança no ambiente econômico que não tem impacto imediato nos custos.

Métodos de variáveis instrumentais fazem parte de uma definição ampla conhecido como *Método Generalizado dos Momentos* ou pela sigla GMM (*Generalized Method of Moments*). Esta classificação é considerado geral pois o estimador de mínimos quadrados ordinários (MQO ou OLS) pode ser apresentado como um caso particular do GMM.

Resumidamente, a necessidade do uso de variáveis instrumentais decorre da existência de alguns tipos de viés no uso do estimador de MQO. As classes de viés de estimação por MQO que levam a necessidade de aplicar um estimador GMM são:

- viés de variável omitida;
- viés de erros nas variáveis;

- viés de equação simultânea.

A existência de um destes viés não permite a sustentabilidade das hipóteses de MQO e sua subsequente aplicação. A existência destes problemas e como solucionar faz parte do problema de identificação econométrica. Outra forma de pensar o problema de identificação econométrica é a seguinte. Podemos logicamente deduzir o valor desconhecido dos parâmetros da distribuição dos dados que observamos? Se a resposta para questão é **não**, sob certo conjunto de hipóteses, então os estimadores consistentes não existem sob este conjunto de hipóteses. Por outro lado, se a resposta para esta questão for **sim** sob um certo conjunto de hipóteses então o estimador consistente pode existir. Portanto, quando violamos as hipóteses de MQO frente a existência de um dos três viés, precisamos utilizar outro modelo para estimar os parâmetros necessários para analisar mercados.

## 6.1 Exemplos

Condição para este capítulo é o conhecimento do modelo de mínimos quadrados ordinários. O modelo MQO pode ser utilizado a partir do método dos momentos, que caracterizam a estimação. Suponha o seguinte modelo de regressão:

$$Y_i = \alpha + X_i\beta + \varepsilon_i, \quad (6.1)$$

Os momentos do MQO são:

$$E(\varepsilon_i | X_i) = 0 \quad [M.1]$$

e

$$E(\varepsilon_i X_i) = 0 \quad [M.2]$$

O primeiro momento [M.1] diz que a média condicional do termo de erro é zero. O segundo momento [M.2] diz que a covariância entre os regressores e o termo de erro é zero. Isto significa que o regressor  $X$  é exógeno. Se estes momentos não são respeitados então o estimador MQO não deve ser aplicado.

Apresentamos abaixo exemplos de viés que impedem a aplicação do estimador MQO.

**Viés de variável omitida.** Um dos momentos amostrais do estimador MQO diz que

$$E(X, U) = 0, \quad (6.2)$$

tal que  $X$  é um regressor e  $U$  o termo de erro. Quando este momento<sup>1</sup> é satisfeito dizemos que  $X$  é exógeno.

Agora considere a seguinte situação: o modelo é corretamente identificado se o modelo é estimado com os regressores  $X$  e  $W$ . Neste caso,  $W$  é um determinante importante da variável dependente e também é correlacionado com  $X$ . Quando se inclui  $W$  temos

<sup>1</sup>Este momento amostral é a covariância. Outro momento do estimador MQO é  $E(U) = 0$ .

que  $E(X, U) = 0$ . Por outro lado, se excluirmos  $w$ , então  $E(X, U) \neq 0$ . Dizemos que a omissão de  $W$  no modelo de regressão gera um viés, pois  $X$  deixa de ser exógena – e esta é uma condição para a aplicação do modelo de mínimos quadrados ordinários. Este viés causado pela ausência de  $W$  é chamado de *viés de variável omitida*.

**Viés de simultaneidade.** Considere o exemplo clássico de Working (1927) que analisa um modelo simples de oferta e demanda. Considere por exemplo uma commodity, o mercado internacional de café em grãos. Modelo de oferta e demanda é descrito por três equações:

$$q_i^d = \alpha_0 + \alpha_1 p_i + u_i \quad (6.3)$$

$$q_i^s = \beta_0 + \beta_1 p_i + v_i \quad (6.4)$$

$$q_i^d = q_i^s \quad (6.5)$$

tal que  $q$  = quantidade,  $p$  = preço,  $u$  é o termo de erro da demanda e  $v$  o termo de erro da oferta. A notação indicadora é: tempo  $i$ , demanda  $d$ , e oferta  $s$ . A terceira equação descreve o equilíbrio de mercado; demanda igual a oferta.

Vamos assumir que as médias dos erros sejam  $E(u_i) = 0$  e  $E(v_i) = 0$ , e por simplicidade não existe covariância entre os dois termos de erro:  $Cov(u_i, v_i) = 0$ . Assumindo equilíbrio de mercado, isto é, que  $q_i^d = q_i^s = q_i$ , temos:

$$q_i = \alpha_0 + \alpha_1 p_i + u_i \quad (6.6)$$

$$q_i = \beta_0 + \beta_1 p_i + v_i. \quad (6.7)$$

Podemos dizer que um *regressor é endógeno* se ele não é predeterminado, i.e. não é ortogonal ao termo de erro. Neste sentido, ele viola a hipótese básica de MQO de não-existência de relação entre regressor e erro.

Quando rodados uma regressão de quantidade contra preço (e uma constante), estamos estimando a oferta ou demanda? Como preço é endógeno em ambas equações a resposta não é clara. Do estimador de MQO sabemos que:

$$\alpha_1^{MQO} = \frac{Cov(p_i, q_i)}{Var(p_i)}, \quad (6.8)$$

com  $\alpha_1^{MQO}$  significando a estimativa de  $\alpha_1$  por MQO.

Para explicitar o efeito-preço na curva de demanda use  $q_i = \alpha_0 + \alpha_1 p_i + u_i$  e faça:

$$Cov(p_i, q_i) = \alpha_1 Var(p_i) + Cov(p_i, u_i) \quad (6.9)$$

Substituindo (6.9) em (6.8), temos o **viés assintótico** para  $\alpha_1$ :

$$\text{viés}(\alpha_1^{MQO} - \alpha_1) = \frac{Cov(p_i, u_i)}{Var(p_i)} \quad (6.10)$$

Isto é conhecido como **viés de endogeneidade**. Porque o erro e o regressor são relacionados por meio de um sistema de equações simultâneas, esse viés também é conhecido como **viés de equações simultâneas** ou **viés de simultaneidade**. O viés existe porque a covariância entre o preço e o termo de erro de outra equação não é zero. O ponto fundamental da análise empírica é que estes tipos de problemas surgem frequentemente na análise econômica.

## 6.2 Variáveis Instrumentais

Nesta seção apresentamos o estimador de variáveis instrumentais para casos mais simples de produtos homogêneos. A análise apresentada a seguir não é uma abordagem geral mesmo para o caso de produtos homogêneos, mas é aplicável para muitas situações neste tipo de mercado.

### 6.2.1 O Problema

Considere o seguinte modelo de regressão:

$$Y_i = \alpha + Z_i\beta + \varepsilon_i, \quad (6.11)$$

$$(i = 1, 2, \dots, n).$$

$Y_i$  é a variável objeto de estudo do modelo para a observação  $i$ ,  $Z$  um regressor de interesse e  $\varepsilon$  o termo de erro desta equação. O parâmetro de interesse aqui é  $\beta$ , mas no caso  $Z$  possui algum viés (mencionado anteriormente). A existência do viés viola a exogeneidade de  $Z$ ,<sup>2</sup> fazendo com que o momento cruzado  $Z$  com o termo de erro  $\varepsilon$  não seja zero, i.e.

$$E(Z_i\varepsilon_i) \neq 0. \quad (6.12)$$

Quando vale esta condição, dizemos que  $Z$  é endógeno. Como discutido anteriormente, se este modelo for estimado por OLS ele apresentará viés.

Uma alternativa para resolver este problema é a estimação por variáveis instrumentais.

Exemplo mercado internacional de café em grãos. Este produto é uma commodity é exportado da região produtora, tipicamente tropical, elevada e com muito sol, para a região compradora. Os grandes compradores de café em grão são empresas que processam café ou traders, que geralmente se localizam em regiões que não são as produtoras (e.g. Europa).

Portanto, um instrumento típico para identificar demanda é utilizar um choque de oferta. No exemplo do mercado de café, um instrumento pode ser a variação de temperatura em localidades produtoras. Neste caso vamos ter uma equação estimável que relaciona a variável endógena a variável instrumental.

$$Z_i = \mu + X_i\gamma + V_i, \quad (6.13)$$

<sup>2</sup>Se  $Z$  fosse exógeno o momento cruzado seria  $E(Z_i\varepsilon_i) = 0$  e não seria preciso a solução de variáveis instrumentais.

tal que  $X_i$  é o instrumento, e deve ser exógeno tanto em (6.13) quanto em (6.11). Quando se tem um estimador de variável instrumental válido o momento instrumento e termo de erro estrutural é válido.

$$E(X_i \varepsilon_i) = 0 \quad (6.14)$$

A seguir vamos ver as condições para que a variável instrumental possa ser utilizada em algum método.

### 6.2.2 Validade de Instrumentos e Identificação

Para aplicar um método de variáveis instrumentais é preciso ter instrumentos válidos, ou seja, que eles satisfaçam algumas condições que permitem o seu uso. Estas condições são a de relevância e de exogeneidade. Questões sobre relevância e validade dos instrumentos são relevantes e fazem parte da rotina de se estimar um modelo de variáveis instrumentais.

**Condição de relevância.** O instrumento deve ser correlacionado com a variável endógena. Isto é dada pela correlação entre as variável  $Z_i$  e o instrumento  $X_i$ :  $\text{Corr}(Z_i, X_i) \neq 0$ .

**Condição de exogeneidade.** A condição é  $\text{Corr}(X_i, \varepsilon_i) = 0$ . Isto é, a parte da variância capturada pelo instrumento é exógena no modelo (6.11).

Podemos apresentar as condições de validade dos instrumentos de forma geral. Como se pode ter mais de um instrumento para um regressor endógeno, e também se pode ter mais de um regressor endógeno. Denominamos  $k$  como o número de instrumentos,  $l$  como o número de regressores endógenos. Portanto uma formulação geral de validade dos instrumentos é a seguinte:

- $E[Xe] = 0$
- $E[XX'] > 0$
- $\text{posto}(E[XZ']) \geq k$

A primeira condição diz que todos os instrumentos não são correlacionados com os regressores. Que é mesma da condição de exogeneidade mas pode ser representada de forma geral. A segunda condição diz que não pode existir colinearidade entre os instrumentos – eventualmente é denominada de condição de *posto completo*.

A terceira condição é mais complicada, pois é uma condição de ordem de *identificação* do modelo de variáveis instrumentais.

**Condição de Ordem de Identificação:**<sup>3</sup> A condição para identificação é que

$$k \geq l \quad (6.15)$$

<sup>3</sup>Hayashi (2000) p. 202.

em palavras:

$$\text{número de instrumentos} \geq \text{número de regressores}$$

Portanto um modelo de variáveis instrumentais pode ser classificado como:

**Sobreidentificado:** a condição de ordem é satisfeita com  $k > l$ ;

**Exatamente identificado:** quando a condição de ordem é satisfeita com  $k = l$ ;

**Subidentificado ou não-identificado:** a condição de ordem não é satisfeita,  $k < l$ .

Quando se estima um modelo de variáveis instrumentais estas condições devem ser objeto da análise. Não são apenas condições que se impõem, mas estas servem como guia para entender a dinâmica do modelo econométrico.

Agora vamos estudar um estimador que utiliza variáveis instrumentais. Este estimador é o mínimo quadrado de dois estágios.

### 6.2.3 Estimador 2SLS

O modelo formado pelas equações (6.31) e (6.32) pode ser estimado pelo método de mínimos quadrados de dois estágios (2SLS de Two Stage Least Squares).<sup>4</sup> O modelo composto pelas equações (6.31) e (6.32) tem função de ser didático e servir como um *workhorse*. Em termos práticos os economistas estimam este modelo sem incluir a variável instrumental,  $X$ , na equação estrutural.

O modelo 2SLS é composto por uma equação de primeiro estágio (as vezes chamada de equação forma-reduzida), que estima a relação entre a variável endógena e o instrumento:

$$Z_i = X_i\gamma + V_i \quad (6.16)$$

e pelo segundo estágio (equação estrutural), utilizando o valor previsto da variável endógena,  $\hat{Z}_i$ , em (6.35) como regressor:

$$Y_i = \hat{Z}_i\beta + \varepsilon_i. \quad (6.17)$$

No primeiro estágio  $Z$  é decomposto em dois componentes: um problemático que pode ser correlacionado com o termo de erro e outro livre de erro,  $X_i\gamma$ , que não é correlacionado com o termo de erro. O segundo estágio usa o componente livre de “problema” para estimar o parâmetro de interesse,  $\beta$ .

O uso de ferramentas econométricas requer que a aplicação não viole as hipóteses do modelo escolhido. Portanto, observar as hipóteses do modelo 2SLS é fundamental para a utilização adequada, algumas já vimos anteriormente, mas vamos resumir elas abaixo.

Defina *consistência* de um estimador como a capacidade de convergência do parâmetro estimado para o parâmetro verdadeiro (do modelo). Represente  $\beta_0$  como o parâmetro

<sup>4</sup>Proposto originalmente por Theil (1953).



verdadeiro para a escolha de dados e modelo (aqui sendo o 2SLS). Portanto, a consistência implica que  $\hat{\beta} \rightarrow \beta_0$ . Ou seja, que a estimativa de  $\beta$  converge para o parâmetro verdadeiro  $\beta_0$ .

As hipóteses necessárias para garantir consistência do estimador 2SLS, tais quais apresentadas por Hansen (2022, cap. 12), são resumidas abaixo:

1. As variáveis  $(Y_i, \mathbf{Z}_i, \mathbf{X}_i)$ ,  $i = 1, \dots, n$  são independentes e identicamente distribuídas.
2.  $E[Y^2] < \infty$ .
3.  $E\|\mathbf{X}\|^2 < \infty$ .
4.  $E\|\mathbf{Z}\|^2 < \infty$ .
5.  $E[\mathbf{X}\mathbf{X}']$  é positiva definida.
6.  $E[\mathbf{X}\mathbf{Z}']$  possui posto completo  $k$ .
7.  $E[\mathbf{X}\varepsilon_i] = 0$ .

As quatro primeiras hipóteses dizem que todas as variáveis possuem variância finita – elas não crescem sem limite. Hipótese 5 diz que o vetor/matriz de instrumentos tem uma matriz de desenho invertível, que é idêntica a hipótese central sobre os regressores no modelo MQO. Isto exclui instrumentos que possuem dependência linear – não pode existir dependência linear.

As hipóteses 6 e 7 são fundamentais para identificação das variáveis instrumentais (como visto anteriormente). A hipótese 6 diz que regressores e instrumentos formam uma matriz de momentos cruzados que é posto completo. Esta é a *condição de relevância* mencionada anteriormente. A hipótese 7 diz que a covariância entre instrumentos e termo de erro da equação estrutural é zero.

**Teorema 1** *Sob as hipóteses 1 a 7,  $\hat{\beta}_{2SLS} \rightarrow_p \beta_0$  a medida que  $n \rightarrow \infty$ .*

A prova é a seguinte e é importante pois possuem elementos recorrentes na análise econométrica. O resultado inclui o estimador de variáveis instrumentais  $k = l$  como caso especial.

Tome a equação estrutural (6.11) no formato matricial:  $Y = Z\beta + e$ .<sup>a</sup> A modelo de primeiro estágio é estimado como MQO, então a relação entre variável endógena,  $Z$  e instrumento  $X$  é  $\Gamma = (X'X)^{-1}X'Z$ . Então a equação estrutural será:

$$Y = X'\hat{\Gamma}\beta + e,$$

para simplificar vamos denominar  $X'\hat{\Gamma} \equiv R$ , então a equação estrutural passa a ser  $Y = R\beta + e$ . Assim como o estimador de MQO, o estimativa de  $\beta$  será:

$$\hat{\beta} = (R'R)^{-1}(R'Y) = (\hat{\Gamma}'X'X\hat{\Gamma})^{-1}(\hat{\Gamma}'X'Y).$$

Por sua vez sabemos que  $\Gamma = (X'X)^{-1}X'Z$ , então substituímos na equação acima e temos:

$$\begin{aligned}\hat{\beta} &= (Z'X(X'X)^{-1}X'X(X'X)^{-1}X'Z)^{-1}(Z'X(X'X)^{-1}X'Y). & (6.18) \\ &= (Z'X(X'X)^{-1}X'Z)^{-1}(Z'X(X'X)^{-1}X'Y).\end{aligned}$$

Esta equação é o estimador 2SLS de  $\beta$ .

Agora substitua por  $Y = Z\beta_0 + e$  na (6.18), logo se tem

$$\hat{\beta} = (Z'X(X'X)^{-1}X'Z)^{-1}(Z'X(X'X)^{-1}X'(Z\beta_0 + e)).$$

Reescrevendo se tem:

$$\hat{\beta} - \beta_0 = (Z'X(X'X)^{-1}X'Z)^{-1}(Z'X(X'X)^{-1}X'e). \quad (6.19)$$

Com isto separamos o componente estocástico  $e$ . Esta é a origem ao que chamamos de erro amostral (sampling error). Quando o modelo é bem especificado, então o erro amostral será zero. Para tanto reescrevemos a equação utilizando o teorema do limite central e a lei dos grandes números:

$$\begin{aligned}\hat{\beta} - \beta_0 &= \\ &= \left( \left( \frac{1}{n}Z'X \right) \left( \frac{1}{n}X'X \right)^{-1} \left( \frac{1}{n}X'Z \right) \right)^{-1} \left( \left( \frac{1}{n}Z'X \right) \left( \frac{1}{n}X'X \right)^{-1} \left( \frac{1}{n}X'e \right) \right) \\ &\rightarrow_p (\Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ})^{-1}\Sigma_{ZX}\Sigma_{XX}^{-1}E[Xe] = 0\end{aligned} \quad (6.20)$$

tal que

$$\Sigma_{ZX} = E[ZX'],$$

$$\Sigma_{XZ} = E[XZ'],$$

e

$$\Sigma_{XX} = E[XX'].$$

<sup>a</sup>Aqui a constante passa a fazer parte da matriz de dados  $Z$ .

Algumas explicações sobre a prova. A lei fraca dos grandes números se aplica sob as hipóteses 1 a 4. O teorema do mapeamento contínuo se aplica se as matrizes  $\Sigma_{XX}$  e  $(\Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ})$  são invertíveis, que valem sob as hipóteses 5 e 6 acima. A última igualdade usa a hipótese 7.

**Distribuição assintótica do estimador 2SLS.** Agora vamos olhar para os requisitos e propriedades da estimação da variância do 2SLS. A variância é necessária para inferência estatística utilizando testes t e F.

Vamos mostrar que o estimador 2SLS satisfaz um teorema do limite central. Novamente seguindo Hansen (2022) as hipóteses necessárias são:

1.  $E[Y^4] < \infty$ .
2.  $E\|X\|^4 < \infty$ .
3.  $E\|Y\|^4 < \infty$ .
4.  $S = E[XX'\varepsilon^2]$  é positiva definida.

O conjunto de hipóteses acima, também tratadas anteriormente como *normalidade assintótica* e *homocedasticidade condicional*, fortalecem as hipóteses para consistência necessitando que todas as variáveis tenham o quarto momento finito. Isto é usado para estabelecer o teorema do limite central.

**Teorema 2** Sob as hipóteses 1 a 4 acima, a medida que  $n \rightarrow \infty$ .

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) \rightarrow_d N(0, Var)$$

tal que

$$Var = (\Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ})^{-1} (\Sigma_{ZX}\Sigma_{XX}^{-1}S\Sigma_{XX}^{-1}\Sigma_{XZ}) (\Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ})^{-1}.$$

Esta matriz de variância é robusta à heterocedasticidade.

Isto mostra que o estimador 2SLS converge a taxa  $\sqrt{n}$  para um vetor aleatório normal. O Teorema 2 também mostra como formar a matriz de covariância.

Sob a hipótese de homocedasticidade condicional temos as seguintes simplificações:

$$S = \Sigma_{XX}\sigma^2, \tag{6.21}$$

e

$$Var^{\text{homocedástica}} = (\Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ})^{-1} \sigma^2. \tag{6.22}$$

Vamos analisar os fatores que determinam a variância para entender a precisão, ou a falta, do estimador. Tome a fórmula da variância (6.22) em sua forma teórica:

$$Var = (E[ZX'](E[XX])^{-1}E[XZ'])^{-1} E[\sigma^2]. \tag{6.23}$$

Podemos observar que a variância é crescente na variância do termo de erro,  $E[\sigma^2]$ , e decrescente na variância dos regressores,  $Z$ , e na correlação entre regressores e instrumentos.

Também é útil observar que a expressão da variância não é afetada pela estrutura do instrumento  $X$ . A variância também não é afetada por rotações imposta a  $X$  (se você trocar  $X$  por  $CX$  para a matrix inversível  $C$  a expressão não muda). Isto significa que a expressão da variância não é afetada pela mudança em  $X$  e também não diretamente afetada pela correlação entre os elementos de  $X$ .

Também podemos usar a expressão (6.23) para examinar o impacto de aumentar o conjunto de instrumentos. Suponha que separamos os instrumentos em dois grupos:  $X = (X_a, X_b)$ . Aqui a dimensão é  $\dim(X_a) \geq l$ , tal que podemos construir um estimador 2SLS utilizando apenas  $X_a$ . Faça agora  $\hat{\beta}_a$  ser a estimativa 2SLS utilizando  $X_a$  e  $\hat{\beta}$  ser a estimativa a partir de  $X = (X_a, X_b)$ . Sem perda de generalidade, assumimos que  $X_a$  e  $X_b$  são não-correlacionados. Neste caso, ambos os momentos  $E[XX']$  e  $(E[XX'])^{-1}$  são bloco-diagonal tal que<sup>5</sup>

$$\begin{aligned} Avar(\hat{\beta}) &= \left( E[ZX'](E[XX])^{-1}E[XZ'] \right)^{-1} \sigma^2 \\ &= \left( E[ZX'_a](E[X_aX_a])^{-1}E[X_aZ'] + E[ZX'_b](E[X_bX_b])^{-1}E[X_bZ'] \right)^{-1} \sigma^2 \\ &\leq \left( E[ZX'_a](E[X_aX_a])^{-1}E[X_aZ'] \right)^{-1} \sigma^2 \\ &= Avar(\hat{\beta}_a), \text{ com desigualdade estrita se } E[ZX'_b] \neq 0. \end{aligned}$$

Portanto, se o estimador 2SLS com o conjunto completo de instrumentos possui menor variância assintótica do que o estimador com menor grupo de instrumentos.

O que é mostrado aqui é que a variância do estimador 2SLS é decrescente a medida que o número de instrumentos aumenta. Do ponto de vista da *eficiência assintótica*, isto significa que é melhor utilizar mais instrumentos. Infelizmente existe um custo em se utilizar mais instrumentos. O viés de amostra finita do estimador 2SLS é genericamente crescente linearmente a medida que o número de instrumentos aumenta. Na prática a escolha de instrumentos possui um *trade-off* entre viés e variância.<sup>6</sup>

**Estimação da Matriz de Covariância do 2SLS.** A estimação é realizada pela substituição das matrizes de momento populacional pela contrapartida amostral. Portanto, a estimação da variância  $Var$  será  $\widehat{Var}$ <sup>7</sup>

$$\widehat{Var} = (\hat{\Sigma}_{ZX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XZ})^{-1} \left( \hat{\Sigma}_{ZX} \hat{\Sigma}_{XX}^{-1} \hat{S} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XZ} \right) (\hat{\Sigma}_{ZX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XZ})^{-1}. \quad (6.24)$$

<sup>5</sup>  $Avar$  é a variância assintótica. Variância quando  $(n \rightarrow \infty)$ .

<sup>6</sup> Para mais detalhes veja Hansen (2022), seção 12.37.

<sup>7</sup> A matriz de variância também é chamada de matriz de covariância ou variância-covariância.

As matrizes<sup>8</sup> são definidas como:

$$\hat{\Sigma}_{XX} = \frac{1}{n} \sum_{i=1}^n X_i X_i' = \frac{1}{n} X' X,$$

$$\hat{\Sigma}_{ZX} = \frac{1}{n} \sum_{i=1}^n Z_i X_i' = \frac{1}{n} Z' X,$$

$$\hat{S} = \frac{1}{n} \sum_{i=1}^n X_i X_i' \hat{e}_i^2,$$

$$\hat{e}_i = Y_i - Z_i' \hat{\beta}_{2SLS}.$$

A matriz de covariância homocedástica é mais simples e pode ser estimada por

$$\widehat{Var}^{\text{homocedástica}} = \left( \hat{\Sigma}_{ZX} \hat{\Sigma}_{XX}^{-1} \hat{\Sigma}_{XZ} \right)^{-1} \hat{\sigma}^2 \quad (6.25)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2.$$

Os erro-padrão são obtidos como a raiz quadrada dos elementos da diagonal de  $(1/n)\hat{V}$ . Intervalos de confiança, teste t e F podem ser construídos a partir das estimativas dos coeficientes e da matriz de variância. Por exemplo, o intervalo de confiança a 95% para o regressor  $d$  (qualquer) é escrito como:

$$C.I. \text{ para } \beta_d = \hat{\beta}_d \pm 1.96 \times \sqrt{\frac{1}{n} \widehat{Var}_d}. \quad (6.26)$$

O intervalo de confiança para possui dois “lados”, um negativo e outro positivo, para  $\beta_d$  que contém a estimativa pontual  $\hat{\beta}_d$  com 95% de probabilidade. De forma equivalente este é conjunto de valores de  $\beta_d$  que não pode ser rejeitado por um teste de hipótese de 5% de dois lados.<sup>9</sup>

**Teorema 3** *Sob o conjunto de hipóteses da variância 1 a 4, a medida que  $n \rightarrow \infty$ ,  $\widehat{Var}^{\text{homocedástica}} \rightarrow_p Var^{\text{homocedástica}}$  e  $\widehat{Var} \rightarrow_p Var$ .<sup>10</sup>*

É importante que a matriz de variância seja construída usando a fórmula correta de resíduo  $\hat{e}_i = Y_i - Z_i' \hat{\beta}$ .

<sup>8</sup>X sem o indicador  $i$  representa uma matriz.

<sup>9</sup>Para mais detalhes veja Stock e Watson (2015), seção 5.2. Se for necessário computar o intervalo de confiança para 90%, substitua 1.96 por 1.64 na fórmula de intervalo de confiança.

<sup>10</sup>Para provar este teorema é crucial mostrar que  $\hat{S} \rightarrow_p S$  como os outros resultados de convergência são estabelecidos na prova de consistência.  $\rightarrow_p$  significa convergência em probabilidade.

### Teste de Hipótese

Teste de hipótese é construído diretamente a partir dos teoremas de consistência e da distribuição assintótica. Aqui estamos partindo da familiaridade do leitor com métodos de mínimos quadrados ordinários, portanto o que se segue é em grande parte uma revisão.<sup>11</sup>

O teste-t para testar a hipótese  $\beta_d = \beta_{d,0}$  contra a alternativa  $\beta_j \neq \beta_{d,0}$  para um regressor  $d$  qualquer segue o seguinte procedimento. Tipicamente testamos quando  $\beta_{d,0}$  é 0. É este o caso que apresentado após a estimação utilizando softwares estatísticos. O teste-t é apresentado abaixo:

(1) Calcule o erro-padrão de  $\hat{\beta}_d$ :  $SE(\hat{\beta}_d) = \sqrt{\frac{1}{n} \widehat{Var}_d}$ ;

(2) Calcule a estatística-t:

$$t = \frac{\hat{\beta}_d - \beta_{d,0}}{SE(\hat{\beta}_d)}; \quad (6.27)$$

(3) Calcule o p-valor:

$$\text{p-valor } \alpha = 2\Phi(-|t^{comp}|),$$

tal que  $t^{comp}$  é o valor da estatística-t calculada em (6.27). Rejeite a hipótese nula a 5% de significância se o p-valor é menos do que 0.05 ou, equivalentemente,  $|t^{comp}| > 1.96$ .

Para o teste de conjunto de hipóteses utilizamos a estatística-F. Considere uma função  $r(\beta)$ . Em notação matricial podemos escrever:

$$\mathbf{R}\beta = \mathbf{r} \quad (6.28)$$

sendo  $\mathbf{R}$  uma matriz não-aleatória  $q \times (k+1)$  com posto completo e  $\mathbf{r}$  um vetor não-aleatório  $q \times 1$ . O número de linhas de  $\mathbf{R}$  é  $q$ , que é o número de restrições impostas sob a hipótese nula.

A hipótese nula na equação (6.28) engloba todas as hipóteses nulas. Por exemplo, uma hipótese conjunta seria de que  $\beta_1 = 0, \beta_2 = 0, \dots, \beta_{q-1} = 0$ . Para escrever a hipótese conjunta na forma (6.28) faça  $\mathbf{R} = \begin{bmatrix} \mathbf{I}_q & \mathbf{0}_{q \times (k+1-q)} \end{bmatrix}$  e  $\mathbf{r} = \mathbf{0}_q$ . Por exemplo, para testar que a relação entre dois parâmetros  $\beta_2$  e  $\beta_3$  seja igual a  $\beta_1 + \beta_2 = 1$ , isto pode ser escrito na forma matricial como  $\mathbf{R} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$ ,  $\mathbf{r} = 1$ , e  $q = 1$ .

A estatística-F robusta a heterocedasticidade para testar a hipótese conjunta determinada por (6.28) é

$$F^R = (\mathbf{R}\hat{\beta} - \mathbf{r})' \left[ \widehat{RVarR}' \right]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) / q. \quad (6.29)$$

Relembre aqui que a estimação da variância vem da equação (6.24).

<sup>11</sup>Veja Stock e Watson (2015), capítulo 7, ou Wooldridge (2018), capítulo 4.

Um teste de hipótese normalmente realizado nas rotinas de estimação se assume testar  $\mathbf{r} = \mathbf{0}$  para todos os coeficientes do modelo, logo a estatística-F é:

$$F^R = \beta' [\widehat{Var}]^{-1} \beta / q. \quad (6.30)$$

Na próxima subseção apresentamos a definição de variáveis instrumentais para o caso de uma única variável endógena e um único instrumento. Apesar deste caso parecer simples, ele é importante visto que parte de modelos são bem definidos e aplicados para este caso.

### 6.2.4 Uma Formulação com Único Instrumento

Nesta subseção tratamos do modelo de regressão linear de variáveis instrumentais com uma variável endógena e um instrumento. Portanto este é o caso de um modelo exatamente identificado ( $k = l$ ).

Andrews, Gentzkow e Shapiro (2020) apresentam um exemplo de variáveis instrumentais. Suponha os seguintes dados:  $D = \{Y_i, X_i, Z_i\}_{i=1}^n$ ; estes são observações de uma variável resultado  $Y_i$ , um regressor endógeno  $Z_i$  e um candidato a instrumento  $X_i$ , sendo todas escalares.

$$Y_i = Z_i\beta + X_i a + \varepsilon_i \quad (6.31)$$

$$Z_i = X_i\gamma + V_i \quad (6.32)$$

tal que os instrumentos  $X_i$  são fixados. É consagrado denominar a equação (6.31) de forma estrutural e a equação (6.32) de forma reduzida. No estimador de variáveis instrumentais 2SLS é forma reduzida é chamada de primeiro estágio e a forma estrutural de segundo estágio.

Se assume que o erro  $V_i$  é normal i.i.d. ao longo de  $i$ ,  $V_i \sim N(0, \Sigma)$ , com  $\Sigma$  (variância do primeiro estágio) comumente conhecido.

A estimativa de variáveis instrumentais de  $\beta$  é:

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i Z_i}. \quad (6.33)$$

Por sua vez a estimativa do parâmetro da regressão de primeiro estágio é:

$$\hat{\gamma} = \frac{\sum X_i Z_i}{\sum X_i^2}$$

A estimativa de  $\hat{\beta}$  será válida sob as seguintes condições:

1.  $\hat{\gamma} \neq 0$ , significa que a relação entre a variável endógena  $Z_i$  e o instrumento  $X_i$  existe e é estatisticamente diferente de zero. Esta é chamada de *condição de relevância*.
2.  $a = 0$ , que significa que a variável instrumental  $X_i$  não é correlacionada com a variável de interesse  $Y_i$ . Está denominada de *restrição de exclusão* ou condição de exogeneidade.

3.  $X_i$  deve ser ortogonal ao termo de erro da equação de interesse (6.31), isto é  $E(X_i\varepsilon_i) = 0$ .

### 6.2.5 Validade dos Instrumentos

Instrumentos precisam ser validos para produzir estimações confiáveis. Sem instrumentos com as características adequadas as estimativas não possuem validade. A primeira característica de um bom instrumento é a sua *validade*. Validade do instrumento está relacionado com sua capacidade de explicar a variável endógena e com ortogonalidade em relação à equação estrutural.

Stock e Watson (2015, p. 489) explicam um a validade de instrumento com argumento de tamanho da amostra na estimativa MQO. Quanto mais relevante forem os instrumentos, mas informação é disponível para se usar na estimação de regressão com variáveis instrumentais. Um instrumento mais relevante produz uma estimativa mais precisa. Além disso, a inferência estatística do 2SLS é prevista como este estimador tendo uma distribuição amostral normal, mas de acordo com o teorema do limite central é uma boa aproximação em amostras grandes. Se ter um instrumento mais relevante é como ter uma amostra maior, então quanto mais relevante for o instrumento melhor é a distribuição amostral do estimador 2SLS e de sua estatística  $t$ .

Instrumentos que explicam pouco da variação na variável endógena são chamados de *instrumentos fracos*. Em um modelo de regressão de dois estágios (2SLS), instrumento é fraco quando é pouco relacionado com a variável endógena que precisa ser controlada, isto é, quando o coeficiente de regressão do instrumento é muito pequeno.

Suponha que o o modelo 2SLS é aplicado para a demanda por um grão utilizando dados de volume e preços comercializados. Uma variável candidata a ser instrumento seria a variação da temperatura onde se produz o grão. Intuitivamente os choques climáticos afetariam apenas a oferta e não a demanda. Neste exemplo, a elasticidade da demanda poderia ser estimada utilizando variação na temperatura como instrumento (excluído)  $Z_i$ . Vamos supor que podemos substituir a equação do primeiro estágio (6.35) na equação que desejamos estimar, (6.34). Primeiro considere as equações atualizadas

$$Y_i = Z_i'\beta + W_i'\kappa + \varepsilon_i \quad (6.34)$$

$$Z_i = X_i'\gamma + W_i'\varphi + V_i \quad (6.35)$$

Substituindo  $Z_i$  em (6.34), obtemos:

$$Y_i = X_i'\delta + W_i'\tau + U_i. \quad (6.36)$$

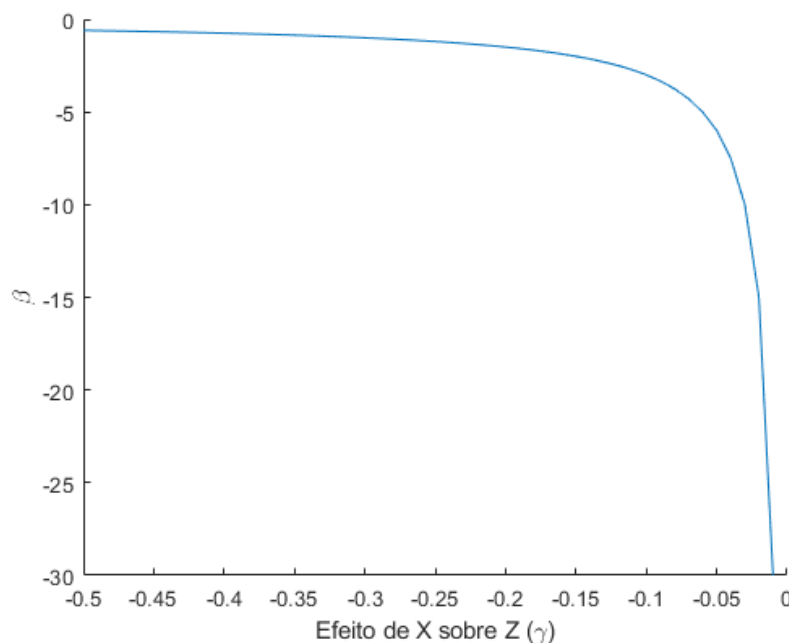
como em Andrews, Stock, e Sun (2019), se tem  $\delta = \gamma\beta$ . A relação entre instrumentos e estimação poderia ser

$$\beta = \frac{\delta}{\gamma}. \quad (6.37)$$

Se a condição de relevância falha, então a variável instrumental não nos ajuda a recuperar o valor verdadeiro da elasticidade de demanda, uma vez que a variação no



Figura 6.1: Efeito de Instrumentos Fracos sobre  $\beta$



instrumento diz pouco sobre a variação no preço sobre a quantidade demandada. Se a relação instrumento preço for zero, então o estimador de  $\beta$  não é definido.

Considere agora que o coeficiente do instrumento no primeiro estágio é bem pequeno, mas diferente de zero. A princípio a elasticidade poderia ser estimada, mas o estimador de variáveis instrumentais deveria ser muito sensível a pequenas variações no denominador, pois a medida que  $\gamma$  se aproxima de zero, o valor do estimador pode degenerar para infinito negativo ou positivo (veja Figura 1 de Mikusheva e Shapiro, 2022, p. 181). A sensibilidade do estimador de variáveis instrumentais ao seu denominador é importante porque, como em qualquer média estatística, o denominador IV é sujeito a variação aleatória devido a uma amostra particular dos dados. A variabilidade estatística resultante do estimador IV é capturada bem pelos erro-padrão e intervalos de confiança tradicionais (Mikusheva e Shapiro, 2022, pp. 180-182).

A situação descrita acima é um caso especial do amplo fenômeno chamado instrumentos fracos, que surge quando o instrumento tem impacto pequeno sobre a variável que ele deveria afetar. Existem algumas alternativas para se testar a validade dos instrumentos.

Para se testar a validade dos instrumentos, quando se tem apenas uma variável endógena, Stock e Yogo (2005)<sup>12</sup> sugerem olhar para o primeiro estágio da regressão de variáveis instrumentais. Especificamente, eles sugerem avaliar uma estatística  $F$  do primeiro estágio. Esta estatística  $F$  testa a hipótese de que os coeficientes dos instrumentos

<sup>12</sup>Veja Stock e Watson (2015, cap. 12).

$Z_{1i}, \dots, Z_{mi}$  são iguais a zero. Quando os instrumentos são fracos, a estimativa 2SLS é viesada e a estatística  $t$ , bem como os intervalos de confiança, não são confiáveis. Se os instrumentos são fracos então, a distribuição normal fornece uma aproximação pobre à distribuição amostral do estimador de dois estágios – mesmo se a amostra for grande. Portanto, não teríamos justificativa teórica para os métodos usuais para realizar inferência estatística.

Para detectar a presença de instrumentos fracos, a literatura propõe utilizar uma estatística  $F$  aplicada a regressão de primeiro estágio do 2SLS.<sup>13</sup> Esta estatística  $F$  é aplicada para todos os regressores do primeiro estágio, que são os instrumentos e as demais variáveis exógenas. Para a variável instrumental não ser fraca, esta estatística  $F$  a 10.

Este valor de 10 é uma regra de bolso definida por Stock e Yogo (2005) e Staiger e Stock (1997). Existem condições importantes que devem ser respeitadas para a aplicação *direta* desta regra. As condições são as seguintes:

- Os erros do primeiro estágio devem ser homocedásticos;
- Deve ter apenas uma variável endógena.

A seguinte condição não é uma grande restrição para o interesse de estimar funções demanda no escopo deste capítulo. A primeira restrição é um problema, pois em análise de mercado não podemos garantir que os erros sejam sempre homocedásticos. Portanto, vamos introduzir um outro teste  $F$  para obter a análise adequada. Recentemente, para o caso de modelos com uma variável endógena e exatamente identificados, Lee, McCrary, Moreira, e Porter (2022) mostraram que o valor de corte deveriam ser 104 se fosse utilizado corretamente a análise proposta por Stock e Yogo (2005). *Para resolver o problema estes autores propõem uma nova análise baseada no teste  $F$  e no teste  $t$  para o modelo 2SLS. Não vamos utilizar o método proposto por Lee, McCrary, Moreira, e Porter (2022), pois frequentemente temos que utilizar modelos sobre-identificados.*

É importante o teste de instrumentos fracos, porque na presença deles o estimador 2SLS é viesado e tem uma distribuição não-normal. Além disso, como mostrado acima, os instrumentos fracos afetam fortemente a estimativa do parâmetro de interesse.

### Teste Anderson-Rubin

Uma alternativa ao teste  $F$  do primeiro estágio seria a estatística Anderson-Rubin (1950). O teste é composto por dois passos: (i) calcule uma nova variável  $Y_i^* = Y_i - \beta_{1,0}x_i$ , tal que  $\beta_{1,0}$  é a estimativa do parâmetro  $\beta_1$  usando instrumentos fracos. (ii) No segundo passo, faça a regressão de  $Y_i^*$  contra os regressores exógenos incluídos ( $x$ ) e os instrumentos ( $z$ ). A estatística Anderson-Rubin é a  $F$  testando a hipótese de o coeficiente dos  $z$ 's são todos zero. Sob a hipótese nula de que  $\beta_1 = \beta_{1,0}$ , se os instrumen-

<sup>13</sup>Veja Stock e Watson (2015, p.490), Andrews, Stock e Sun (2019), e Lee, McCrary, Moreira, e Porter (2022).

tos satisfazem a condição de exogeneidade, eles serão não-correlacionados com o termo de erro nessa regressão, e a hipótese nula será rejeitada em 5% de todas as amostras.

Um conjunto de confiança pode ser construído a partir do conjunto dos parâmetros que não são rejeitados pelo teste de hipótese. O conjunto de valores de  $\beta_1$  que não são rejeitados pelo teste Anderson-Rubin de 5% compõe um intervalo de confiança de 95% para  $\beta_1$ . Quando a estatística é calculada usando a fórmula de homocedasticidade, o intervalo de confiança Anderson-Rubin pode ser construído solucionando uma equação quadrática. Este teste nunca assume relevância dos instrumentos. O intervalo de confiança do teste terá probabilidade cobertura de 95% em amostras grandes, não importando se os instrumentos são fortes, fracos ou mesmo irrelevantes.<sup>14</sup>

**Detalhes (Andrews, Stock, e Sun (2019), seção 5.1).** Especificamente o modelo implica que  $\beta = \frac{\delta}{\gamma}$ , veja em (6.37). Portanto, sob a hipótese nula  $H_0 : \beta = \beta_0$  sabemos que  $\delta - \gamma\beta_0 = 0$  e por conseguinte

$$g(\beta_0) = \hat{\delta} - \hat{\gamma}\beta_0 \sim N(0, \Omega(\beta_0)), \quad (6.38)$$

para

$$\Omega(\beta_0) = \Sigma_{\delta\delta} - \beta(\Sigma_{\delta\gamma} + \Sigma_{\gamma\delta}) + \beta^2\Sigma_{\gamma\gamma}.$$

As matrizes que compõem  $\Omega(\beta_0)$  são:

- $\Sigma_{\delta\delta}$  é a variância de  $\hat{\delta}$ ,
- $\Sigma_{\gamma\gamma}$  é a variância de  $\hat{\gamma}$ , e
- $\Sigma_{\delta\gamma}$  é a covariância entre  $\hat{\delta}$  e  $\hat{\gamma}$ .

A estatística Anderson-Rubin  $AR$  é definida como

$$AR(\beta) = g(\beta)' \Omega(\beta)^{-1} g(\beta), \quad (6.39)$$

e ela segue uma distribuição  $\chi_k^2$  sob  $H_0 : \beta = \beta_0$  não importando o valor de  $\gamma$ . Andrews, Stock e Sun (2019) observam que a Anderson e Rubin (1949) consideram o caso com erro normal homocedásticos. A versão apresentada acima (6.39) é formalmente uma generalização do teste  $AR$  que permite erros não-homocedásticos.

Usando a estatística  $AR$  (6.39) podemos formar um teste  $AR$  de  $H_0 : \beta = \beta_0$  como

$$\phi_{AR} = 1 \left\{ AR(\beta_0) > \chi_{k,1-\alpha}^2 \right\}, \quad (6.40)$$

para a quantil  $1 - \alpha$  de uma distribuição  $\chi_k^2$ . Isto resulta um teste tamanho- $\alpha$  que é robusto a instrumentos fracos (Andrews, Stock e Sun, 2019, Staiger e Stock, 1997).

<sup>14</sup>Quando os instrumentos são fortes, com a regressão 2SLS valendo, e os coeficientes são sobre-identificados, o teste Anderson-Rubin é ineficiente no sentido de ser menos poderoso do que o teste  $t$  do 2SLS. Um teste alternativo é o teste de razão de verossimilhança condicional de Moreira (2003).

Nesta fórmula  $1\{\cdot\}$  representa uma variável indicadora que será igual a 1 em função da regra definida entre as chaves.

Podemos formar um conjunto de confiança robusto a instrumentos fracos,  $CS_{AR}$ , coletando os valores não rejeitados por (6.40). É mais direta a formação do conjunto de confiança no caso homocedástico.<sup>15</sup>

### Testes com erros não-homocedásticos

Os resultados de Stock e Yogo (2005) dependem fortemente da hipótese de homocedasticidade.<sup>16</sup> Andrews, Stock e Sun (2019) recomendam o uso de uma estatística F que seja robusta e aplicável a casos não-homocedásticos. A estatística F robusta sugerida foi proposta por Montiel Olea e Pflueger (2013) e é chamada de estatística F eficiente  $F^{Ef}$ :

$$F^{Ef} = \frac{\hat{\gamma}' \hat{\Sigma}_{\gamma\gamma, N}^{-1} \hat{\gamma}}{tr(\hat{\Sigma}_{\gamma\gamma} \hat{\Sigma}_{XX})} \quad (6.41)$$

aqui  $tr(\cdot)$  é o traço da matriz entre parêntesis, com as seguintes definições de matrizes:

$$\hat{\Sigma}_{XX} = \frac{1}{n} \sum_{i=1}^n X_i X_i',$$

$$\hat{\Sigma}_{\gamma\gamma, N}^{-1} = \frac{\hat{\sigma}_V^2}{n} \hat{\Sigma}_{XX}^{-1}.$$

Montiel Olea e Pflueger (2013) mostram que a distribuição de  $E^{Ef}$  pode ser aproximada por uma distribuição não-central  $\chi^2$ , mas os resultados somente se aplicam ao estimadores 2SLS e LIML. Os testes deles rejeitam quando a estatística-F efetiva excede um valor crítico.

Andrews, Stock e Sun (2019), consideram que para o caso com  $k > 1$  o melhor é utilizar a estatística  $F^{Ef}$  aplicando a regra de corte de 10. A principal conclusão deles é de que  $F^{Ef}$  é a melhor estatística para detectar instrumentos fracos nos casos sobre-identificados não-homocedásticos. Em suma,  $F^{Ef}$  pode ser comparada com os valores críticos de Stock e Yogo (2005) quando  $k = 1$ , e com os valores críticos de Montiel Olea e Pflueger (2013) para  $k > 1$ , ou ainda, com a regra de bolso de valor de corte de 10.<sup>17</sup>

<sup>15</sup>Uma vez que os conjunto de confiança AR possui a cobertura correta independente da força dos instrumentos, sabemos que ele tem tamanho infinito com probabilidade positiva. Especificamente, o  $CS_{AR}$  pode assumir uma de três formas em especificações com um único instrumento: (i) um intervalo limitado  $[a, b]$ , (ii) uma linha real  $(-\infty, \infty)$ , e (iii) a linha real excluindo o intervalo  $(-\infty, a] \cup [b, \infty)$ . Em especificações com mais de um instrumento mas com erros homocedásticos, o conjunto de confiança AR pode assumir uma das três formas ou ainda pode ser vazio.

<sup>16</sup>No caso homocedástico a matriz de variância  $\Sigma$  para  $(\hat{\delta}, \hat{\pi})$  pode ser escrita como produto de Kronecker de uma matriz  $2 \times 2$  com uma matriz  $k \times k$ . Portanto, a análise de Stock e Yogo (2005) não se aplica aqui.

<sup>17</sup>De acordo com Andrews, Stock e Sun (2019), a literatura econômica em alguns casos reporta

Por exemplo, para se estimar a estatística  $F^{Ef}$  no programa Stata é preciso utilizar o comando `weakivtest`. Este comando deve ser implementado após a estimação do modelo de variáveis instrumentais, seja por `ivreg2` ou `ivregress`.<sup>18</sup>

estatísticas F que são robustas a não-homocedasticidade ( $F^R$ ):

$$F^R = \frac{1}{k} \hat{\gamma}' \hat{\Sigma}_{\gamma\gamma}^{-1} \hat{\gamma}, \quad (6.42)$$

enquanto outros relatam estatísticas F tradicional que *não* é robusta à não-homocedasticidade.

$$F^N = \frac{1}{k} \hat{\gamma}' \hat{\Sigma}_{\gamma\gamma}^{-1} \hat{\gamma} = \frac{n}{k \hat{\sigma}_V^2} \hat{\gamma}' \hat{\Sigma}_{XX} \hat{\gamma}, \quad (6.43)$$

tal que

$$\hat{\Sigma}_{\gamma\gamma} = \frac{\hat{\sigma}_V^2}{n} \hat{\Sigma}_{XX}^{-1}$$

e  $\hat{\sigma}_V^2$  é um estimador para  $E[V_i^2]$ .

O uso de estatística-F em setup não-homocedástico é comum em pacotes estatísticos. Por exemplo, quando se estima um modelo 2SLS utilizando o comando `ivreg2` no programa estatístico Stata, automaticamente é gerado a estatística Wald de Kleibergen-Paap (2007) de que  $\gamma$  tem posto reduzido em relação aos valores críticos de Stock-Yogo. Cabe observar que o usuário é avisado de que os valores críticos de Stock e Yogo são para o caso homocedástico. (Ver Baum et al. 2007. Veja também a tabela de valores críticos de Kleibergen-Paap na Tabela 12.5 de Hansen (2022, cap. 12).)

Em configurações com uma única variável endógena, a estatística Wald de Kleibergen-Paap é equivalente a estatística F robusta à não-homocedasticidade  $F^R$  para testar  $\gamma = 0$ . Quando se tem múltiplos regressores endógenos ela é análoga a estatística robusta Cragg-Donald.

A estatística F eficiente de Montiel Olea e Pflueger (2013) é relacionada com  $F^R$  e  $F^N$ :

$$F^{Ef} = \frac{\hat{\pi}' \hat{\Sigma}_{N,\gamma\gamma}^{-1} \hat{\gamma}}{\text{traco}(\hat{\Sigma}_{\gamma\gamma} \hat{\Sigma}_{XX})} = \frac{\text{traco}(\hat{\Sigma}_{\gamma\gamma, N} \hat{Q}_{ZZ})}{\text{traco}(\hat{\Sigma}_{\gamma\gamma} \hat{\Sigma}_{XX})} F^N = \frac{k \hat{\sigma}^2}{\text{traco}(\hat{\Sigma}_{\gamma\gamma}^* \hat{\Sigma}_{XX})} F^N \quad (6.44)$$

Em casos com erros homocedásticos, a estatística  $F^{Ef}$  se reduz na  $F^N$ , enquanto que com erros não-homocedásticos se adiciona uma correção multiplicativa da estimativa robusta da variância. De forma similar, no caso exatamente identificado  $F^{Ef}$  se reduz na estatística  $F^R$  (que é igual a estatística Kleibergen-Paap) enquanto que no caso não-homocedástico  $\hat{\gamma}$  é ponderado por  $\hat{\Sigma}_{XX}$  ao invés de  $\hat{\Sigma}_{\gamma\gamma}^{-1}$ .

As equações do estimador 2SLS (??) e as dos testes  $F^N$ ,  $F^R$  e  $F^{Ef}$  fornecem a intuição porque  $F^{Ef}$  é a estatística apropriada para testar a força do instrumento em 2SLS no caso não-homocedástico. O estimador 2SLS não se comporta bem quando o seu denominador,  $\hat{\gamma}' \hat{\Sigma}_{XX} \hat{\gamma}$ , é próximo de zero. A estatística  $F^N$  mede o mesmo objeto, mas calcula o erro-padrão equivocado e não tem uma distribuição F não-central como em Stock-Yogo. O problema aqui é  $F^N$  pode ser muito grande mesmo quando  $\hat{\gamma}' \hat{\Sigma}_{XX} \hat{\gamma}$  é muito pequeno. Por outro lado, a estatística  $F^R$  mede o objeto populacional errado,  $\hat{\gamma}' \hat{\Sigma}_{\gamma\gamma}^{-1} \hat{\gamma}$ , ao invés de  $\hat{\gamma}' \hat{\Sigma}_{XX}^{-1} \hat{\gamma}$ , embora ela tenha uma distribuição não-central não é correspondente ao estimador 2SLS.

Por último, a estatística  $F^{Ef}$  mede o objeto correto e consegue capturar o erro-padrão correto na média.  $F^{Ef}$  é distribuída como uma média ponderada de variáveis  $\chi^2$  não-central tal que os pesos são positivos e somam um. Os pesos para formar  $F^{Ef}$  são dados pelos autovalores de  $\hat{\Sigma}_{\gamma\gamma}^{1/2} \hat{\Sigma}_{XX} \hat{\Sigma}_{\gamma\gamma}^{1/2} / \text{tr}(\hat{\Sigma}_{\gamma\pi} \hat{\Sigma}_{XX})$ . Montiel Olea e Pflueger (2013) mostram que a distribuição de  $F^{Ef}$  pode ser aproximada por uma distribuição não-central  $\chi^2$ , mas os resultados somente se aplicam ao estimadores 2SLS e LIML. Os testes deles rejeitam quando a estatística-F efetiva excede um valor crítico.

Quando  $k = 1$ , i.e. com apenas um instrumento,  $\Sigma_{\gamma\gamma}$ ,  $\Sigma_{\gamma\gamma, N}$ , e  $\Sigma_{XX}$  são todos escalares e, portanto,  $F^R = F^{Ef}$ . Ambas as estatísticas possuem distribuição F não-central com o mesmo parâmetro de não-centralidade que governa a distribuição do estimador de variáveis instrumentais. Portanto, nesta situação se pode usar os valores críticos de Stock e Yogo (2005).

<sup>18</sup>Para mais detalhes sobre a implementação do teste veja Pflueger e Wang (2015). Importante rodar

**Teste J restrições sobre-identificadoras.** Quando a equação é sobre-identificada então a diferença entre as funções objetivo de duas estimações GMM não é exatamente zero, i.e.,  $J(\tilde{\delta}, \widehat{W}) \equiv n \left[ \mathbf{g}_n(\tilde{\delta})' \widehat{W} \mathbf{g}_n(\tilde{\delta}) \right]$ . Na abordagem GMM ótima, a distância minimizada é assintoticamente qui-quadrada.

- **(Teste de Hansen de sobre-identificação de restrições:)** Suponha que disponível um estimador consistente de  $\mathbf{S}$ ,  $\hat{\mathbf{S}} (= E(\mathbf{g}_i \mathbf{g}_i')$ ). Sob as hipóteses GMM,

$$J(\hat{\delta}(\hat{\mathbf{S}}^{-1}), \hat{\mathbf{S}}^{-1}) = n \cdot \mathbf{g}_n(\hat{\delta}(\hat{\mathbf{S}}^{-1}))' \hat{\mathbf{S}}^{-1} \mathbf{g}_n(\hat{\delta}(\hat{\mathbf{S}}^{-1})) \rightarrow_d \chi^2(K - L). \quad (6.45)$$

Comentários:

- Este é um teste de especificação: se todas as restrições do modelo são satisfeitas. Se a estatística  $J$  é grande significa que tanto as condições de ortogonalidade e/ou outras hipóteses são falsas. Apenas quando temos confiança nas outras hipóteses podemos interpretar a estatística  $J$  grande como evidência para endogeneidade de algum dos  $K$  instrumento incluídos em  $\mathbf{x}_i$ .
- O teste não é consistente contra algumas falhas das condições de ortogonalidade.
- Existem preocupações quanto a aplicação do teste para amostras pequenas.

**Estimação com instrumentos fracos.** Para trabalhar na presença de instrumentos fracos, o economista possui algumas alternativas. Se o modelo for exatamente identificado, então a saída extrema é encontrar instrumentos melhores (fortes). Se o problema empírico tiver muitos instrumentos, tente excluir as variáveis instrumentais mais fracas e observar se o erro-padrão vai crescendo (ele cresce a medida que se exclui os instrumentos mais fracos).

No caso não em que não se pode excluir instrumentos fracos é aconselhável não utilizar o modelo de mínimos quadrados de dois estágios (2SLS). Uma alternativa na presença de instrumentos fracos é o modelo LIML (*limited information maximum likelihood*) – que é contrapartida usando máxima verossimilhança ao 2SLS. O estimador LIML também é o valor de  $\beta_{1,0}$  que minimiza a estatística Anderson-Rubin homocedástica. Portanto, se o intervalo de confiança Anderson-Rubin não é vazio, ele irá conter o estimador LIML. Quando os instrumentos são fracos, o estimador LIML é adequado pois ele é mais centrado no valor verdadeiro  $\beta_1$  do que o 2SLS. Para uma introdução ao estimador LIML veja o capítulo 12 de Hansen (2022).

### 6.2.6 Limites

A abordagem da estimação de um modelo de demanda é limitada e não deve ser vista como uma formulação geral. No sistema de demanda com produtos homogêneos o preço é função dos deslocadores da demanda e da oferta, bem como dos termos de erro. Seguindo a notação de (6.34) e (6.35), faça  $Z_i$  ser o preço, então ele é  $Z_i = Z(W_i, X_i, \varepsilon_i, V_i)$ . O

os modelos utilizando a opção `robust` ou `vce(cluster id)`.

ponto crítico é que o preço depende dos erros estruturais e eles aqui são modelados como escalares. Este tipo de hipótese é mais simples do que modelos considerados completos para produto diferenciado (BLP).

O limite para aplicabilidade de modelos lineares mais simples de variáveis instrumentais reside no grau de substituição entre produtos de firmas diferentes. Em produtos homogêneos isso não costuma ser um problema, mas no mundo prático não existe uma linha clara que divide produtos homogêneos dos diferenciados. Neste sentido, em uma aplicação que a priori os produtos parecem ser homogêneos, mas começam a surgir algum grau de substituição com outros produtos, então podem começar a surgir problemas sérios (e desconhecidos) na aplicação deste tipo de modelo empírico.

## 6.3 Exemplos

### 6.3.1 Demanda por Galeto

Exemplo de demanda agregada por galeto de Epple e McCallum (2006) e Reiss (2016). Suponha um modelo de demanda por um produto homogêneo (na forma log-log):

$$q_c = \beta_0 + \beta_p p_f + \beta_I y + \beta_b p_b + \varepsilon \quad (6.46)$$

nesta equação  $q_c$  é o logaritmo natural da quantidade consumida per capita,  $p_f$  é o log natural do preço real do galeto,  $y$  é o log natural renda real per capita do consumidor e  $p_b$  é o log natural do preço do produto substituto, que no caso é a carne vermelha. Por simplicidade o subscrito das unidades da amostra é omitido. Geralmente os economistas precisam estimar  $\beta_p$ , que representa a elasticidade da demanda (que seria constante). Em outras palavras, seria o percentual constante pela qual a quantidade irá cair com um por cento de aumento no preço, mantendo todas as demais variáveis constantes.

O problema de se estimar  $\beta_p$  por MQO é que  $p_f$  e o termo de erro  $\varepsilon$  são potencialmente correlacionados. Este viés leva a estimativa da elasticidade da demanda,  $\beta_p$ , ser inconsistente e viesada.

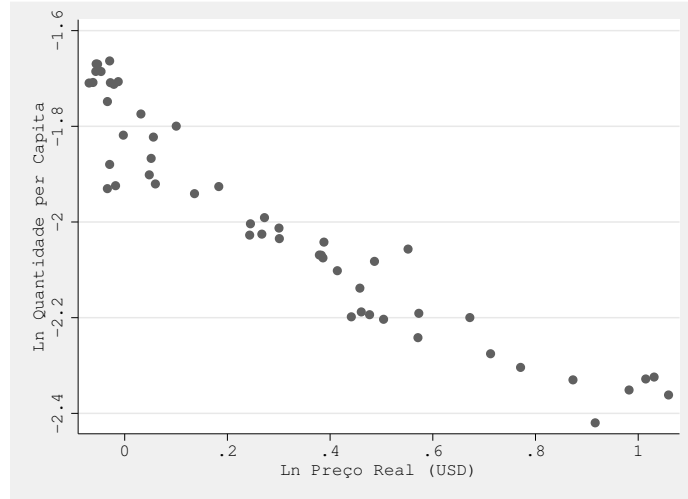
Intuitivamente poderíamos observar a Figura 6.2, onde é apresentado o valor real do frango e o consumo per capital (em libras), e achar que a estimativa de demanda por MQO seria satisfatória. Entretanto, a questão relacionada ao viés de simultaneidade permanece: onde estaria a curva de oferta?

Epple e McCallum (2006) usam preços anuais e quantidades vendidas de galeto nos EUA de 1950 até 2001. Como variável instrumental os autores utilizam o log natural do preço real do milho.<sup>19</sup> Para testar a validade dos instrumentos podemos estimar uma equação em que se relaciona o log natural do preço real do galeto com a renda (log natural real per capita) e o log natural do preço real da carne vermelha. Esta regressão tem a seguinte forma:

$$p_f = \pi_0 + \pi_1 p_c + \pi_2 y + \pi_3 p_b + \eta \quad (6.47)$$

<sup>19</sup>Neste exemplo o preço do galeto, da carne bovina e do milho, são deflacionados pelo índice de preços ao consumidor nos EUA (CPI).

Figura 6.2: Quantidade e Preço do Galeto, EUA: 1950-2001 Fonte:



Os valores estimados por MQO da equação 6.47 são:

$$p_f = \frac{0.2425}{(0.0227)} + \frac{0.2397p_c}{(0.0987)} + \frac{0.4185y}{(0.0793)} - \frac{0.2724p_b}{(0.3191)} \quad (6.48)$$

os erros-padrão dos coeficientes da equação 6.47 estão em parêntesis. O  $R^2$  da equação é .8136 com estatística F de 5.90.

A variável instrumental preço real do milho pode ser considerado um bom instrumento pode ser julgado pela estatística F desta regressão. O valor de 132.97 é alto em bem acima da regra adotada por Stock e Watson (10). Outra evidência de  $Cov(p_f, p_c) \neq 0$  é devido ao coeficiente  $\pi_1$  ser diferente de zero (mesmo que estatisticamente não-significante) no primeiro estágio com F bom.

Na Tabela 6.1 apresentamos o resultado da estimação por 2SLS e MQO do modelo 6.46.<sup>20</sup> O ponto principal que deve ser observado é a diferença no valor da elasticidade preço. Além disso, na presença do viés de simultaneidade não se pode garantir que a estimativa MQO seja ao menos consistente. Observamos que o coeficiente estimado por 2SLS é -1.71, que apresenta muito mais elasticidade do que a estimativa (inconsistente) de MQO.

### 6.3.2 Demanda por Cigarros

Parte do interesse na demanda por cigarros está relacionada com saúde pública. Como o uso de cigarros está relacionado a diversas doenças e uso da rede pública de saúde, a aplicação de impostos pode efetivamente reduzir o consumo de cigarros. Aqui vamos

<sup>20</sup>Os erro-padrão são erros robustos a heterocedasticidade. Como descrito anteriormente, os erro-padrão são substancialmente maiores na estimativa 2SLS.



Tabela 6.1: Estimativa da Demanda por Galeto: 2SLS vs. MQO

	2SLS	MQO
$\beta_p$	-1.71 (.9406)	-.2647 (.0671)
$\beta_I$	-.416 (.8361)	.8519 (.0297)
$\beta_b$	.09355 (.321)	-.1183 (.0738)
$\beta_0$	7.869 (.8.272)	-4.679 (.577)
F	70.86	623.59
N	52	52

seguir o exemplo dado por Stock e Watson (2015, cap. 12) de estimação da demanda por cigarros.

Por exemplo, qual deve ser a mudança de imposto para gerar uma queda de 20% no consumo de cigarros? A resposta a esta questão depende da elasticidade da demanda por cigarros. Se a elasticidade é -1 então a redução de 20% pode ser atingida com um aumento de 20% no preço final. Se a elasticidade for -0.5, então o preço final deve subir 40% para gerar a redução de 20% no consumo.

Stock e Watson (2015, p. 480) estimam a demanda por cigarro usando o modelo 2SLS. Eles utilizam dados anuais para 48 estados americanos de 1985 até 1995. Consumo de cigarros é  $Q_i$ , medido pelo numero de pacotes vendidos de cigarro per capita em cada estado  $i$ . O preço,  $P_i$  é o preço real médio por maço, incluindo todos os impostos.

Inicialmente Stock e Watson consideram como variável instrumental o imposto vendas sobre o cigarro. No caso desta aplicação os impostos são medidos em dólares por maço (em valor real, deflacionado pelo índice de preço ao consumidor dos EUA (CPI)).<sup>21</sup>

Segundo Stock e Watson, primeiro estimamos a elasticidade-preço de longo prazo dos cigarros, ou pelo menos um pouco mais longa do que a variação anual. Será estimada a elasticidade-preço para o período de 10 anos. Como o que interessa é o comportamento

<sup>21</sup>Artigo de Dearing (2022) mostra que condicional ao preço pago pelo compradores ou recebido pelos vendedores – a alíquota de imposto (tax rate) é excluída de ambas as equações. Inicialmente para usar impostos como instrumentos deveria se assumir demanda linear, concorrência perfeita, etc. Dearing (2022) mostra que se o setup é de produtos diferenciados (similar a modelos estruturais), você pode restabelecer identificação para um sistema multi-produto com uma *única* mudança na alíquota de imposto. Este resultado é muito útil porque mostra que um dos instrumentos mais comuns é válido. Em grande parte, a dúvida comum é se bastaria apenas uma mudança de alíquota para identificar o sistema.

para esta janela de tempo, a variável dependente quantidade será  $(\ln Q_{i,t} - \ln Q_{i,t-10})$ . Por exemplo, se  $t$  é medido em anos e o último ano de observação é 1995, então teremos  $(\ln Q_{i,1995} - \ln Q_{i,1985})$ . As demais variáveis são definidas analogamente:  $(\ln P_{i,t} - \ln P_{i,t-10})$  e  $(\ln R_{i,t} - \ln R_{i,t-10})$ , tal que  $P$  é o preço real médio por maço e  $R$  é a renda real per capita de cada estado americano  $i$ .

Primeiro observamos a relação do preço médio, a variável endógena, com os instrumentos: com dois instrumentos e a relação do preço com cada instrumento por vez. A relação apresentada é a seguinte:

$$(\ln P_{i,t} - \ln P_{i,t-10}) = \eta_0 + \sum_{j=1}^2 \eta_j (Z_{i,t}^j - Z_{i,t-10}^j) + \eta_3 (\ln R_{i,t} - \ln R_{i,t-10}) + V_{i,t}$$

Observe que esta equação tem dois instrumentos,  $Z^1$  e  $Z^2$ . Faça  $Z^1 = \tau^{\text{cigarros}}$  e  $Z^2 = \tau^{\text{vendas}}$ , tal que  $\tau^{\text{cigarros}}$  é o imposto específico sobre cigarros e  $\tau^{\text{vendas}}$  é o imposto sobre as vendas aplicado a todos os bens de consumo. Especificamente, o imposto sobre vendas é o valor do imposto médio em *cents* por maço. Observe que os instrumentos não estão em ln. Estas regressões são apresentadas na Tabela 6.2.

Tabela 6.2: Regressão de Primeiro Estágio: Demanda por Cigarros

Variáveis	2 Instrumentos	Apenas:	
		$\tau^{\text{vendas}}$	$\tau^{\text{cigarros}}$
$\tau_{i,t}^{\text{cigarros}} - \tau_{i,t-10}^{\text{cigarros}}$	0.0076 (0.0009)		0.0101 (0.0010)
$\tau_{i,t}^{\text{vendas}} - \tau_{i,t-10}^{\text{vendas}}$	0.0135 (0.0031)	0.0255 (0.0044)	
$\ln R_{i,t} - \ln R_{i,t-10}$	-0.0290 (0.1242)	-0.2241 (0.2189)	0.0294 (0.1310)
Constante	0.1439 (0.0183)	0.1841 (0.0295)	0.1548 (0.0201)
N	48	48	48
$F^{Ef}$ -Instrumentos	82.97	33.67	107.18
R2	0.7779	0.5146	0.6796

Fonte: Erro-padrão robusto de White entre parêntesis.  $F^{Ef}$  é a estatística  $F$  efetiva de Montiel-Pflueger.

A primeira análise que deve ser feita é sobre a relevância dos instrumentos utilizados no modelo de demanda por cigarros (como discutido na seção 6.2.5). O melhor modelo que consideramos é o com dois instrumentos, apresentado na primeira coluna da Tabela 6.2. Para se testar validade dos instrumentos se observa o teste  $F^{Ef}$  exclusivo para os dois instrumentos,  $\tau^{\text{vendas}}$  e  $\tau^{\text{cigarros}}$ , que é 82.97. Como o valor é bem acima de 10, se conclui

que os instrumentos não são fracos. O teste  $F^{Ef}$  também pode ser calculado para os outros dois modelos, mesmo tendo apenas um instrumento cada. Para estes dois casos os valores do teste  $F^{Ef}$  são 33.7 e 107.2, para os modelos apenas com  $\tau^{\text{vendas}}$  e  $\tau^{\text{cigarros}}$ , respectivamente. Neste exemplo, podemos concluir que os instrumentos não são fracos.

A estimativa de 2SLS da relação de longo prazo da demanda por cigarros

$$[\ln Q_{i,t} - \ln Q_{i,t-10}] = \beta (\ln P_{i,t} - \ln P_{i,t-10}) + \alpha_I (\ln R_{i,t} - \ln R_{i,t-10}) + \varepsilon_{i,t} \quad (6.49)$$

Os resultados para este modelo com diferentes regressões 2SLS são apresentados na Tabela 6.3. Cada modelo de regressão utiliza um conjunto diferente de instrumentos. Abaixo dos coeficientes estão os intervalos de confiança a 5%. Observe que todos os parâmetros da elasticidade são significantes.

Podemos perguntar se os instrumentos são exógenos. Podemos usar a estatística  $J$  de Hansen para testar a exogeneidade, lembrando que ela pode ser aplicado apenas ao caso sobre-identificado. No modelo 2SLS com dois instrumentos a estatística é 4.085, com p-valor de 0.0433. Como a hipótese nula é de que ambos os instrumentos são exógenos, o p-valor nos diz que a rejeitamos. A rejeição da hipótese nula nos diz que pelo menos UM instrumento não é exógeno.

Tabela 6.3: Estimativa da Demanda por Cigarro

	2SLS (imposto específico)	2SLS (imposto geral)	2SLS (2 Instr.)
$\beta_p$	-1.3425 [-1.77645, -0.9085794]	-0.9381 [-1.331797, -0.5442319]	-1.202 [-1.576148, -0.8286589]
$\beta_I$	0.4281 [-0.138739, 0.9950303]	0.5259 [-0.1182979, 1.170236]	0.4620 [-0.1250138, 1.049074]
Estatística-J			4.085
(p-valor)			0.0433
F	20.57	12.31	21.30
N	48	48	48

A rejeição da estatística J diz que pelo menos um dos instrumentos é endógeno. Neste caso se tem três possibilidades: o imposto sobre bens é exógeno mas o imposto sobre cigarros não é; outra hipótese seria o reverso; por fim, ambos os impostos seriam endógenos. A estatística não pode nos dizer qual o instrumento possui problemas.

Observe que a estimativa do parâmetro da elasticidade é menor do que um (em modulo) quando se utiliza o imposto pago sobre todos os bens. Nos demais modelos o parâmetro é maior do que a unidade. Stock e Watson sugerem que o instrumento do segundo modelo possui problemas, pois geralmente se espera que a demanda seja elástica em períodos mais longos.

## 6.4 GMM<sup>22</sup>

GMM é o Método Generalizado dos Momentos (*Generalized Method of Moments*) é o método geral para estimação de parâmetros de modelos lineares e não-lineares, em que os parâmetros são escolhidos para fornecer a melhor aproximação para um conjunto de equações múltiplas, cujo momentos amostrais são zero. Estas equações, que no contexto do GMM são denominadas condições dos momentos, tipicamente não podem ser todas satisfeitas simultaneamente. O estimador GMM *trade off* o desejo de satisfazer cada equação pela minimização de uma função objetivo quadrática. O GMM generaliza os métodos clássicos de estimação ao permitir que os modelos tenham mais equações do que parâmetros desconhecidos (caso sobre-identificado) introduzido no artigo seminal de Lars Peter Hansen (1982). GMM trata MQO, variáveis instrumentais, regressão multivariada e 2SLS como casos especiais, além de permitir a estimação de modelos lineares e não-lineares.

Se os erros são heterocedásticos, então o estimador 2SLS não é o mais eficiente na classe dos estimadores de variáveis instrumentais que usa combinações lineares dos instrumentos. O estimador eficiente neste caso é o *GMM eficiente*.

Todos os modelos lineares de econometria para a graduação podem ser escritos como modelos de equações de momentos, tal que os parâmetros populacionais resolvem um sistema de equações dos momentos. Normalmente os modelos de equações dos momentos assumem a seguinte forma: tome  $g_i(\beta)$  ser uma função conhecida  $l \times 1$  da  $i$ -ésima observação e  $\beta$  ser um vetor de parâmetros com dimensão  $k \times 1$ . Um modelo de equação dos momentos é resumido como:

$$E[g_i(\beta)] = 0, \quad (6.50)$$

e um espaço dos parâmetros  $\beta \in B$ . Por exemplo, no caso do modelo de variáveis instrumentais o momento é a condição de ortogonalidade:  $g_i(\beta) = X_i \varepsilon_i = X_i(Y_i - Z_i' \beta)$ .

**Identificação.** Dizemos que o parâmetro  $\beta$  é identificado se existe um mapeamento único da distribuição dos dados para  $\beta$ . No contexto do modelo (6.50), isto significa que existe um único  $\beta$  satisfazendo o sistema de  $l$  equações com  $k$  incógnitas. Portanto, a condição para solucionar o modelo é  $l \geq k$ .

### 6.4.1 Estimador do Método dos Momentos

Primeiro vamos considerar o caso em que  $l = k$ , i.e., o caso exatamente identificado. Definimos a equação amostral análoga a (6.50) como:

$$\bar{g}_n(\beta) = \frac{1}{N} \sum_{i=1}^n g_i(\beta). \quad (6.51)$$

<sup>22</sup>Adaptado de Hansen (2022), cap. 13. Veja também Stock e Watson (2015), seção 18.7.

O estimador do método dos momentos (MM)  $\hat{\beta}_{mm}$  de  $\beta$  é definido como o valor dos parâmetros que satisfaz  $\bar{g}_n(\beta) = 0$ . Portanto

$$\bar{g}_n(\hat{\beta}_{mm}) = \frac{1}{N} \sum_{i=1}^N g_i(\hat{\beta}_{mm}) = 0. \quad (6.52)$$

Em alguns contextos existe solução explícita para  $\hat{\beta}_{mm}$  em (6.52), em outros casos a solução deve ser encontrada numericamente (por otimização). A maioria dos estimadores pode ser escrito utilizando o método dos momentos, fazendo do GMM uma abordagem *unificadora*. Por exemplo, o estimador de variáveis instrumentais pode ser escrito da seguinte forma: como anteriormente faça  $g_i(\beta) = X_i \varepsilon_i = X_i(Y_i - Z_i' \beta)$ . Portanto o estimador MM é  $\hat{\beta}_{mm} = \left( \sum_{i=1}^N X_i Z_i' \right)^{-1} \left( \sum_{i=1}^N X_i Y_i \right)$ .

#### 6.4.2 Equações de Momentos Sobre-identificados

No caso do modelo de variáveis instrumentais linear,  $g_i(\beta) = Z_i \varepsilon_i = Z_i(Y_i - X_i' \beta)$ , o sistema empírico pode ser escrito como:

$$\bar{g}_n(\beta) = \frac{1}{N} \sum_{i=1}^N g_i(\beta) = \frac{1}{N} \sum_{i=1}^N X_i(Y_i - Z_i \beta) = \frac{1}{N} (X'Y - X'Z\beta). \quad (6.53)$$

A solução para o sistema é direta no caso do modelo exatamente identificado, entretanto, no caso sobre-identificado existem mais equações do que parâmetros livres. Isto significa que não existe escolha de  $\beta$  que faça (6.50) igual a zero. Apesar de não conseguir achar os parâmetros que satisfaçam (6.50), podemos encontrar os parâmetros que façam com que  $\bar{g}_n(\beta)$  seja o *o mais perto possível* de zero.

Para mostrar como podemos pensar este problema podemos reescrever os elementos de (6.53) como:

- $\mu = X'Y$ ;
- $G = X'Z$ ;
- e um “novo” termo de erro  $\eta = \mu - G\beta$ . Relembre que o termo de erro nesta formulação é  $\bar{g}_n(\beta)$ .

Neste caso, podemos reescrever (6.53) como

$$\mu = G\beta + \eta.$$

Esta equação se parece com uma equação de regressão com a variável dependente  $\mu$  ( $l \times 1$ ), o matrix de regressores  $G$  com dimensão  $l \times k$  e o termo de erro  $\eta$ . O objetivo aqui é fazer com que o vetor de erro seja o menor possível. Sabemos que uma solução simples é estimar por MQO a relação entre  $\mu$  e  $G$ , que minimiza a soma dos quadrados

$\eta' \eta$ . Esta é uma forma de fazer com que  $\eta$  seja pequeno, com a estimação “MQO”  $\hat{\beta} = (G'G)^{-1} (G'\mu)$ .

Sabendo que os erros são não-homogêneos, uma solução mais eficiente é estimar o modelo por mínimos quadrados ponderados.<sup>23</sup> O estimador por mínimos quadrados ponderados é  $\hat{\beta} = (G'WG)^{-1} (G'W\mu)$ , para uma matriz ponderadora  $W$ . Em termos das variáveis originais o estimador será:

$$\hat{\beta} = (Z'XWX'Z)^{-1} (Z'XWX'Y).$$

Essa estimação minimiza a soma ponderada do quadrado dos erros:  $\eta'W\eta$ . Relembre que  $\eta = \bar{g}_n(\beta)$ . Esta solução é conhecida como método generalizado dos momentos (GMM). Partindo desta estrutura de erro quadrático e assumindo matriz de pesos  $W > 0$  ( $l \times l$ ), a função objetivo ou critério do problema GMM é definida como:

$$J(\beta) = n \cdot \bar{g}_n(\beta)' W \bar{g}_n(\beta). \quad (6.54)$$

Quando  $W = I_l$ , então  $J(\beta) = n \|\bar{g}_n(\beta)\|^2$ , o quadrado da distância Euclidiana. Como a atenção é restrita a matrizes positivas definidas  $W$ ,  $J(\beta)$  é sempre não-negativo. Cabe observar que o fator  $n$  em (6.54) não é importante para a definição do estimador, mas é conveniente para a teoria distribucional.

O estimador GMM é definido como o minimizador da função objetivo em (6.54):

$$\hat{\beta}_{gmm} = \min_{\beta} J_n(\beta). \quad (6.55)$$

Alguns comentários. No caso exatamente identificado  $k = l$ , o estimador do método dos momentos,  $\hat{\beta}_{mm}$ , é igual ao estimador GMM,  $\hat{\beta}_{gmm}$ . Isto significa que o estimador do método dos momentos é um caso particular do estimador GMM. Todavia, o estimador GMM depende da matriz de pesos  $W$  e veremos a seguir como estimá-la.

Portanto o estimador GMM é encontrado ao minimizar a função objetivo (6.54):

$$\min_{\beta} J(\beta) = \min_{\beta} \{n \cdot (X'(Y - Z\beta))' W (X'(Y - Z\beta))\}. \quad (6.56)$$

A solução para o problema de primeira ordem da função objetivo GMM é:<sup>24</sup>

$$\hat{\beta}_{gmm} = (Z'XWX'Z)^{-1} (Z'XWX'Y). \quad (6.57)$$

<sup>23</sup>Sobre mínimos quadrados ponderados veja Stock e Watson (2015) seção 17.5 ou Wooldridge (2016) seção 8.4.

<sup>24</sup>A condição de primeira ordem para o problema é:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta} J(\hat{\beta}) \\ 0 &= 2 \frac{\partial}{\partial \beta} \bar{g}_n(\beta)' W \bar{g}_n(\beta) \\ 0 &= -2 \left( \frac{1}{n} Z'X \right) W \left( \frac{1}{n} X'(Y - Z\hat{\beta}) \right). \end{aligned}$$

Observem que o estimador  $\hat{\beta}_{gmm}$  depende de  $W$ , mas ele não muda se  $W$  for substituída por  $cW$  para algum  $c > 0$ . Comumente fazemos:

$$W = \left( \frac{1}{N} X'X \right)^{-1}.$$

Quando  $W$  é fixa, este estimador GMM é denominado de estimador GMM de um passo (*one-step GMM estimator*). A matriz para este caso é  $W = (X'X)^{-1}$ . No caso exatamente identificado  $k = l$ , este é o estimador mais eficiente. Neste caso, o estimador resultando também é igual ao 2SLS:  $\hat{\beta}_{gmm} = \hat{\beta}_{2sls}$ . Os dois estimadores são iguais quando  $W = (X'X)^{-1}$ .

Sobre hipóteses padrão, o estimador GMM apresenta características similares ao 2SLS. Portanto,

$$\sqrt{n} (\hat{\beta} - \beta) \rightarrow_d N(0, V_\beta), \quad (6.58)$$

tal que

$$\widehat{Var} = (\hat{\Sigma}_{ZX} W \hat{\Sigma}_{XZ})^{-1} (\hat{\Sigma}_{ZX} W \hat{S} W \hat{\Sigma}_{XZ}) (\hat{\Sigma}_{ZX} W \hat{\Sigma}_{XZ})^{-1}. \quad (6.59)$$

### 6.4.3 GMM Eficiente

Quando o estimador é construído com  $W = W_0 = S^{-1}$ , chamamos o estimador de GMM eficiente. Não apresentaremos a prova da construção da matriz eficiente.

O estimador eficiente é construído em duas etapas, uma vez que é preciso o termo de erro (resíduo) da primeira estimação GMM. O estimador GMM eficiente funciona para o caso sobre-identificado. No caso exatamente identificado  $k = l$  a primeira etapa já eficiente.

Faça o termo de erro ser  $\hat{\varepsilon} = y_i - z_i \hat{\beta}$ , o momento ser  $\tilde{g}_i(\tilde{\beta}) = X_i \hat{\varepsilon}$  e  $\bar{g}_n = \sum_{i=1}^N \tilde{g}_i$ . Nesse caso a matriz eficiente é

$$\hat{S} = \frac{1}{N} \sum_{i=1}^N \tilde{g}_i' \tilde{g}_i \quad (6.60)$$

Procedimento de dois estágios de estimador GMM eficiente.

1. Escolha uma matriz  $\hat{W}$  que converge em probabilidade para uma matriz simétrica positiva definida, e minimize  $J(\tilde{\beta}, \hat{W})$  sobre  $\tilde{\beta}$  para obter  $\beta(\hat{W})$ . Usualmente fazemos  $\hat{W} = ((1/N)X'X)^{-1}$ . Esse estimador é o conhecido método de mínimos quadrados de dois estágios (2SLS). Use isto para calcular o resíduo  $\hat{\varepsilon}$  e obter um estimador consistente  $\hat{S}$  de  $S$  como em (6.60).
2. Minimize  $J(\tilde{\beta}, \hat{S}^{-1})$  sobre  $\tilde{\beta}$ . O minimizador é o estimador GMM eficiente.

Para rodar esse modelo use o comando `gmm` ou `ivregress gmm`, para ambos os casos o modelo GMM é estimado em duas etapas.

## 6.5 Estimação do Modelo de Cournot Linear<sup>25</sup>

Imagine uma cross-section de mercados, indexados por  $n$ . Vamos começar com uma função demanda linear:

$$p_n = x_n \beta_n - \beta_q Q_n + \varepsilon_n \quad (6.61)$$

tal que, como no modelo concorrencial,  $\varepsilon$  e  $x$  são determinantes não-observáveis e observáveis, respectivamente, do nível da demanda.

Por sua vez, se  $w$  são determinantes observáveis dos custos e garantindo que  $E[\varepsilon | x, w] = 0$ , então podemos estimar os parâmetros da função demanda utilizando técnicas de variáveis instrumentais. Para o lado da oferta temos um conjunto de  $J$  CPOs para cada mercado. Assuma custo marginal linear tal que

$$CMg_{n,j} = w_{n,j} \gamma + \lambda q_{n,j} + \omega_{n,j} \quad (6.62)$$

Então, se  $q_{n,j} > 0$ , a CPO do modelo de Cournot ( $p - CMg_j - q_j(\partial p / \partial q)$ ) pode ser reduzida para

$$p_n - w_{n,j} \gamma - (\lambda + \beta_q) q_{n,j} - \omega_{n,j} = 0. \quad (6.63)$$

tal que  $\omega_{n,j}$  é a produtividade da firma  $j$  no mercado  $n$ .

Dado que  $E[\omega | x, w] = 0$ , podemos estimar a equação (6.63). Para estimar este modelo, observe que se tem  $N \times J$  condições na forma  $E[\omega_{n,j} | x_j, w_{n,j}] = 0$ . Primeiro escreva a (6.63) em função de  $\omega$

$$\omega_{n,j}(\theta) = p_n - w_{n,j} \gamma - (\lambda + \beta_q) q_{n,j},$$

nesta equação  $\theta = (\gamma, \lambda, \beta_q)$  é o vetor de parâmetros a ser estimado.

Para estimar encontre um vetor de instrumentos suficientemente rico da função  $f(x, w)$  e forme as condições de momento amostral  $G_n(\theta)$ :

$$G_n(\theta) = \frac{1}{NJ} \sum_{n,j} \omega_{n,j}(\theta) f(x_n, w_{n,j}) \quad (6.64)$$

Para estimar, use um algoritmo de busca para encontrar os valores de  $\theta$  que façam  $\|G_n(\theta)\|$  ser bem próximo de zero.

Além disso, observe que produto e preços dependem de toda a distribuição dos deslocadores de custo que se está usando como instrumento. Isto pode nos fornecer mais instrumentos do que os disponíveis para um modelo de concorrência perfeita.

O quanto a quantidade ótima será zero depende dos custos fixos e dos parâmetros que o pesquisador estima. Se a quantidade for zero então você apenas observa  $q$  para um subconjunto particular de  $\omega$ , então se a amostra original foi “sorteada” da população satisfazendo  $E[\omega | x, w] = 0$  então geralmente  $E[\omega | x, w, q > 0] \neq 0$  então você terá

<sup>25</sup>Seguindo notas de aula de Ariel Pakes.



um problema similar ao problema de seleção que será discutido no capítulo 12, sobre estimação de função de produção.

Para lidar com o caso de  $q = 0$  é preciso construir um modelo de quando e porque produção zero ocorre – produção zero também poderia ser chamada de hibernação. Se o economista utilizar o modelo acima e especificar uma distribuição conjunta para o custo fixo e  $\omega$ , pode então trabalhar a condição  $E[\omega \mid x, w, q > 0]$ . Se o pesquisador tiver dados quando planta/firma faz  $q = 0$ , então o modelo pode ser mais direto e estimar a probabilidade com que  $q = 0$ , i.e.  $\text{Prob}(q = 0 \mid x, w)$ .

**Identificação.** Observe que a CPO para quantidade (6.63) não tem o poder *per se* de determinar a inclinação do custo marginal ou da função demanda. É apenas pela combinação dos coeficientes de inclinação do custo marginal e da demanda que podemos identificar *ambos*. Outra forma de dizer isso, é que no caso linear as CPOs sozinhas não podem nos dizer quando o mercado age “como se” ele fosse populado por firmas tomadoras de preço ou por competidores Cournot.

## Apêndice

### Construção das Variáveis: Demanda por Cigarros

No modelo estimado todas as variáveis são reais, i.e. deflacionadas pelo índice de preços ao consumidor. Os dados utilizados por Stock e Watson (2015) foram cedidos por Jonathan Gruber. O preço médio do cigarro, impostos (geral e específico sobre cigarro) e renda são deflacionados pelo índice de preços ao consumidor (CPI). Como a variável renda não é per capita, ela deve ser dividida pela população dos estados americanos para construção da renda per capita.

## 6.6 Bootstrap

Bootstrap é um método de simulação para calcular erros-padrão e distribuições de estatísticas de interesse, que emprega um processo gerar dos dados (DGP) estimado para gerar amostras artificiais (bootstrap) e computar as estatísticas por meio de sorteios de dados (bootstrap). Bootstrap empírico ou não-paramétrico se apoia em estimativas não-paramétricas do DGP, enquanto bootstrap paramétrico usa estimativas paramétricas do DGP (Chernozhukov e Fernández-Val, 2017). O foco da literatura empírica em economia tem sido no bootstrap empírico.

Em resumo, a distribuição bootstrap é obtida pela estimação baseada em amostras independentes criada pela amostragem iid a partir dos dados originais. Este sorteio iid é baseada em amostragem com reposição.

**Amostragem com repetição.** Para ficar claro o conceito é possível utilizar um exemplo com 20 observações utilizando dados do CPS, seguindo Hansen (2022), cap. 3.

Tabela 6.4: Observações da CPS

Observação	Wage	log(Wage)	Education	Education <sup>2</sup>	Education*log(Wage)
1	37,93	3,64	18	324	65,44
2	40,87	3,71	18	324	66,79
3	14,18	2,65	13	169	34,48
4	16,83	2,82	16	256	45,17
5	33,17	3,50	16	256	56,03
6	29,81	3,39	18	324	61,11
7	54,62	4,00	16	256	64,00
8	43,08	3,76	18	324	67,73
9	14,42	2,67	12	144	32,03
10	14,90	2,70	16	256	43,23
11	21,63	3,07	18	324	55,44
12	11,09	2,41	16	256	38,50
13	10,00	2,30	13	169	29,93
14	31,73	3,46	14	196	48,40
15	11,06	2,40	12	144	28,84
16	18,75	2,93	16	256	46,90
17	27,35	3,31	14	196	46,32
18	24,04	3,18	16	256	50,76
19	36,06	3,59	18	324	64,53
20	23,08	3,14	16	256	50,22
Sum		62,64	314	5010	995,86

A amostragem a partir desta distribuição significa selecionar aleatoriamente uma linha desta tabela. Por exemplo, use um gerador de números aleatórios entre 1 e 20. Suponha que a seleção seja 16, o que implica que devemos selecionar a linha 16. O indivíduo da linha 16 possui salário de 18,75 USD e 16 anos de educação. Continuando o processo de amostragem se seleciona a linha 5, isto significa a observação da linha 5, cujo indivíduo possui salário de 33,17 USD com 16 anos de educação. Continuando com o processo aleatório de sorteio das linhas no total de 20 sorteios e chegamos a  $\{16, 5, 17, 20, 20, 10, 13, 16, 13, 15, 1, 6, 2, 18, 8, 14, 6, 7, 1, 8\}$ . Este sorteio aleatório das linhas da base de dados é denominada de amostra bootstrap.

No procedimento de sorteio aleatório algumas linhas se repetem. 1, 6, 8, 13, 16, 20 aparecem duas vezes na nova amostra, enquanto que outras linhas não aparecem. Esta é a natureza da amostragem com reposição. Uma vez que o número 16 foi sorteado, ele retorna para a base de dados. Como o processo de sorteio é aleatório este número pode ser sorteado novamente. Portanto, a amostra bootstrap é uma reamostragem com reposição da amostra original. Seguindo a notação da literatura de econometria, se escreve a observação  $i$  da amostra bootstrap como  $(y_i^*, z_i^*)$  e a amostra bootstrap como  $\{(y_1^*, z_1^*), \dots, (y_n^*, z_n^*)\}$ .

*Estimação.* Considere a estimação do parâmetro de interesse  $\hat{\beta}^*$  por MQO a partir da amostra bootstrap. A partir da estimação realizada por MQO se tem



## Capítulo 7

# Determinação de Preços: Produto Diferenciado

Anteriormente assumimos demanda por produtos homogêneos em situações de oligopólio ou monopólio. Esta é uma abordagem razoável para mercado de commodities, onde o preço governa a demanda. Agora, vamos deixar exclusivamente o “espaço das quantidades” e apresentar discussões de determinação de estratégias de mercado em que o preço não é a único determinante do volume demandado.

Primeiro vamos apresentar princípios de demanda por produto diferenciado. Em seguida uma aplicação de sistemas de demanda inspirado em produto homogêneo para o caso de diferenciado. Por fim apresentamos modelos de demanda de produtos diferenciados. Também explicamos quando utilizar cada modelo e porque são necessários.

### 7.1 Demanda por Produtos Diferenciados I

Inicialmente um bom exemplo é o mercado de casas. Cabral (2017, p. 350) considera um exemplo interessante pois seria o oposto ao mercado de commodities: geralmente cada casa tem características que são únicas – mesmo casas similares em uma mesma vizinhança possuem localização distintas. Desconsiderando a localização, cada casa poderia ser bem descrita por um conjunto de características: área total, número de quartos, número de pisos, estilo, etc.

Suponha que todos os consumidores concordam sobre quanto cada característica vale em reais; e que um único vendedor domina o mercado. Estas duas hipóteses são fortes e serão relaxadas depois. Sob estas condições, podemos estimar um modelo de regressão linear do preço como função destas características. Por exemplo, na Tabela 7.1 apresentamos coeficientes selecionados de um modelo para preços de imóveis em Brasília.<sup>1</sup> Observe que uma variável muito importante é o tamanho da residência, com coeficiente de 0.4422

---

<sup>1</sup>Variáveis selecionadas porque este tipo de modelo geralmente possui muitas variáveis.

(e significativo). A medida que a residência fica mais distante da região central da cidade, o seu preço vai se reduzindo mesmo controlando por outras características do imóvel. Também é bom explicar que este tipo de modelo geralmente possui R2 elevado, i.e. ele é capaz de contabilizar por grande parte da variação nos preços. Entretanto, não podemos realizar muitas interpretações a partir dos coeficientes estimados por uma regressão hedônica. Aqui estamos baseando a interpretação no uso de hipóteses fortes.

Tabela 7.1: Determinantes do (ln) Preço de Residências em Brasília (MQO)  
Variáveis Seleccionadas

Variável	Coefficiente	Estatística T (EP White)
Constante	9,3327	34,76
Área	0,4422	11,41
Número de Quartos	0,0373	1,08
Número de Banheiros	0,0218	0,84
Apartamento	-0,0753	-0,73
Andar	-0,0075	-0,67
Idade	-0,0038	-3,00
Distância do Centro (Km)	-0,0547	-11,92

Variáveis estão em log natural. Fonte: Rezende (2009), p. 27..

Em casos com poucos produtos e marcas o modelo pode ser mais informativo para entender estratégias de mercado. Cabral (2017, p. 351) considera o caso da demanda por Diet Pepsi e Coca-Cola Diet. Neste caso é possível estimar diretamente a curva de demanda considerando diretamente o preço destes produto, no segmento de refrigerante diet. Esta estratégia não seria adequada para o caso de demanda por residências.

### 7.1.1 Introdução ao BLP<sup>2</sup>

O conceito da estimação de demanda por residência é denominado de **abordagem de características**. Ao invés de estimar por cada casa individual estimamos a demanda por cada característica. O preço implícito em cada característica que foi utilizado no modelo de regressão de imóveis é referido como *preço hedônico*. A utilização de regressão de preços hedônicos não é adequada para qualquer caso, tal que muitas complicações podem surgir quando aplicamos uma estimativa como esta. Considere por exemplo a demanda por carros – que é um caso bem estudado na literatura de economia industrial. Distintamente do assumido anteriormente, carros não são vendidos por um monopolista, mas o mais importante é que todos *os consumidores não valorizam igualmente as características dos veículos*. Um modelo mais realístico de demanda permite que a possibilidade do coeficiente da utilidade de uma dada característica varie entre tipos de consumidores. Importante observar que o modelo de demanda não precisa definir uma

<sup>2</sup>Inicialmente, seguindo a versão intuitiva apresentada por Cabral (2017), seção 14.1.

função utilidade, mas por conveniência é mais simples utilizar uma função demanda derivada a partir de uma função utilidade (veja Berry e Haile, 2021, seção 2.7). Por exemplo, uma família grande tende a dar maior valor ao tamanho do carro do que uma família menor. Uma análise formal (teórica) simplificada que permite essa heterogeneidade pode ser a seguinte.

Considere a utilidade de um consumidor do tipo  $i$  pelo carro  $j$  ser dada por:

$$U_{ij} = x_{j1}\beta_{i1} + \dots + x_{jK}\beta_{iK} - p_j + \epsilon_{ij}, \quad (7.1)$$

tal que  $\beta_{ik}$  é o valor dado pelo consumidor para a característica  $x_{.k}$ ,  $k = 1, \dots, K$ , e  $x_{jk}$  é o quanto da característica  $k$  o produto  $j$  possui e  $p_j$  é o preço do produto  $j$ . Por fim,  $\epsilon_{ij}$  é mede as preferências consumidor/modelo específica que não são capturadas pela pelas características dos produtos.

Suponha que cada consumidor compra no máximo um carro, i.e. existe uma escolha discreta de compra do carro: 0 ou 1. A opção de não compra é chamada de opção externa (*outside good*). A opção de não comprar não um carro resulta na utilidade  $U_{i0} = 0$ . Dada a função utilidade (7.1), a “probabilidade” de que um consumidor do tipo  $i$  escolha comprar o carro  $j$  é dada por:

$$Prob(U_{ij} > U_{il}) \text{ para qualquer } l \neq j, \quad (7.2)$$

tal que  $Prob$  é a notação de probabilidade.

Estimar a distribuição de probabilidade destes coeficientes pode ser difícil. Se o pesquisador tiver acesso a dados de compras individuais, a tarefa é bem mais simples. Podemos estimar o quanto cada característica demográfica influencia a escolha do consumidor por cada veículo. Por exemplo, a relação tamanho da família e tamanho do carro. Finalmente, a distribuição de características dos consumidores leva a distribuição da demanda do comprador pelas características dos produtos.

A tarefa de estimar esta relação é complicada quando apenas dados agregados por produto são disponíveis. O metodologia denominada de BLP, de Steve Berry, Jim Levinsohn e Ariel Pakes (1995), foi desenhado para estimar a função de demanda quando se conhece a característica de cada produto, o seu market share e se possui características demográficas da população. Por exemplo, se a média do tamanho das famílias no Distrito Federal é maior do que no Espírito Santo, e se – tudo o mais constante – carros maiores são mais demandados no DF, podemos inferir que, na média, famílias maiores preferem carros maiores.

Na Tabela 7.2 é apresentado uma parte dos resultados da estimação do BLP para automóveis nos EUA (Berry, Levinsohn e Pakes, 1995). Na segunda e terceira e coluna estão os valores mínimos e máximos de cada característica, respectivamente, para os 997 modelos considerados. Por exemplo, o carro menos potente na amostra é Plymouth Gran Fury 1985, com 0.170 HP/Peso, enquanto que o mais poderoso é o Porsche 911 Turbo 1989 com 0.948 HP/Peso.

Na quarta e quinta coluna da Tabela 7.2 estão as estimativas da utilidade média e do desvio-padrão dos coeficientes da utilidade média. Na média, os consumidores americanos não colocam muito peso na escolha de potência ou consumo. Por outro lado, um consumidor cuja valoração de HP é um desvio-padrão acima da média, i.e. um consumidor com  $\beta_{jh} = 4.628$  ( $h = \text{HP}$ ), tem a disposição à pagar de  $(0.948 - 0.170) \times 4.628 = \$3600$  mais por um Porsche ( $x_{jh} = .948$ ) do que por um Plymouth ( $x_{jh} = .170$ ), considerando apenas o HP/peso.

Tabela 7.2: Características dos Produtos Seleccionados e Valoração do Consumidor, Modelo BLP

Característica	Min $x_{jk}$	Max $x_{jk}$	Média $\beta_{jk}$	Desvio-padrão $\beta_{jk}$
HP/Peso	.170	.948	0	4.628
Ar Cond.	0	1	1.521	0.619
Milhas / \$	8.46	64.37	0	1.050
Tamanho	.756	1.888	3.460	2.056

Fonte: Berry, Levinsohn e Pakes (1995), Tabela II e Tabela IV.

Se o valor médio de  $\beta_h$  é próximo de zero e o desvio-padrão é positivo, então para alguns consumidores um valor elevado de HP/peso é utilidade-decrescente. Geralmente as características dos produtos são consideradas como bons atributos, mesmo que em alguns casos a percepção de valor seja bem diferente entre consumidores.

Estes resultados também nos permitem comparar a importância relativa de cada característica. Por exemplo, possuir ar-condicionado como uma característica básica do veículo vale tanto quanto adicionalmente 0.44 ( $= 1.521/3.460$ ) unidades de tamanho.

Tendo feito a estimativa dos parâmetros da função utilidade, podemos calcular a demanda do consumidor. Especificamente, vamos considerar a demanda de um consumidor que valoriza as características na média (da Tabela 7.2). Suponha que este consumidor compara Nissan Sentra com o Ford Escort. Nenhum dos dois possui ar condicionado no modelo básico. Em relação a tamanho, Sentra possui 1.092 e o Escort 1.116. Neste exemplo, a utilidade do Sentra é  $1.092 \times 3.46 = 3.778$ , enquanto que a do Escort é dado por  $1.116 \times 3.46 = 3.861$ .

Suponha que o valor de  $\epsilon_{ij}$  em (7.1) é distribuído como um valor extremo tipo I, ou seja, como um duplo exponencial:  $-\exp(\exp(\epsilon_{ij}/\sigma))$ . A distribuição valor extremo tipo I é parecida com uma normal, mas com caudas mais grossas. Também suponha por simplicidade que o Sentra e o Escort sejam as únicas opções disponíveis para os consumidores. Então o market share do Sentra entre os consumidores com *valoração média* é dado por:

$$s_{Sentra} = \frac{\exp(3.378/\sigma)}{\exp(3.378/\sigma) + \exp(3.861/\sigma)}. \quad (7.3)$$

Se somarmos o market share do Sentra para todos os demais tipos de consumidores,



ponderado pelo tamanho de cada grupo no total, podemos computar o market share total do Sentra.

No processo de estimar a distribuição de  $\beta_k$ , podemos estimar também a variância do termo de erro  $\epsilon_{ij}$ , que permite calcular o erro-padrão  $\sigma$ . Relembre que  $\epsilon_{ij}$  mede a utilidade do consumidor que não é explicada pelas características do produto que foram incluídas no modelo. Na equação do market share (7.3), a medida que  $\sigma \rightarrow \infty$ , a participação de mercado do Sentra seria  $s \rightarrow \frac{1}{2}$ . Este seria o caso em que as características do produto não são úteis para explicar a participação de mercado do Sentra. No extremo oposto, quando  $s \rightarrow 0$ , mesmo uma pequena diferença na característica do produto desloca a demanda de um produto para outro.

Preço também é uma característica do carro, e é a característica que os consumidores mais se importam. Por simplicidade assumimos que o parâmetro do preço seria -1 na função utilidade, mas dada o coeficiente estimado é possível calcular uma matriz de elasticidade para os produtos. A principal característica do BLP é ser adequado para estimar elasticidades preço própria e cruzada. É a abordagem de maior sucesso para estimar a demanda para produtos diferenciados.

Para mostrar a capacidade do modelo em estimar elasticidades cruzada, Cabral (2017, p. 354) fez uma seleção destas elasticidades, que estão na Tabela 7.3. Sem surpresa, os resultados indicam que a elasticidade-cruzada entre Sentra e Escort é maior do que a elasticidade cruzada entre Sentra e Lexus. Isto é, Sentra e Escort são substitutos e um consumidor migra para um destes devido a mudanças nos preços. Também é possível observar que a elasticidade-própria dos veículos de luxo, BMW e Lexus, é menor do que Sentra e Escort que são carros menos dispendiosos. A visão geral, é de que o estimador BLP consegue mostrar os submercados relevantes pela correta estimação da função demanda.

Tabela 7.3: Elasticidades Preço da Demanda por Modelos Seleccionados, BLP

Característica	Nissan Sentra	Ford Escort	Toyota Lexus	BMW 735i
Nissan Sentra	-6.5282	.4544	.0008	.0000
Ford Escort	.0778	-6.0309	.0008	.0000
Toyota Lexus	.0002	.0010	-3.0847	.0322
BMW 735i	.0001	.0005	0.0926	-3.5151

A quantidade muda na linha e o preço na coluna. A amostra inclui 77 carros sub-compacto e compacto, bem como 24 modelos de luxo. Fonte: Berry, Levinsohn e Pakes (1995), Tabela VI.

Aqui é útil a definição de características verticais e horizontais. Característica vertical diz respeito quando os consumidores concordam na preferência por um tipo de característica, por exemplo, os consumidores sempre preferem pagar mais por um carro com ar condicionado. Conforto é uma característica que todos os consumidores preferem, embora seja difícil mensurar. Característica horizontal é quando alguns consumidores

preferem certa característica e isto pode incluir até um sinal diferente. Isto pode significar que alguns consumidores avaliam positivamente uma característica enquanto outros não. Um exemplo é HP/peso: alguns consumidores avaliam positivamente enquanto outros não. Na seção 7.4, veremos mais detalhes do modelo BLP.

## 7.2 Sistema de Demanda Multi-Produto: Espaço do Produto

Uma questão básica aqui é escolha entre modelar assumindo hipótese de agente representativo ou heterogêneo.

### 7.2.1 Modelo AIDS

O modelo AIDS (Almost Ideal Demand System) de Deaton e Mullbauer (1980) é baseado na teoria microeconômica do consumidor. O ponto de partida é que temos que lidar com múltiplos produtos no espaço das quantidades:

$$\begin{aligned} \ln q_1 &= \beta_1 p_1 + \alpha_{12} p_2 + \dots + \alpha_{1J} p_J + \gamma_1 y + e_1 \\ &\vdots \\ \ln q_J &= \beta_J p_J + \alpha_{J2} p_1 + \dots + \alpha_{JJ-1} p_{J-1} + \gamma_J y + e_J \end{aligned} \quad (7.4)$$

Como existem mais parâmetros do que equações para estimar, para ser aplicado o modelo precisa de diversas restrições. O ponto inicial do modelo é a função dispêndio  $e(p_t, U_t)$ . O logaritmo da função dispêndio, escrevendo as funções sem indicadores das observações para simplificar, é

$$\ln(e(p, U)) = (1 - U) \log(a(p)) + U \ln(b(p)). \quad (7.5)$$

$U$  é escrito a partir de função de utilidade indireta, tendo renda e preços como argumentos. As funções  $a(p)$  e  $b(p)$  são referidas como índices de preços pois são construídas a partir dos preços (formas polares de Gorman, 1961). As funções são:

$$\ln(a(p)) = \alpha_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \ln(p_k) \ln(p_j), \quad (7.6)$$

e

$$\ln(b(p)) = \ln(a(p)) + \beta_0 \prod_k p_k^{\beta_k}. \quad (7.7)$$

Usando o lema de Sheppard<sup>3</sup> podemos obter as participações de dispêndio dos grupos ( $g$ ) de produtos como:

$$w_g = \frac{\partial \ln e(p, U)}{\partial \ln p_g} = \alpha_g + \sum_j \gamma_{gj} \ln(p_j) + \beta_g \ln \left( \frac{x}{P} \right), \quad (7.8)$$

<sup>3</sup>O lema de Sheppard é uma forma fácil de se obter uma função de demanda Hicksiana a partir das funções dispêndio.

tal que  $x$  é o dispêndio total sobre o grupo,  $\gamma_{gj} = \frac{1}{2}(\gamma_{gj}^* + \gamma_{jg}^*)$ ,  $P$  é um índice de preços para o grupo.

O índice de preços  $P$  deflaciona a renda e em termos práticos pode ser complicado lidar com esta variável. A especificação apropriada do índice de preços é a seguinte:

$$\ln P = \alpha_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_j \sum_k \ln(p_k) \ln(p_j). \quad (7.9)$$

Em termos práticos, este índice de preços é substituído por outro denominado de índice de preços de Stone (Stone, 1954). A vantagem deste índice é que ele não depende dos parâmetros do modelo.

$$\ln P = \sum_k w_k \ln p_k. \quad (7.10)$$

Com o índice de Stone as participações de dispêndio estimadas são lineares nos parâmetros  $(\alpha_g, \gamma_{g1}, \dots, \gamma_{gJ}, \beta_g)$ . Modelos que são lineares nos parâmetros são mais simples de serem estimados. Além disso, o índice de Stone não depende de todos os parâmetros e preços do modelo, fazendo com que não seja preciso estimar o modelo inteiro.

Na prática o modelo AIDS é estimado seguindo a seguinte receita:<sup>4</sup>

1. Calcule  $w_g$ , i.e. a participação de dispêndio do bem  $g$ , usando o preço de  $g$ ,  $p_g$ , a quantidade demanda de  $g$ ,  $q_g$ . O dispêndio total é definido como  $x = \sum_j p_k q_k$ .
2. Calcule o índice de preços de Stone:  $\ln P = \sum_k w_k \ln(p_k)$
3. Rode a seguinte regressão linear (geralmente com variáveis instrumentais):

$$w_g = \alpha_g + \sum_k \gamma_{jk} \ln(p_k) + \beta_g \ln\left(\frac{x}{P}\right) + \varepsilon_g,$$

tal que  $\varepsilon_g$  é o termo de erro.

4. Recupere os  $J + 2$  parâmetros  $(\alpha_g, \gamma_{g1}, \dots, \gamma_{gJ}, \beta_g)$ .

As elasticidades preço-cruzada da demanda podem ser recuperadas a partir dos parâmetros estimados pelo modelo AIDS. Primeiro cabe observar que o dispêndio se relaciona com preços, quantidades e renda:

$$\ln w_j = \ln p_j + \ln q_j - \ln y \Leftrightarrow \ln q_j = \ln w_j - \ln p_j + \ln y.$$

Portanto, a elasticidade da demanda é calculada como:

$$\eta_{jk} \begin{cases} \frac{\partial q_j}{\partial p_k} = \frac{\partial w_j}{\partial p_k} - 1, & \text{se } j = k, \\ \frac{\partial q_j}{\partial p_k} = \frac{\partial w_j}{\partial p_k}, & \text{se } j \neq k. \end{cases} \quad (7.11)$$

<sup>4</sup>Estes passos são comuns na literatura sobre modelos AIDS, mas aqui estou seguindo Davis e Garcés (2010), cap. 9.

Diferenciando a equação de despesa em relação ao preço, se tem:

$$\frac{\partial q_j}{\partial p_k} = \frac{\gamma_{jk} - w_k \beta_j}{w_j}, \quad (7.12)$$

substituindo na equação (7.11), observamos que elasticidades dependem apenas dos parâmetros estimados no modelo AIDS:

$$\eta_{jk} \begin{cases} \frac{\gamma_{jk}}{w_k} - \beta_j - 1 & \text{se } j = k, \\ \frac{\gamma_{jk}}{w_j} - \frac{w_k}{w_j} \beta_j, & \text{se } j \neq k. \end{cases} \quad (7.13)$$

### Restrições nos Parâmetros

No sistema de demanda AIDS podemos impor algumas restrições para reduzir o número de parâmetros a serem estimados. Não é necessário impor todas as restrições, geralmente o uso depende do tamanho do sistema.

Uma restrição utilizada é a de aditividade, i.e. o requerimento que o sistema de demanda deve satisfazer a restrição orçamentária. Isto implica que a soma das participações do produto deve ser igual a 1

$$\sum_{j=1}^J w_j(p, y) = 1.$$

Neste caso as seguintes restrições podem ser aplicadas:

$$\sum_{j=1}^J \alpha_j = 1, \quad \sum_{j=1}^J \gamma_{jk} = 0, \quad \sum_{j=1}^J \beta_j = 0. \quad (7.14)$$

O uso destas restrições requer que o termo de erro do sistema some zero. Isto significa que a matriz de variância-covariância ( $Var$ ) será singular. Este fato não torna possível a estimação de todas as equações em conjunto. Portanto, uma equação deverá ser excluída. Por exemplo, na estimação de quatro produtos, um deverá ser quando excluído quando se estima o modelo restrito (com as restrições impostas).

Outro conjunto de restrições que pode ser utilizado vem de quando se assume homogeneidade na equação de participação na despesa ( $w$ ). Um exemplo comum é utilizar uma função homogênea de grau zero. Por sua vez, sabemos que a soma funções homogêneas de grau zero é homogênea de grau zero.

A homogeneidade de grau zero implica que

$$w_j(\lambda p_1, \dots, \lambda p_J, \lambda y) = w_j(p_1, \dots, p_J, y) \text{ para } \lambda > 0.$$

Isto implica que a inclusão de um termo comum em todas os elementos do problema não afeta a escolha dos consumidores. Neste caso as seguintes restrições podem ser aplicadas:

$$\sum_{j=1}^J \alpha_j = 1, \quad \sum_{j=1}^J \gamma_{jk} = 0, \quad \sum_{k=1}^J \gamma_{jk} = 0. \quad (7.15)$$

Neste conjunto de restrições a soma em relação à  $j$  significa restringir entre equações e a soma em relação a  $k$  implica em restrição na mesma equação, mas entre variáveis.

Outra restrição muito utilizada é a simetria. Esta restrição implica que o parâmetro entre produtos em um mercado é simétrico a outro mercado. Por exemplo, o parâmetro que forma a elasticidade cruzada da Pepsi na equação da Coca-Cola deve ser igual ao parâmetro da Coca-Cola na equação da Pepsi. Observe que os parâmetros serem iguais não implica em elasticidades-cruzada iguais, pois esta ainda depende dos market shares dois dois produtos. Portanto, a restrição de simetria é:

$$\gamma_{jk} = \gamma_{kj}.$$

### 7.2.2 Exemplo Mercado de Cerveja

Uma aplicação conhecido do sistema de demanda AIDS para produtos diferenciados é o trabalho sobre demanda por cerveja de Hausman, Leonard, e Zona (1994). Neste trabalho eles estimam um sistema de demanda para poder simular fusão de empresas de cerveja que operam no mercado americano. Esse artigo também é muito conhecido pelos instrumentos que eles usam.

Nesta abordagem os autores utilizam um modelo de três níveis. O nível mais alto (mais agregado) trata da modelagem da demanda pelo produto cerveja. No nível intermediário eles identificam a demanda pelos grupos de cerveja, enquanto que no último estágio é estimada a demanda por cada marca/tipo de cerveja.

O nível mais baixo utiliza uma especificação similar a equação de participação (7.8):

$$w_{itn} = \alpha_{in} + \sum_j \gamma_{ij} \ln(p_{jnt}) + \beta_i \ln\left(\frac{y_{Gnt}}{P_{nt}}\right) + \varepsilon_{itn}, \quad (7.16)$$

tal que  $i$  é a marca de cada cerveja,  $n$  o mercado local (cidade) e  $t$  o tempo.  $y_{Gnt}$  é o dispêndio no segmento de cerveja  $G$ . No segundo nível, estimação do segmento, os autores utilizam a especificação ln-ln:

$$\ln q_{mnt} = \beta_m \ln y_{Bnt} + \sum_k \delta_k \ln \pi_{knt} + \alpha_{mn} + \varepsilon_{mnt}. \quad (7.17)$$

Neste equação  $q_{mnt}$  representa o volume comercializado no segmento  $G$ ,  $y_{Bnt}$  é o gasto total com cerveja,  $\pi$  é o índice de preços de cerveja em cada segmento e  $\alpha$  é uma constante. No nível mais agregado a equação é similar ao intermediário. Novamente utilizando uma formulação ln-ln se tem:

$$\ln u_t = \beta_0 + \beta_1 \ln y_t + \beta_2 \ln \Pi_t + Z_t \delta + \varepsilon_t, \quad (7.18)$$

tal que  $u$  é o dispêndio total em cerveja,  $y_t$  é a renda disponível,  $\Pi_t$  é o índice de preços (geral) para cerveja, por fim,  $Z$  são as variáveis demográficas, fatores de deslocamento mensais, bem como requerimento de idade.

A identificação dos coeficientes de preço segue a seguinte estratégia. Mesmo em estruturas mais complexas a endogenia do preço permanece e o problema de escolha de instrumentos permanece. Os autores propõem utilizar o preço de uma cidade como instrumento para o preço em outra cidade. Isto funciona sob a hipótese de que a regra de preço se parece com:

$$\ln p_{jnt} = \delta_j \ln c_{jt} + \alpha_{jn} + \omega_{jnt}. \quad (7.19)$$

Nesta equação  $p_{jnt}$  é o preço do produto  $j$  na cidade  $n$  no período  $t$ ,  $c_{jt}$  representa o custo nacional do produto  $j$ ,  $\alpha_{jn}$  é o deslocador do preço específico a uma localidade  $n$  em  $t$ . Esse deslocador por representar custos locais de transporte ou diferencial de salário, por exemplo. Por fim,  $\omega_{jnt}$  é distúrbio estocástico com média zero.

Uma hipótese importante de Hausman, Leonard e Zona (1994) é a de que os choques de demanda locais  $\omega_{jnt}$  não são correlacionados. Isto permite utilizar os preços de outros mercados para para o mesmo produto (no mesmo  $t$ ) como instrumentos, se for possível obter efeito fixo por mercado). Este tipo de instrumento é comumente chamado de *instrumentos de Hausman*. Uma das críticas a este tipo de instrumento é que ele pode não cumprir com o seu papel se existir uma campanha de marketing nacional. Além dos instrumentos de Hausman, diversos pesquisadores também usam preços dos fatores, tais como salários, preço do malte ou açúcar como variáveis que deslocam o custo marginal.

### 7.2.3 Problemas com Sistemas Multi-Produto no Espaço dos Produtos

Estes são problemas gerais que afetam este tipo de arcabouço e não apenas o modelo AIDS e suas variações. O primeiro problema é ter muitos parâmetros para serem estimados:  $J$  bens implicam na ordem de  $J^2$  parâmetros. Além disso, o agrupamento de produtos depende do tópico e interesse em estudo.

Quando se utiliza o modelo AIDS para produto diferenciado estamos tipicamente utilizando dados de produto ( $j$ ) e temos preços e quantidades,  $\{q_j, p_j\}$ , algumas características dos produtos e alguns atributos dos consumidores ( $z$ ).

A teoria entrega:

$$q_j = f(p_j, y), \quad (7.20)$$

para um dado mercado. Mesmo um sistema log-linear contém parâmetros demais para se estimar. Para conseguir estimações é preciso agregar sobre produtos. As possibilidades incluem:

- Orçamento multi-nível e formas polar de Gorman (1961). Ideia básica de utilidade em árvore, como no exemplo do mercado de cerveja:
  - Primeiro aloca gastos para grupos;
  - Aloque os gastos dentro dos grupos;
  - $J$  bens em  $K$  grupos fornece  $J^2 + J$  versus  $J^2/K + K^2$  parâmetros a serem estimados.

- Preferências Dixit-Stiglitz (1977). Usualmente é algo como  $(\sum x_i^\rho)^{1/\rho}$ . Isto tem sido utilizado extensivamente para a análise de benefícios de variedade. Esse modelo impõe hipóteses de formas funcionais fortes para ter uma abordagem simples para questões que tratam do número de produtos comercializados. Além disso, a forma funcional elimina o fato de que alguns bens são substitutos mais próximos do que outros bens. Neste sentido esta abordagem elimina a possibilidade de analisar padrões de substituição e mesmo análises dinâmicas de comportamento de mercado.

Outro problema importante aqui é a questão da introdução de bens novos. Como o modelo é baseado em quantidades e preços de produtos comercializados, é difícil avaliar adequadamente a introdução de novos bens.

O melhor sistema para um dado problema depende das questões que se pretende analisar e o tipo de dado que é disponível. Um pequeno número de bens, muitos dados e não existir maior característica do produto que dirige a escolha do consumidor, é uma situação adequada para estimativa no espaço dos produtos.

Uma forma funcional suficientemente flexível (veja abaixo a seção 7.3) e um tratamento adequado da questão da simultaneidade, deve gerar elasticidades razoáveis em pontos próximos do centro dos dados – na maioria dos casos é disto que precisamos.

Se estamos interessados em medidas de bem estar total, precisamos mais do que boas aproximações próximas do centro dos dados, e as hipóteses são muito importantes.

### 7.3 Agentes Heterogêneos: Agregação e Simulação em Estimativa

Digressão sobre simulação em estimativa. Método introduzido em estimativa por Pakes (1986), mas usado por MacFadden e outros para fazer previsões. Tipicamente apenas possuímos demanda agregada (por mercado), preços e características. Para fazer o *match* entre teoria e dados para valores de parâmetros diferentes em um modelo com agentes heterogêneos, vamos precisar de

$$q_t(p_t; \theta, f_t) \equiv \int D(p_t, z_i; \theta) f_t(z) dz, \quad (7.21)$$

que não é uma integral analítica. Nós aproximamos esta integral por simulação. No fundo, a integral aqui diz que podemos agregar utilizando informações sobre  $z_i$ .

Nós conhecemos  $f_t(\cdot)$  ou temos uma estimativa muito boa a partir de domicílios, como PNAD ou CPS para os EUA, ou mesmo de outra fonte pública de dados. Uma forma de lidar com isto é tomar diversas amostragens (*draws*) de  $z_t$ , de tamanho  $ns$ , e para cada uma calcule  $D(p_t, z_i; \theta)$ .

A expectativa de cada saque aleatório da base de dados é  $q_t(p_t; \theta, f_t)$ , e a variância de

$$\frac{1}{ns} \sum_i D(p_t, z_i; \theta)$$

em torno de sua média com respeito a distribuição  $f_t(\cdot)$  é proporcional a  $1/ns$ . Então, a medida que  $ns$  aumenta nós temos um estimador cada vez mais preciso. Portanto, o procedimento de saques aleatórios deve ser grande para garantir precisão.

Alternativamente, quando se tem estrutura de painel, se fixamos  $ns$  temos um estimador mais preciso de

$$\frac{1}{T} \frac{1}{ns} \sum_i D(p_t, z_i; \theta)$$

pelo aumento de  $T$ . Esta é uma forma de usar computadores via simulação para entender o comportamento dos consumidores.

## 7.4 Demanda por Produtos Diferenciados II<sup>5</sup>

A demanda por produtos diferenciados é um sistema de demanda, mas neste setup o produto é apenas uma cesta de características. Se especificamos completamente as características de todos os produtos, podemos determinar a utilidade de cada indivíduo e, conseqüentemente, a demanda.

### 7.4.1 História

O fundamento desta teoria da demanda é devido a Lancaster (1966), enquanto que o avanço econométrico a McFadden. Em organização industrial, a literatura focou mais sobre localização de produtos e apenas recentemente ocorreu o retorno ao estudo da natureza da demanda, e o seu impacto sobre precificação condicional à “localização do produto.” Entre os trabalhos sobre localização dos produtos podemos destacar: modelo da cidade linear de Hotelling, cidade circular de Salop e os modelos de características horizontais ou verticais de Mussa-Rosen, Gabszwwicz-Thisse, Shaked-Sutton. Por fim, a abordagem de Anderson, De Palma, e Thisse resume o uso de modelo logit em organização industrial.

A visão moderna é a seguinte. Transfira todas as abordagens anteriores para um modelo teórico de demanda de mercado que possa ser utilizado empiricamente e forneça as estratégias necessárias para estimação (Lancaster/McFadden). Esta abordagem é facilitada devido a grande capacidade dos computadores. Dois dos principais expoentes da abordagem moderna são:

- O modelo de Bresnahan (1987); que estima um modelo vertical com uma única dimensão;

<sup>5</sup>Baseado em notas de aula de Ariel Pakes.



- BLP; que resgata a questão de simultaneidade e fornece uma solução, estendendo o modelo logit para permitir características horizontais e verticais. Este framework é uma linha de pesquisa ativa gerando muita contribuição para o entendimento do mercado.

Assim como todo método, a abordagem no espaço das características tem os seus problemas. Não é sempre que se tem bons dados sobre as características e é difícil lidar com o caso de muitas características. Este é um problema recorrente em bens de consumo. Para o modelo estimar adequadamente os parâmetros é necessário definir pelo menos uma característica não-observada. Ao mesmo tempo que a característica não-observada é solução ela se torna um problema devido a complexidade computacional.

Introdução de novos bens e características totalmente novas não são bem capturadas no ambiente do espaço das características. Novas múltiplas características também são difíceis de serem captadas nesta abordagem. Um exemplo clássico é a introdução do laptop. Os primeiros laptops possuíam tela menor, menor capacidade de processamento e memória do que computadores desktop. Como se introduz um bem novo mais caro e com características piores do que os produtos comercializados? Também existem problemas técnicos em expandir o modelo para múltiplas escolhas e produtos complementares.

### 7.4.2 Arcabouço Básico: Escolha do Consumidor

Nesta seção descrevemos o modelo básico no espaço das características. Estes componentes teóricos são utilizados em diversas estruturas.

Aqui os produtos são cestas de características<sup>6</sup> e as preferências são definidas sobre estas características. Estes são os primeiros pressupostos teóricos de um modelo no espaço das características.

Além disso, cada consumidor escolhe uma cesta que maximiza a sua utilidade. Os consumidores possuem preferências relativas diferentes (usualmente apenas preferências marginais) para características diferentes. Isto implica que escolhas diferentes são feitas por consumidores diferentes. Também assumimos que se consome uma unidade de um bem ou serviço (embora esta hipótese possa ser relaxada depois).

Aqui a demanda agregada é a soma individual sobre todas as demanda individuais. O indivíduo pode ser um representante por tipo de consumidor. Portanto, a demanda depende totalmente da distribuição da preferência dos consumidores.

Formalmente, a utilidade de um indivíduo  $i$  para o produto  $j$  é dado por

$$U_{ij} = U(x_j, p_j, \nu_i; \theta), \quad (7.22)$$

para  $j = 0, 1, 2, \dots, J$ .  $J$  é o número de bens competindo em uma indústria. Aqui  $x_j$  representa o conjunto de características do produto  $j$ ,  $p_j$  o preço do mesmo produto. O preço pode ser interpretado como uma característica do bem, e é a característica

<sup>6</sup>Inspiração em Lancaster (1966).

que geralmente os consumidores mais se importam. Aqui também assumimos que as características não variam com os consumidores.  $\nu_i$  representa os fatores que geram diferentes preferências e, por sua vez,  $\theta$  determina o impacto destas preferências na utilidade. Observe que  $\nu_i$  varia entre os indivíduos  $i$ .

A escolha do individuo é discreta, sendo feita entre as possibilidades  $J$ . Um modelo de escolha discreta extremo poderia ser de compra ou não-compra, neste casos as possibilidades seriam  $\{0, 1\}$ . Importante deixar claro que as abordagens empíricas são casos de modelos de escolha discreta.

**Não-Compra.** O produto  $j = 0$  é tipicamente a opção de não-compra, chamada também de *outside good*. A não-compra é a procura pelo bem que não pode competir com os bens ofertados pela indústria, e, portanto, possuem preços e quantidades que são completamente exógenos. Se não existir o outside good, então não podemos usar o modelo para estudar a demanda agregada, pois estaríamos assumindo que todos os indivíduos deveriam comprar bens comercializados no mercado. O sistema de demanda para  $J$  bens funciona como se fosse um sistema em que se escolhe um bem de um grupo superior, condicional a compra de algum bem. Nesta estrutura assumimos que existe um bem abaixo dos bens que devem ser escolhidos.

**Escolha.** Consumidores do tipo  $\nu$  fazem escolhas exclusivas por produtos valorando a cesta de características. Isto significa que eles comprem apenas um bem. Os indivíduos que comprem o bem  $j$  formam um subconjunto  $A_j(\theta)$  que é:

$$A_j(\theta) = \{\nu : U_{i,j} > U_{i,k}, \forall k\}. \quad (7.23)$$

Observe que não existe “empate” quando se compara as utilidades de escolha entre o bem  $j$  e qualquer outro. Fazendo com que  $f(\nu)$  seja a distribuição de preferências na população de interesse, as probabilidades de escolha, que também são as previsões do modelo para *market shares* (uma vez que  $f(\nu)$  for especificada corretamente), e são dadas por

$$s_j(x, p; \theta) = \text{Prob}\{\nu \in A_j(\theta)\}, \quad (7.24)$$

tal que  $(x, p)$  sem subscritos significam os vetores contendo todas as características de todos os produtos  $J$  e não apenas do produto  $j$ . Esta especificação decorre do fato de que o consumidor leva em conta todas as características dos bens para fazer a sua escolha (inclusive o preço).

Por fim, a demanda total será

$$Ms_j(x, p; \theta)$$

sendo que  $M$  é o número total de indivíduos ou domicílios.

A escolha de cada consumidor é invariante à adição de um número a utilidade individual. O que importa no processo de escolha é a comparação entre os benefícios com o consumo. Isto implica que algumas normalizações precisam ser feitas. Esta normalização é a

determinação de uma constante, que naturalmente afeta a interpretação do modelo. A normalização usualmente feita na literatura é determinar o valor da não-compra ser zero. Isto equivale a subtrair  $U_{i,0}$  de todas as outras utilidades dos demais bens (inside goods)  $U_{i,j}$ .

**Exemplos.** Algumas formas funcionais podem ser utilizadas para estimar a demanda. Isto resume grande parte das aplicações com a estimação de demanda no espaço das características. Os modelos são: horizontal puro, vertical puro, logit, e BLP. BLP é uma generalização destas abordagens.

No modelo horizontal puro, chamado também de modelo de Hotelling, a característica do produto é a sua localização e a característica do consumidor é a sua localização. Suponha que existe um custo de transporte para o consumidor adquirir a mercado, dado que ela está em outra localidade. Neste caso a função utilidade é

$$U_{i,j} = \bar{u} + (y_i - p_j) + \theta(\delta_j - \nu_i)^2. \quad (7.25)$$

Versões deste modelo tem sido extensivamente utilizada em modelos teóricos para ganhar intuição sobre problemas de colocação dos produtos no mercado. Observe que na função utilidade (7.25) existe apenas uma característica com um coeficiente aleatório ( $\delta_j$ ): localização. Mesmo em modelos com bens relativamente homogêneos com custos de transporte você pode querer que pessoas diferentes coloquem valor em moeda distintos para os bens (heterogeneidade).

No modelo vertical puro existe apenas uma característica: a qualidade.<sup>7</sup> Nesta estrutura, todos os consumidores concordam em ordenar os produtos por qualidade e desejam mais qualidade. Os consumidores não gostam do preço – desutilidade marginal em pagar. Uma função utilidade neste setup pode uma em que diferentes níveis de renda interagem com o preço

$$U_{i,j} = \bar{u} - \nu_i p_j + \delta_j, \text{ para } \nu_i > 0. \quad (7.26)$$

Esta talvez seja a outra forma funcional mais usada para tratar da questão de colocação do produto no mercado. Bresnahan usou esta forma funcional em trabalhos empíricos. Apesar dela ir na direção correta, ele encontrou muitos problemas de aplicação por causa da simplicidade da especificação.

O modelo de escolha logit prove uma forma direta de estimação da escolha. Uma possível forma funcional para o logit é a qualidade média mais uma preferência idiossincrática por um produto:

$$U_{i,j} = \delta_j + \epsilon_{i,j}. \quad (7.27)$$

se o termo idiossincrático  $\epsilon_{i,j}$  é distribuído i.i.d. na forma do duplo exponencial  $F(\epsilon) = \exp[-\exp(-\epsilon)]$ , então temos uma forma fechada para o máximo e a integral do máximo.

<sup>7</sup>Mussa-Rosen, Gabszewicz-Thisse, Shaked-Sutton, Bresnahan.

Por esta razão o modelo logit tem sido largamente utilizado no trabalho empírico. Observe que implicitamente existe a hipótese distribucional das preferências, e como mostraremos abaixo ela não é uma hipótese muito atrativa.

Por fim, temos a generalização do BLP. A forma funcional é

$$U_{i,j} = f(\nu_i, p_j) + \sum_k x_{j,k} \nu_{r,k} \theta_{k,r} + \epsilon_{i,j} \quad (7.28)$$

tal que  $\epsilon_{i,j}$  é distribuído como no logit. A seguir vamos ver estes modelos em maior detalhes. Em todos estes modelos a estrutura de dados é a mesma. Partimos de informações agregadas de venda por produto  $j$ , preços, e características individuais dos produtos:

$$(s_j^o, p_j, x_j)$$

com  $s_j^o$  sendo o market share observado do produto  $j$ .

### 7.4.3 Demanda Logit

O modelo de regressão logit é uma abordagem de escolha discreta popularmente utilizada para estimar a demanda. Ele pode ser aplicado para escolhas binárias ou multinomiais.

Em um modelo de escolha discreta o interesse principal é na *probabilidade de escolha*. Por exemplo, considere um modelo de escolha binária, tal que  $y$  pode assumir dois valores:  $\{0, 1\}$ . Então a probabilidade condicional a  $x$  é:  $\text{Prob}(y = 1 | x)$ . No caso do modelo multinomial escrevemos a probabilidade de escolha como  $\text{Prob}(y = j | x)$ . Como no logit se modela a probabilidade de escolha, o resultado estimado estará entre zero e um.

Vamos indexar os consumidores como  $i$  e os produtos/escolhas como  $j$ . Suponha que cada opção é descrita por um conjunto de características  $X_j$ . A utilidade individual é dada por uma parte fixa (linear nas características) e uma parte aleatória ou não-observada:

$$U_{ij} = X_j' \beta - \alpha p_j + \epsilon_{ij}. \quad (7.29)$$

Algumas observações. A linearidade é assumida por conveniência, e é restritiva porque limita os possíveis padrões de substituição que se pode prever. Algumas propostas lidam com o preço de forma diferente, como veremos a seguir. Importante observar que as características não variam por indivíduo.<sup>8</sup>

A probabilidade de escolha logit para o tipo de indivíduo  $i$  escolhendo o produto  $j$  é a expressão de forma-fechada:

$$P_{ij} = \frac{e^{(\beta' x_j - \alpha p_j)}}{\sum_k e^{(\beta' x_k - \alpha p_j)}} \quad (7.30)$$

<sup>8</sup>Isso é diferente de um produto ou serviço que é desenhado especificamente para um tipo de consumidor.

$k$  contabilizando por todos os produtos  $J$  e  $P_{ij}$  é a probabilidade de escolha do produto  $j$  pelo agente  $i$ . Quando se tem apenas um único tipo de consumidor, a probabilidade de compra é igual ao market share. Para mais tipos de consumidores, o market share é igual a soma das probabilidades dos tipos de indivíduos.

A função de escolha logit é derivada a partir da distribuição dos erros como valor extremo tipo I, ou simplesmente o duplo exponencial. A função de distribuição cumulativa de  $\epsilon$  é:

$$F(\epsilon_{ij}) = \exp(-\exp(-\epsilon_{ij})). \quad (7.31)$$

Com explicado anteriormente, quando se soma as escolhas para todos os indivíduos  $i$  se obtém a demanda pelo produto  $j$ . As propriedades das probabilidades logit são:

1.  $0 < P_{ij} < 1$ . A probabilidade de uma alternativa nunca é exatamente 0 e somente será 1 com uma alternativa.
2.  $\sum_{j=1}^J P_{ij} = 1$ .
3. Relação da probabilidade logit com a utilidade representativa é *sigmoid*. Relação em formato de S. Mudanças no regressores tem pouco impacto quando a probabilidade é elevada ou baixa. Os regressores tem impacto quando a probabilidade está mais próxima de 0.5.

Como assumimos na estimação da demanda que a utilidade da não-compra é zero, a fórmula logit pode ser simplificada para:

$$P_{ij} = \frac{e^{(\beta'x_j - \alpha p_j)}}{1 + \sum_{k=1}^J e^{(\beta'x_k - \alpha p_j)}} \quad (7.32)$$

e o *market-share* da não-compra é

$$P_{i0} = \frac{1}{1 + \sum_{k=1}^J e^{(\beta'x_k - \alpha p_j)}}. \quad (7.33)$$

### Elasticidades

Apresentamos aqui como calcular as elasticidades para o modelo logit de forma geral, bem como assumindo linearidade nos regressores. A questão colocada é o quanto muda a probabilidade de compra logit em relação a alteração de algum fator observado na função utilidade. Geralmente queremos observar o comportamento da demanda em relação ao preço, mas pode ser em relação a alguma característica do produto.

**Derivadas.** Primeiro temos que definir as derivadas da função utilidade em relação a mudança no preço/atributo para em seguida definir as elasticidades. Portanto, a mudança na probabilidade de escolha da alternativa  $j$  dado mudança no preço  $p_j$  (mantendo a utilidade representativa das demais alternativas constante) é:

$$\frac{\partial P_{ij}}{\partial p_j} = \alpha P_{ij}(1 - P_{ij}). \quad (7.34)$$

Esta é a fórmula da derivada em relação ao preço (assumindo relação linear entre os regressores).

O quanto muda a probabilidade de compra quando muda um atributo de outra alternativa? Por exemplo, o preço de outro produto aumenta 10%, qual será o impacto sobre o volume de venda do produto  $j$ , para  $j \neq l$ ? Portanto, calculamos a elasticidade cruzada para este fim. A derivada parcial é:

$$\frac{\partial P_{ij}}{\partial p_l} = -\frac{\partial V_{il}}{\partial p_l} P_{ij} P_{il} = -\alpha_l P_{ij} P_{il} \quad (7.35)$$

**Elasticidades.** A elasticidade de  $P_{ij}$  com respeito a  $p_j$  é:

$$\eta_{jp_j} = \frac{\partial P_{ij}}{\partial p_j} \frac{p_j}{P_{ij}} = \alpha p_j (1 - P_{ij}). \quad (7.36)$$

A elasticidade-cruzada ao preço do produto concorrente  $l$  é:

$$\eta_{jp_l} = \frac{\partial P_{ij}}{\partial p_l} \frac{p_l}{P_{ij}} = -\alpha_l p_l (P_{il}). \quad (7.37)$$

Como as probabilidades de compra são iguais aos market shares, podemos fazer  $\sum_i P_{ij} = s_j$ .

**Identificação.** Uma das principais razões para se assumir utilidade linear nas características é a identificação econométrica.<sup>9</sup> Uma forma geral de escrever a utilidade do logit multinomial seria

$$U_{ij} = \delta_j - p_j + \mu + \sigma \epsilon_{ij},$$

Neste caso,  $\delta_j$  é uma representação no espaço do produto e  $\delta_j - p_j$  a utilidade média. Neste caso seria preciso estimar os  $\delta_j$ . Com o total de parâmetros  $\theta = (\delta_1, \dots, \delta_J, \mu, \sigma)$ , temos  $J$  observações independentes por produto para estimar  $J + 2$  parâmetros.<sup>10</sup> Assumindo linearidade nas características os parâmetros a serem estimados são  $\theta = (\beta_1, \dots, \beta_K, \mu, \sigma)$ , tal que  $K$  é o número total de características. O número total de características serão menores do que o número total de produtos, por exemplo, quatro características em comparação a 2217 modelos/ano no BLP.

A distribuição valor extremo Tipo-1 do termo de erro  $\epsilon$  possui média  $\gamma \approx 0.577$  e desvio-padrão de  $\pi^2/6$ . Suponha a seguinte especificação para o modelo logit no espaço das características:

$$U_{ij} = X'_j \beta - p_j \alpha + \mu + \sigma \bar{\epsilon}_{ij}, \quad (7.38)$$

com  $\bar{\epsilon}$  distribuído como valor extremo Tipo-1. Dividindo (7.38) por  $\sigma$  obtemos a mesma utilidade padrão (7.29). Isto significa que podemos apenas identificar ou estimar  $\beta/\sigma$ ,

<sup>9</sup>Veja Moszkowski (2019).

<sup>10</sup>Relembre que  $s_0 = 1 - \sum_{j=1}^J s_j$ .

que não possui interpretação independente.  $\beta^*$  e  $\sigma$  não são separadamente identificados. Usualmente, o modelo é expressado na sua forma re-escalada:  $\beta = (\beta^*/\sigma)$ , sendo  $\beta^*$  o parâmetro corretamente identificado. Portanto a forma logit é a (7.33). Os parâmetros  $\beta$  são estimados, mas importante ter em mente que o coeficiente original,  $\beta^*$ , é dividido pelo parâmetro de escala,  $\sigma$ . Os coeficientes que são estimados indicam os efeitos de cada variável observada relativo à variância dos fatores não-observados. Variância grande nos fatores não observados leva a coeficientes menores, mesmo se os fatores observados possuem o mesmo efeito sobre a utilidade. Isto é,  $\sigma$  maior significa em parâmetro  $\beta$  menor mesmo que o  $\beta^*$  seja o mesmo.

Vale observar que o parâmetro de escala é comum ele não afeta a razão entre dois coeficientes. Por exemplo:

$$\frac{\beta_1}{\beta_2} = \frac{\beta_1^*/\sigma}{\beta_2^*/\sigma} = \frac{\beta_1^*}{\beta_2^*}.$$

Por outro lado, as taxas marginais de substituição podem ser utilizadas normalmente.

#### 7.4.4 Estimação do Logit

Estimação do modelo logit é por máxima verossimilhança (ML), pois as probabilidades logit não são lineares. Tradicionalmente livros introdutórios apresentam estimativas de máxima verossimilhança e mostram que sob as hipóteses clássicas o estimador MQO é o estimador ML, condicional nas variáveis explicativas.

Considere uma amostra aleatória de tamanho  $n$ , para obter o estimador ML condicional precisamos da densidade de  $y$  dado  $X$ , tal que  $X$  é a matriz de variáveis explicativas.

$$f(y | X; \theta) = [\Phi(X; \theta)]^y [1 - \Phi(X; \theta)]^{1-y}, \quad (7.39)$$

para  $y = \{0, 1\}$  e  $\Phi(\cdot)$  a probabilidade logit (7.33). Observe que quando  $y = 1$  a probabilidade de escolha é  $\Phi(X; \theta)$  e quando  $y = 0$  é  $[1 - \Phi(X; \theta)]$ . Quando aplicamos o log natural a esta equação a denominamos de log-verossimilhança:

$$L(\theta) = y \log[\Phi(X; \theta)] + (1 - y) \log[1 - \Phi(X; \theta)]. \quad (7.40)$$

Generalizando para para o caso multinomial a função de log-verossimilhança passa a ser:

$$L(\theta) = \sum_{i=1}^N \sum_{j=1}^J y_{ij} \ln P_{ij}. \quad (7.41)$$

Importante lembrar que geralmente é incluída uma constante comum às alternativas. Nesse sentido o modelo estima corretamente os parâmetros com a exceção das constantes normalizadoras. O modelo é solucionado por máxima-verossimilhança que significa encontrar o máximo da função (7.41):  $\min\{L(\theta)\}$ .

### 7.4.5 Modelo Vertical Puro

O modelo vertical puro foi primeiro estimado por Bresnahan (1987) para a indústria automotiva dos EUA. Ele começa com um puzzle: em 1955 ocorreu uma queda aparente no preço dos automóveis em relação a 1954 e 1956. Além disso, 1955 foi um ano de crescimento econômico elevado (“boom”). A pergunta que Bresnahan faz é se a natureza da concorrência explica este possível paradoxo.

A ideia que Bresnahan deseja testar é a de que o período de boom econômico tornou o cartel insustentável. Existem teorias de comportamento de cartel que mostram que quando se tem um período de crescimento da demanda, os incentivos para romper com o acordo são elevados. Particularmente se o período de demanda elevada for curto, as firmas tem mais incentivos para romper com o acordo de conluio. Neste sentido, o trabalho de Bresnahan tenta entender a natureza da concorrência a partir dos dados da indústria de automóveis. Mais especificamente, ele testa previsões para diferentes modelos de equilíbrio e pergunta qual a capacidade do modelo reproduzir o observado neste mercado. Ele conclui que o modelo de determinação de preços com conluio se aproxima mais dos dados em 1954 e 1955, mas um modelo Bertrand-Nash se adequa melhor aos dados observados em 1955.<sup>11</sup>

O procedimento de estimação que apresentamos é apenas a estimação da equação de demanda. Por modelo vertical entendemos

$$U_{ij} = \delta_j - \nu p_j, \quad (7.42)$$

com  $\delta$  sendo a qualidade dos bens e  $\nu$  sendo diferente entre indivíduos. Observe que todos os consumidores concordam sobre quais produtos tem a maior qualidade e, conseqüentemente, a ordenação dos bens.

#### Estimação

Para obter os *market shares* previstos pelo modelo para diferentes valores de parâmetros fazemos:

Ordene os bens por preço crescente. Também deve ser crescente em  $\delta_j$  se o bem  $j$  se o bem for ter demanda positiva. Um produto com preço mais elevado em relação a qualidade nunca será comprado. Portanto, preço e qualidade devem estar corretamente ordenados.

Começamos a análise com a não-compra. Escolha  $j = 0$  se  $0 > \max_{j \geq 1} (-\nu + \delta_j)$ . Dado o ordenamento dos bens, pode ser mostrado que isto implica em

$$\text{escolha } 0 \Leftrightarrow \nu < \frac{\delta_1}{p_1}.$$

<sup>11</sup>Bresnahan obtém dados de automóveis da *Automotive News and Wards* e utiliza procedimento de estimação similar, mas não idêntico ao subconjunto de procedimentos que vamos discutir aqui.



Assim podemos chegar ao conjunto de consumidores que demandam o *outside good*:<sup>12</sup>  
 $\Rightarrow A_0 = \{\nu : \nu < (\delta_1/p_1)\}$ . Se  $\nu$  segue a distribuição log-normal, ou seja,  $\nu = \exp[\sigma\nu + \mu]$   
 com  $\sigma$  sendo normal padronizada, então escolha 0 se

$$\exp[\sigma\nu + \mu] < \frac{\delta_1}{p_1}, \quad (7.43)$$

isolando  $\nu$  chegamos em

$$\nu \leq \frac{1}{\sigma} \left[ \ln \left( \frac{\delta_1}{p_1} \right) - \mu_1 \right] \Leftrightarrow \nu \leq \psi_0(\theta), \quad (7.44)$$

com  $\psi_0(\theta) \equiv \frac{1}{\sigma} [\ln(\delta_1/p_1) - \mu_1]$ . Isto é, este modelo tem  $s_0 = F(\psi_0(\theta))$ , com  $F(\cdot)$ ,  
 função de distribuição cumulativa, sendo distribuída como normal padronizada e  $s_0$   
 como o *market-share* do *outside good*.

De forma análoga é feita a escolha dos demais bens. Escolha o bem 1 se e somente se

$$0 < \nu p_1 + \delta_1 \text{ e } \nu p_1 + \delta_1 > \nu p_2 + \delta_2.$$

Isto implica em

$$\frac{\delta_2 - \delta_1}{p_2 - p_1} < \nu < \frac{\delta_1}{p_1} \quad (7.45)$$

ou em

$$\psi_0(\theta) < \nu < \psi_1(\theta), \quad (7.46)$$

com  $\psi_1(\theta) = \psi_1(\delta_1, \delta_2, p_1, p_2, \sigma, \mu)$ . Isto implica que a demanda pelo bem 1 é a diferença  
 de escolha entre as alternativas 0 e 2:

$$s_1(\theta) = F(\psi_1(\theta)) - F(\psi_0(\theta)). \quad (7.47)$$

Generalizando obtemos a demanda pelo bem  $j$ :

$$\frac{\delta_{j+1} - \delta_j}{p_{j+1} - p_j} < \nu < \frac{\delta_j - \delta_{j-1}}{p_j - p_{j-1}} \quad (7.48)$$

$$s_j(\theta) = F(\psi_j(\theta)) - F(\psi_{j-1}(\theta)), \quad (7.49)$$

para  $j = 1, \dots, J$ , e com o conjunto de parâmetros a ser estimado

$$\theta = (\delta_1, \dots, \delta_J, \mu, \sigma).$$

Se você desenhar um gráfico com a distribuição de  $\nu$ , os valores  $\psi_j$  representam pontos  
 de corte na distribuição e as participações de mercado ( $s$ ) são a integral da densidade  
 entre estes pontos de corte. Observe as condições necessárias para esta solução. Em  
 particular  $U_{ij} < U_{ij+1} \Rightarrow U_{ij} < U_{ik}$ , para qualquer  $k > (j + 1)$  e analogamente para a  
 desigualdade reversa, se ambos  $\delta_j$  é crescente em  $j$  e  $(\delta_{j+1} - \delta_j)/(p_{j+1} - p_j)$  é crescente

<sup>12</sup>Como os bens são ordenados, então  $-\nu p_1 + \delta_1 < 0 \Rightarrow \nu p_1 < \delta_1 \Rightarrow \nu < (\delta_1/p_1)$ .

em  $j$ . Se um dos bens estiver fora de ordem, o modelo não fornecerá demanda pelo bem – ele nunca será comprado. Portanto, as previsões do modelo são condicionais a  $(\delta, p)$ .

Para estimar precisamos de um processo gerador dos dados ou um processo gerador para os *market shares*. Aqui simplesmente assumimos que observamos as escolhas de uma amostra aleatória de tamanho  $n = \sum_{j=0}^J n_j$ . Cada indivíduo escolhe um bem de um número finito de células (entradas no modelo). As escolhas são mutuamente exclusivas e exaustivas, implicando na distribuição multinomial dos resultados e

$$L_j \propto \prod_j s_j(\theta)^{n_j}, \quad (7.50)$$

então escolha os parâmetros  $\theta$  que maximiza

$$\max_{(\theta)} n \sum_j s_j^o \log [s_j(\theta)], \quad (7.51)$$

que é equivalente a seguinte minimização:<sup>13</sup>

$$\min_{(\theta)} \sum_j \frac{(s_j^o - s_j(\theta))^2}{s_j(\theta)}. \quad (7.52)$$

Esta função é chamada de  $\chi^2$ -mínimo ou, se substituir  $s$  pelos *market shares* observados  $s^o$ , é chamada de  $\chi^2$ -mínimo modificado. Neste problema encontramos os parâmetros que forneça a melhor aproximação entre  $s$  e  $s^o$ . A estimação de (7.51) pode ser realizada por máxima verossimilhança e a estimação de (7.52) por mínimos quadrados (não-linear). Se o pesquisador tiver instrumentos para os preços ela pode ser estimada por algum método de variáveis instrumentais. Por qualquer método escolhido a estimativa é não-linear, uma vez que os *market shares* previstos possuem relações não-lineares com os parâmetros  $\theta$ .

#### 7.4.6 Problemas de Estimação com Modelos Simples: Logit e Vertical Puro

Existem problemas específicos para cada um dos modelos logit e vertical e um problema comum.

Modelo vertical. Neste caso as elasticidades preço-cruzada apenas mostram relações com os bens “vizinhos”. Os vizinhos são definidos pelo ordenamento dos preços. Também existe problema com a elasticidade-própria. Geralmente a estimativa entrega valores elevados para a elasticidade-própria de bens mais caros, embora intuitivamente isto faça sentido – consumidores de alta renda são menos sensíveis a preço.

No modelo logit existe o problema devido a IIA (independência e irrelevância das alternativas). A distribuição das preferências de um consumidor sobre outros produtos, que

<sup>13</sup>Esta é equivalente devido a equivalência assintótica de primeira-ordem.

não seja o que ele compra, não depende completamente dos demais bens. A propriedade IIA implica que a derivada própria do preço é  $(\partial s_j / \partial p_j) = -s_j(1 - s_j)$  (veja (??)). No logit ela depende apenas do *market share* do produto. Isto implica que dois bens diferentes com a mesma participação de mercado devem ter o mesmo *markup* quando firma de um único produto operam em equilíbrio “Nash em preços.”

No logit as derivadas parciais são  $s_j s_k$  (veja (??)). Dois bens com o mesmo *market share* possuem as mesmas elasticidades preço-cruzada com qualquer outro bem.

Nenhum dado nunca irá mudar as implicações/resultados destes dois modelos. Se suas estimativas não satisfazem estes resultados, então o está incorrendo em algum erro de estimação (provavelmente no código).<sup>14</sup>

O problema comum a estes modelos são simultaneidade e “overfitting.” Se  $n$  é grande o pesquisador terá o problema de “overfitting” – o modelo estará errado. Geralmente este problema aparece claramente em produtos de consumo, pois possuem muitas características. Ele não será tão óbvio em mercados de produtos para produtores, i.e. mercados de insumos. Em seguir vamos tratar de estratégias de soluções para estes problemas.

#### 7.4.7 B-Logit: Solução de Berry

Uma possível fonte de erro dos problema de simultaneidade e de “overfitting” é a característica não-observada ou não-medida. Em certas situações a característica não-observada é um atributo de qualidade ou percepção que não pode ser capturado diretamente por meio de dados. Em outras situações, o pesquisador sabe que existe uma característica quantificável, mas não possui dados.

Bens de consumo tipicamente possuem um número de características que importam para os consumidores. Por exemplo, consumidores se importam com ar-condicionado, potência e consumo dos veículos. Além destas características, a confiabilidade do carro é importante. Esta é tipicamente uma característica não-observada. Além disso, um ponto importante em relação às características é de que não devemos utilizar todas as características de um produto, pois caso contrário retornamos ao problema de muitos parâmetros. Usualmente se incluem as principais características dos produtos de consumo, deixando os demais efeitos para serem contabilizados pela característica não-observada. Para bens intermediários isto é um problema menor, pois a introdução de característica não-observada parece ser um problema menor.

O primeiro tratamento explícito sobre o problema de simultaneidade e “overfitting” foi dado por Berry (1994). Considere o conjunto de características observadas e não-observadas de um produto de consumo  $j$ :

$$\delta_j = \sum_{\kappa} x_{j,\kappa} \beta_{\kappa} - \alpha p_j + \xi_j, \quad (7.53)$$

<sup>14</sup>Veja Pakes (2020).

como antes  $x_{j,\kappa}$  são as  $\kappa$  características observadas do produto  $j$  e  $\xi_j$  a característica não-observada. Não tente estimar  $\{\xi_j\}$ , mas assuma que elas são saques aleatórios de alguma distribuição e use as propriedades da distribuição para estimar  $\beta$ .

Uma hipótese tradicional para este tipo de problema é assumir que  $\xi_j$  seja independente da média dos  $x$ , mas não em relação ao preço  $p$ . Neste caso estamos assumindo que os consumidores conhecem  $\xi_j$  e uma vez que os consumidores conhecem, naturalmente a firma também sabe da característica não-observada e a leva em conta no processo de escolha de preços.

Berry (1994) propôs uma estratégia de identificação da característica não-observada dos produtos baseado no modelo logit. Ele assume que  $n$  é bem grande e então que  $s_j^o = s_j(\xi, \dots; \theta_0)$ , tal que  $\theta_0$  representa o valor verdadeiro de  $\theta$ . Para cada  $\theta$  existe apenas um  $\xi$  que torna o *market share* previsto  $s$  o mesmo que o observado  $s^o$ . Isto forma um sistema de  $J$  equações em  $J$  incógnitas.

Intuitivamente dizemos que **invertemos** o sistema de demanda para encontrar  $\xi$  como função do vetor de parâmetros. A forma como fazemos isto depende da forma funcional do modelo de demanda. Uma vez que se tem  $\xi(\theta)$ , temos o distúrbio isolado e podemos estimar como em processos tradicionais. Ou seja, fazemos hipóteses sobre as propriedades do distúrbio verdadeiro, as propriedades para  $\xi(\theta_0)$ , e encontramos que valores de  $\theta$  que faz com que a análoga amostral destas propriedades seja o mais próximo possível do valor verdadeiro.

**Exemplo 1: Modelo Vertical Puro.** Considere a inversão na relação entre o *market share* observado  $s^o$  e os parâmetros de interesse  $\theta$ :

$$s_0^o = F(\psi_0(\theta, \xi)) \Rightarrow y_0 \equiv F^{-1}(s_0^o) = \psi(\theta, \xi).$$

Observe que  $y_0$  pode ser calculado a partir do conhecimento dos dados de mercado dos produtos. De forma similar ao modelo vertical puro original, podemos calcular a participação de mercado do produto 1 em relação ao conhecimento da participação do produto 0. Então

$$s_1^o = F(\psi_1(\cdot)) - F(\psi_0(\cdot)),$$

$$\Rightarrow y_1 \equiv F^{-1}(s_0^o + s_1^o) = \psi_1(\cdot) \dots$$

$$\Rightarrow \{y_j\} = \{\psi_j(\theta, \delta_j)\}.$$

Uma vez que se tem esta sequência o economista será hábil para calcular  $\xi_j$  como uma função de  $(y, x, p; \beta, \mu, \sigma)$ , diga que seja  $\xi_j(y, x, p; \beta, \mu, \sigma)$ . De forma resumida escrevemos  $\xi_j(s, x, p; \theta)$ , para  $\theta = (\beta, \mu, \sigma)$ .

**Exemplo 2: Logit.** A fórmula de inversão é mais simples para o modelo logit.<sup>15</sup> Considere a relação dos *market shares* com as características:

$$\ln[s_j] - \ln[s_0] = \delta_j(\theta) \equiv x_j\beta - \alpha p_j + \xi_j. \quad (7.54)$$

Observe que o que está do lado esquerdo são dados – as participações de mercado dos bens considerados e o tamanho do *outside good*. Então a inversão é simples:

$$\xi_j(s, x, p; \theta) = \ln[s_j] - \ln[s_0] - x_j\beta + \alpha p_j. \quad (7.55)$$

Esta inversão funciona, pois se assume que a diferença entre a participação de mercado observada e a simulada é contabilizado pela característica não-observada, por isso ele possui a interpretação de erro estrutural. Esta inversão somente funciona se o tamanho do mercado é grande, i.e.  $N \rightarrow \infty$ . Isto é, acreditamos que os dados são observados sem qualquer erro amostral.

O modelo de Berry (1994) funciona muito bem para o caso do logit com grupamentos ( $g$  – chamado também de *nested logit*). O modelo é:

$$\ln[s_j] - \ln[s_0] = \delta_j(\theta) \equiv x_j\beta - \alpha p_j - \sigma \ln(s_{j|gt}) + \xi_j. \quad (7.56)$$

Este essencialmente o mesmo caso do logit mas com o termo de erro que contabiliza a participação em cada sub-grupo. Agora temos um segundo parâmetro endógeno além do preço ( $s_{j|gt}$ ). Um bom instrumento para identificar  $\sigma$  é o número de produtos em cada sub-grupo  $g$ .

Estimação dado  $\xi_j(\cdot)$ . Em ambos os modelos podemos recuperar a característica não-observada  $\{\xi_j(s, x, p; \theta)\}$ . Em seguida desejamos escolher nossas restrições de momentos. Observe que a característica não-observada tem interpretação de termo de erro – um resíduo. Aqui a volta o problema tradicional de simultaneidade na estimação da demanda: uma vez que consumidores conhecem a característica não-observada  $\xi_j$  de cada produto, assumimos que as firmas também a conhecem. Logo para qualquer solução de equilíbrio de mercado teremos:

$$p_j = p(x_j, \xi_j, x_{-j}, \xi_{-j}),$$

com  $x_{-j}$  representando as características dos demais bens que não seja o  $j$  e, analogamente, o mesmo para  $\xi_{-j}$ .

<sup>15</sup> Considere o market share do produto  $j$  com a característica não-observada:

$$s_{jt} = \int \frac{\exp[x_{jt}\beta_i + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta_i + \xi_{kt}]} f(\beta_i | \theta),$$

adicionando a característica não-observada na função utilidade se tem:

$$u_{ijt} = x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ij},$$

tome o log natural do outside good:  $\ln s_{0t} = \ln \{1 + \sum_k \exp[x_{kt}\beta_i + \xi_{kt}]\}$  e agora o log natural do produto  $j$ :  $\ln s_{jt} = [x_{jt}\beta_i + \xi_{jt}] - \ln \{1 + \sum_k \exp[x_{kt}\beta_i + \xi_{kt}]\}$ . Tomando a diferença entre os dois se tem:  $\ln[s_j] - \ln[s_0] = x_{jt}\beta_i + \xi_{jt}$ .

Uma vez que  $p$  é função de  $\xi$ , não podemos usar condições de momento que tenha interação de  $p$  com  $\xi(\theta)$ . Este é o problema de simultaneidade que vimos no capítulo anterior e implica que devemos utilizar uma solução de variáveis instrumentais para controlar pela endogeneidade do preço. Formalmente, o problema é que  $E[p\xi(\theta)] \neq 0$ .<sup>16</sup> Também é importante que o número de produtos varie entre mercados (no caso de painel longitudinal, veja abaixo).

Como todo modelo aplicado, a estimação da demanda utiliza algumas hipóteses para identificação econométrica. Tradicionalmente se assumem algumas hipóteses sobre exogeneidade. Neste tipo de modelo a principal hipótese é  $E[\xi_j | x, w] = 0$ , tal que  $x$  é o vetor que contém todas as características de todos os produtos, não apenas do bem  $j$ . O vetor  $w$  são todos os fatores pelo lado do custo, que tipicamente assumimos que são deslocadores ortogonais aos choques de demanda. Essas hipóteses são fortes e deixam o economista com componentes dos vetores  $x$  e  $w$  como instrumentos em potencial.

Outro tipo de hipótese identificadora pode ser assumida a partir da estrutura longitudinal de uma base de dados. Dada a estrutura de painel poderia se assumir que a variável não-observada de cada produto muda ao longo do tempo com a seguinte estrutura

$$\xi_{jt} = \rho\xi_{jt-1} + u_{jt}, \text{ com } E[u_{jt} | x_1, \dots, x_j] = 0.$$

Isso requer que os produtos sejam observados ao longo do tempo, mas adicionalmente pode trazer preocupações sobre problemas de seleção se ocorrer saída de produto dos mercados.

Se a estrutura do painel for múltiplos mercados para o mesmo produto  $j$ , então podemos fazer a distinção entre choques locais e característica não-observada do produto que independe da localidade e tempo. Uma estrutura apropriada para esta hipótese é:

$$\xi_{jr} = \xi_j + u_{jr},$$

com hipóteses apropriadas sobre  $u_{jr}$ .<sup>17</sup>

De forma geral estas hipóteses boas para trabalhos aplicados. O importante é economista saber escolher qual estrutura usar frente ao caso em análise.

#### 7.4.8 Complementares Estratégicos e Substitutos

Assuma uma firma 1 de produto único e escreva a função lucro em um problema de Bertrand com produtos diferenciados como

$$\pi^1(p_1, p_2; \cdot) = D_1(p_1, p_2; \cdot)[p_1 - c_1] \quad (7.57)$$

<sup>16</sup>Lembrando que  $p$  sem o indicador  $j$  representa um vetor com todos os preços. Observe também que se ocorrer o mesmo problema com alguma característica, é preciso ter um instrumento específico para ela. Geralmente a literatura assume que as características são exógenas, pois elas mudam mais lentamente do que o preço do bem de consumo.

<sup>17</sup>Observe que se existem diferenças inter-regionais ou inter-grupo então existe algo gerando estas e o economista deve assumir que estas são ortogonais aos instrumentos.

assumindo, por simplicidade, custo marginal  $c_1$  como constante e  $D^1(\cdot)$  representando a função demanda da firma 1. Portanto, os preços serão complementos estratégicos ou substitutos estratégicos de acordo com a segunda derivada parcial da função lucro:

$$\frac{\partial^2 \pi^1(\cdot)}{\partial p_1 \partial p_2} = \pi_{1,2}^1(\cdot). \quad (7.58)$$

Se a derivada parcial cruzada for maior do que zero então o preço é estratégico complementar, se for menor do que zero ele é substituto.

No modelo BLP conseguimos acessar corretamente se dois produtos são complementos ou substitutos estratégicos. Vamos mostrar o porque: no modelo de Bertrand com produto diferenciado (Nash em preços), a condição de primeira ordem da firma com um único produto é

$$\frac{\partial \pi^1}{\partial p_1} = \frac{\partial D_1(\cdot)}{\partial p_1} [p_1 - c_1] + D^1(\cdot) = 0, \quad (7.59)$$

reorganizando em função da margem  $p_1 - c_1$ , se tem:

$$p_1 - c_1 = \frac{D^1(\cdot)}{-\frac{\partial D_1(\cdot)}{\partial p_1}}. \quad (7.60)$$

Esta é a relação entre margem e elasticidade.

Agora vamos derivar a condição de primeira ordem (7.59) em relação ao produto da firma 2:

$$\frac{\partial^2 \pi^1}{\partial p_1 \partial p_2} = \frac{\partial^2 D_1(\cdot)}{\partial p_1 \partial p_2} [p_1 - c_1] + \frac{\partial D^1(\cdot)}{\partial p_2}, \quad (7.61)$$

com a margem  $p_1 - c_1$  como definida anteriormente.

Algumas observações podem ser feitas. Primeiro, uma vez que se especifica a forma do sistema de demanda e o vetor corrente de preços saberemos o resultado (complementar ou substituto). Em segundo lugar, o último termo de (7.61),  $\frac{\partial D^1(\cdot)}{\partial p_2}$ , será sempre positivo ou zero em todos os modelos de produtos diferenciados, uma vez que todos os bens são substitutos. Isto apenas diz que se o preço do bem 2 sobe mais consumidores compram o bem 1.

A visão convencional diz que se o primeiro termo não for muito negativo, os preços serão *complementos* no modelo de produto diferenciado. Além disso, o quanto o primeiro termo é negativo ou não depende do quanto pode parecer ser “detalhes técnicos” da função demanda. Neste sentido, a maior tendência é que o último termo domine o sinal.

Entretanto, simples generalizações de modelos teóricos mostram que os preços não costumam ser substitutos. A intuição é a seguinte. É verdade que quando o preço do bem 2 sobe maior será a demanda pelo bem 1. Entretanto, a medida que o preço do bem 2 sobe, o aumento de consumo do bem 1 é realizado pelos consumidores que são mais sensíveis a mudança nos preços. Isto é exatamente porque eles deixam de consumir o bem 2 e migram para o bem 1. Se a firma 1 receber demanda de consumidores mais sensíveis a

preço, então ela tem incentivo para reduzir o seu preço. O quão importante são estes dois efeitos e, portanto, qual domina, isto é deixado para os dados decidirem. Para o exemplo de automóveis (BLP), metade dos pares de preços são substitutos estratégicos e metade são complementos estratégicos.

#### 7.4.9 BLP

BLP é um modelo de demanda micro, uma generalização do logit que permite características observadas e não-observada, que agregada explicitamente para obter a demanda de mercado por um produto  $j$ . O fato de que este modelo começa com um modelo microeconômico permite que o mesmo *framework* seja usado para estruturar a análise de demanda que utiliza micro-dados, amostras de extratos, ou dados agregados de produtos, bem como qualquer combinação deles.

Começamos a exposição com uma função utilidade (7.1) alterada para representar o framework micro:

$$U_{ij} = \sum_k x_{jk} \beta_{ik} + \xi_j + \epsilon_{ij}, \quad (7.62)$$

$$\beta_{ik} = \lambda_k + \beta_k^o z_i + \beta_k^u \nu_i, \quad (7.63)$$

tal que  $x_{jk}$  e  $\xi_j$  são as características observadas e não-observadas dos produtos, respectivamente. As variáveis  $z_i$  e  $\nu_i$  são os vetores de atributos observados e não-observados dos consumidores. Observe que estes atributos podem mudar dependendo da aplicação. Os vetores  $\beta_k^o$  e  $\beta_k^u$  determinam os impactos das características observadas e não-observadas do consumidores sobre a utilidade da característica  $k$ . Os  $\lambda_k$  fornecem o impacto da característica do produto  $k$  sobre o termo constante do produto na função utilidade (utilidade média). Por fim, o termo  $\epsilon_{ij}$  representa a preferência individual idiossincrática sobre os diferentes bens (preferências que são independentes dos atributos dos produtos e indivíduos que contabilizamos no modelo).

Substituíamos (7.63) em (7.62) para obter

$$U_{ij} = \delta_j + \sum_{kr} x_{jk} z_{ir} \beta_{rk}^o + \sum_{kl} x_{jk} \nu_{il} \beta_{kl}^u + \epsilon_{ij}, \quad (7.64)$$

com

$$\delta_j = \sum_k x_{jk} \lambda_k + \xi_j. \quad (7.65)$$

Observe que a (7.65) é a equação estrutural do modelo de Berry (1994) para o caso logit.<sup>18</sup> Além disso, o modelo tem dois tipos de termos de interações entre as características do produto e as características do consumidor. Elas são:

**Termo**  $\sum_{kr} x_{jk} z_{ir} \beta_{rk}^o$ : interações entre características observadas dos consumidores ( $z_i$ ) e a características dos produtos.

<sup>18</sup>Nesta formulação, o preço está contido na matriz de características.



**Termo**  $\sum_{kl} x_{jk} \nu_{il} \beta_{kl}^u$ : interações entre características não-observadas dos consumidores ( $\nu_i$ ) e a características dos produtos.

São estes termos de interação que geram elasticidades próprias e cruzadas razoáveis, eliminando o problema gerado pelo IIA. No BLP, se aumentamos o preço de um carro, um tipo muito específico de consumidor deixa este carro, o consumidor que valoriza as características apresentadas pelo veículo. Consequentemente eles irão substituir este veículo por outro com características similares. Portanto, neste ambiente é possível construir modelos com elasticidades próprias e cruzadas razoáveis.

Além disso, aumentos de preços de veículos irá forçar diferentes tipos de consumidores deixarem de consumir tipos diferentes de carros. Automóveis com preços altos tendem a ser os carros que são preferidos por consumidores de alta renda, que não respondem muito a mudanças nos preços. Isto implica que o efeito próprio do preço para este tipo de consumidor será mais baixo para carros mais caros – utilizando equilíbrio Bertrand-Nash para produtos diferenciados, teremos markup mais elevado para este tipo de veículo.

Quando temos apenas dados no nível dos produtos, i.e, não sabemos quem compra qual bem, então o melhor que se pode ter é a informação sobre a distribuição de  $z_i$  utilizando uma base de dados sobre características dos domicílios/famílias. Como o BLP é um modelo de decisão individual, é possível adicionar dados individuais de compra. Quando adicionamos esses micro-dados é importante saber o quanto os atributos observados do consumidor são suficientemente ricos para capturar as fontes de diferença nas preferências pelas características. Se não houver essa riqueza de detalhes, como por exemplo, tamanho da família para relacionar ao tamanho do carro, o pesquisador precisa manter o controle de características não-observadas de atributos do consumidor ( $\sum_{kl} x_{jk} \nu_{il}$ ) ou irá terminar incorrendo nos problemas que surgem com o modelo vertical ou logit.<sup>19</sup>

#### 7.4.10 Markups

No contexto de estimação de produtos diferenciados o markup pode ser recuperado utilizando modelagem do lado da oferta. Para estimação apropriada é preciso especificar um modelo de como as firmas competem. O modelo mais usado é o de Bertrand com produtos diferenciados – BLP (1995, 1999).

Assumimos que existem  $F$  firmas, cada qual produz um subconjunto  $J$  dos produtos comercializados,  $J_f$ , e a decisão de quais produtos vender é assumida ser predeterminada. As hipóteses deste modelo são padrão na literatura de OI.

Para cada firma  $f$ , a função lucro é

$$\pi_f = \sum_{j \in J_f} (p_j - mc_j) s_j(\cdot) \times M \quad (7.66)$$

<sup>19</sup>Para uma introdução ao BLP veja Conlon e Gortmaker (2020).

tal que  $p_j$  e  $mc_j$  são os preço e custo marginal para o produto  $j$ , respectivamente.  $s_j$  é o *market share*, e  $M$  é o tamanho do mercado.

Assuma que o equilíbrio é Nash em preços. Dado que tal equilíbrio existe, os preços resultantes devem satisfazer as condições de primeira ordem:

$$s_j(\cdot) + \sum_{j \in J_f} (p_j - mc_j) \frac{\partial s_j(\cdot)}{\partial p_j} = 0 \quad (7.67)$$

O *markup* para a firma multi-produto,  $b$ , é

$$b(\cdot)_f = \tilde{\delta}(\cdot)^{-1} s_j(\cdot). \quad (7.68)$$

tal que  $\tilde{\delta}$  é uma matriz  $J$  por  $J$ , cujo  $(j, r)$  elemento é dado por:

$$\Delta_{j,r} = \begin{cases} -\frac{\partial s_r}{\partial p_j}, & \text{se } r \text{ e } j \text{ são produzidos pela mesma firma;} \\ 0, & \text{caso contrário.} \end{cases}$$

## 7.5 Concorrência Monopolística

Por exemplo, considere o mercado de restaurantes em uma cidade. O produto oferecido pelos restaurantes é um combinado de refeição e serviços, embora sejam produtos parecidos eles possuem distinções tal que a hipótese de homogeneidade do produto não seja a melhor hipótese.

O modelo de concorrência monopolística<sup>20</sup> foi construído para descrever indústrias como a de restaurantes. Se diz que se tem concorrência monopolística quando as seguintes hipóteses valem:

1. Existe grande número de firmas, tal que o impacto de cada empresa sobre as demais é negligível;
2. Devido a diferenciação de produtos cada firma pode determinar o seu preço, ou seja, a demanda que cada firma tem não é “horizontal” como na concorrência perfeita (com firmas tomadoras de preços);
3. Existe livre entrada e livre acesso a todas as tecnologias disponíveis.

Isto significa que o modelo de concorrência monopolística adota todas as hipóteses da concorrência perfeita com a exceto pela diferenciação de produtos. Esta característica do modelo implica que os produtos não são substitutos perfeitos, levando a possibilidade das firmas decidirem preços.<sup>21</sup>

Para entender melhor os resultados do modelo de concorrência monopolística é útil fazer a distinção entre resultado de curto e longo-prazo. Um resultado de equilíbrio de curto

<sup>20</sup>Chamberlin (1933) e Robinson (1933).

<sup>21</sup>Quanto menor o grau de substituição entre produtos maior será o preço.

prazo de concorrência monopolística diz que o preço recebido por cada firma é maior do que o custo médio – lembrando que este resultado é uma possibilidade. A diferença entre preço e custo depende do número de firmas e do grau de diferenciação dos produtos.

Como o preço é diferente do custo médio, este equilíbrio não é de longo prazo. Se preço é maior do que o custo médio, então existe incentivo para firmas entrarem no mercado. Uma hipótese subjacente ao modelo é a de que todas as firmas tem acesso a mesma tecnologia, com impacto negligível sobre as demais. Isto implica que a empresa entrante espera lucro similar ao da típica empresa incumbente, i.e.  $\pi = (p - CM(q))/q$ , tal que  $CM$  é o custo médio. Por outro lado, se o preço for menor do que o custo médio firmas incumbentes tem prejuízo e devem deixar o mercado.

Portanto, o equilíbrio de longo prazo é a situação em que firmas maximizam lucros, tal que a receita marginal é igual ao custo marginal, e as empresas fazem lucro-zero (ou seja, quando o preço é igual ao custo médio). A condição de lucratividade do equilíbrio de longo prazo determina o número de firmas ativas.

O modelo de concorrência monopolística é muito utilizado em diversas áreas além da organização industrial, tais como macroeconomia e comércio internacional. Por exemplo, a medida que o grau de substitubilidade entre produtos aumenta nos aproximamos do resultado da concorrência perfeita. Esta propriedade de precificação acima do custo médio no curto prazo é uma hipótese atrativa para estudar algum grau de rigidez de preços em questões de macroeconomia.

## 7.6 Problemas de Índice de Preços

Índices de preços estão entre as mais importantes tarefas que um governo pode fazer. A discussão sobre mensuração do custo de vida e o tipo de problema que pode surgir é fundamental em economia. Vale ressaltar que o problema endereçado aqui surge em todos os índices de preços e existe um número questões de economia industrial para corrigir problemas que surgem em índices de preços.

Primeiro considere um modelo teórico de construção de um índice de preços. O custo de vida ( $e_{i,t}$ ) de obter um dado nível de utilidade,  $\bar{U}$ , tipicamente em um ano-base, é calculado como

$$e_{i,t} = \min_{q_{i,t}} \sum p_{j,t} q_{i,j,t} \quad (7.69)$$

sujeito a

$$U(q_i, z_{i,t}) = \bar{U},$$

aqui  $z$  se refere as características dos indivíduos. Isto implica que o custo de vida é função do conjunto de escolha, das características dos indivíduos e da utilidade de consumo de determinada cesta de bens:

$$e = e(\text{conjunto de escolha}_t, z_{i,t}, \bar{U}).$$

Observe que a frase “conjunto de escolha” se refere ao conjunto de produtos disponíveis e seus preços. Em introdução à macroeconomia ou contabilidade nacional o leitor já deve ter tido contato com as definições de índices de preços de Laspeyres e Paasche. Aqui retornamos a estes índices clássicos.

Como consequência da análise acima, a mudança no custo de vida verdadeira de Laspeyres entre dois períodos, 1 e 0, compara o custo de obtenção da utilidade no período-base (0):

$$IPC^V = \frac{e(\text{conjunto de escolha}_1, z_{i,1}, \bar{U}_{i,0})}{e_{e_i,0}}$$

tal que  $IPC^V$  é o verdadeiro índice de preços ao consumidor.

O índice do custo de vida calculado, entre bens que podem aparecer entre 0 e 1, é essencialmente

$$IPC^C = \frac{w_j p_{j,1}}{w_j p_{j,0}},$$

com  $w_j$  representando o peso do produto  $j$  no período anterior. Aqui os pesos são determinados pela gasto realizado anteriormente pelos consumidores, medido por meio de uma pesquisa específica. No Brasil esta pesquisa é a POF – Pesquisa de Orçamento Familiar – do IBGE que ocorre apenas a cada **cinco** anos. Os preços são obtidos de amostras grandes com vários grupos e pontos de venda selecionados aleatoriamente.<sup>22</sup>

Se existe um único indivíduo, e todos os bens disponíveis no período base são disponíveis naquele dado período, este índice deveria garantir que o consumidor poderia comprar a mesma cesta de bens no período corrente bem como no base. Neste sentido, seria garantido a este consumidor a utilidade do período base. Esta é a ideia para a construção de um índice de preços adequado. Consequentemente, se existem muitos consumidores, e nós ponderamos corretamente os tipos de indivíduos, podemos obter a média ponderada da mudança na renda que poderia garantir que cada indivíduo poderia continuar obtendo os mesmos bens obtidos no passado.

Nos EUA, foi criada pelo Senado americano a Boskin Commission em 1995.<sup>23</sup> A função da comissão era estudar possível presença de viés no cálculo do índice de preços ao consumidor. A comissão mostrou que existia os seguintes vieses: (i) viés de substituição, aproximadamente de 0.4% por ano; (ii) viés de bens novos, aproximadamente 0.7% por ano. A rotação da amostra e os procedimentos de substituição de bens das cestas ao longo do tempo seriam responsáveis por este tipo de problema.

Os dois tipos de viés podem ser explicados da seguinte forma. Viés de substituição ocorre porque se mantém a cesta de consumo fixa quando o vetor de preços muda. Todavia, quando o vetor de preços muda, o consumidor pode melhorar sua alocação mudando quantidades consumidas. Se o pesquisador possui estimativas da distribuição das utilidades, este viés pode ser estimado diretamente.

<sup>22</sup>Um ponto de venda é usado para determinar onde a amostra é produzida.

<sup>23</sup>Os membros da Boskin Commission foram: Michael Boskin, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, e Dale Jorgenson.

O segundo viés tem como causa o fato de bens novos nunca serem avaliados diretamente. Bens novos surgem no índice de preços com o procedimento de rotação da base bem como na troca de bens que compõem as cestas. Neste caso, eles somente são avaliados algum tempo depois que eles entram no mercado. Nunca se compara bens novos com velhos, que é a parte de onde vem os ganhos dos bens novos. Por exemplo, quando *smartphone*, *laptop* ou *TV-flatscreen* passaram a fazer parte do índice de preços?

Os remédios adotados para resolver este tipo de viés seriam os seguintes. Primeiro, procedimentos com maior rotação da amostra.<sup>24</sup> O impacto depende da organização industrial de novos bens, um tópico que não é muito estudado. Se o bem é introduzido com um preço inicial baixo para atrair consumidores, ou se o bem é um produto tecnológico avançado, cujo preço permanece estável nos períodos iniciais, mas cai em seguida, a ideia de rotação mais rápida da amostra pode agravar o problema de viés. Importante se observar aqui que desejamos os novos bens incluídos na base, tais como TVs e geladeiras, por exemplo. Se não introduzirmos novos bens no índice de preços, logo estaremos calculando preços de produtos que as pessoas não consomem mais. A questão é quando introduzir – logo, com preços mais baixos ou mais altos, ou depois de algum tempo após os preços estabilizarem.

Outro remédio seria um índice hedônico, algo como vimos na seção 7.1. No espaço das características podemos fazer isto com um índice hedônico. A ideia básica é manter as características constantes e comparar os preços para um dado conjunto de características. O uso de preços hedônicos são tradicionais em organização industrial em trabalhos clássicos como os de Court (1939) e Griliches (1961).

O que é uma análise empírica de preços hedônicos? Como vimos anteriormente, uma regressão hedônica assume a seguinte forma:

$$p_i = x_{i1}\beta_1 + \dots + x_{ik}\beta_k + e_i. \quad (7.70)$$

Agora vamos utilizar a ideia recorrente de que o preço é o custo marginal mais uma margem aplicada sobre o custo, como feito em Pakes (2003). Se estivermos em um mercado e todas as firmas são produtores de um único bem jogando Nash em preços (Bertrand), portanto os preços são (como a firma é possui um único produto não utilizamos o subscrito  $j$ ):

$$p_i = Cmg(\cdot) + \frac{D_i(\cdot)}{|\partial D_i(\cdot)/\partial p|}, \quad (7.71)$$

aqui  $Cmg(\cdot)$  é o custo marginal e o segundo termo,  $\frac{D_i(\cdot)}{|\partial D_i(\cdot)/\partial p|}$  é o markup que varia inversamente com a elasticidade da demanda no ponto.

Vamos definir uma função hedônica como  $h(x)$ , em que o preço é função das características do bem. Neste sentido, dizemos que na média o preço pode ser previsto como

<sup>24</sup>Geralmente os procedimentos de índices de preços incluem amostras selecionadas de bens e serviços. Se a seleção amostral destes produtos for maior no tempo é possível que os produtos sejam incluídos com maior velocidade no índice de preços.

função de  $x$ . Isto equivale a assumir a esperança matemática condicional de  $p_i | x_i$ . Aplicando a esperança condicional em relação a  $x$  ( $E[\cdot | x_i]$ ) na equação de preços (7.71), temos:

$$h(x_i) \equiv E[p_i | x_i] = E[Cmg(\cdot) | x_i] + E \left[ \frac{D_i(\cdot)}{|\partial D_i(\cdot)/\partial p|} | x_i \right], \quad (7.72)$$

Nesta equação a expectativa integra sobre a aleatoriedade no processo gerador das características de produtos concorrentes, preços dos insumos (que também podem conter markups), e produtividade. Isto significa que nem todos os produtos são iguais. Por exemplo, a TV de uma marca específica possui algumas características distintivas, além de preços de insumos e margens próprias da empresa produtora.

Os markups, sobre insumos bem como sobre os produtos, são determinados pelos bens concorrentes no mercado e pela distribuição dos gostos dos consumidores, portanto também não deve gerar funções estáveis de características ao longo do tempo em um determinado mercado, ou entre mercados. Além disso, dado que qualquer heterogeneidade nas preferências relativas sobre as características, os coeficientes destas características não possuem qualquer interpretação como o valor marginal dos  $x$ . O uso do modelo hedônico para interpretar um coeficiente apenas é provavelmente problemático. A equação de preços possui dois componentes, uma função custo e um termo de markup. O termo de markup será relevante dado que não podemos abdicar da ficção de que preço igual a custo marginal. As um resultado, não podemos nem mesmo sinalizar a relação entre os coeficientes hedônicos e as características verticais nos modelos de utilidade. Neste sentido, nem pense em interpretar as magnitudes dos termos hedônicos.

**Exemplo dos Computadores.** Na Tabela 7.5 é apresentada as características dos computadores comercializados. Todas as características são verticais e crescem ao longo do tempo. Observe a seriedade na composição do índice de preços ao consumidor dos EUA (CPI). A fração de computadores selecionados para a amostra no período base que foram disponíveis para calcular o preço varia de 8 a 24%. Testa é uma indústria tecnológica em que os produtos ficam melhores ao longo do tempo (veja Pakes, 2003).

Na Tabela 7.5 mostramos os resultados de Pakes (2003), em que ele compara a performance dos modelos utilizados oficialmente (MM) com modelos de preços hedônicos. Nestas tabelas se tem dois modelos hedônicos, dois modelos MM (Tornqvist e Laspeyres), e um modelo de regressão de dummies que utiliza características e variáveis indicadoras de tempo. Não apenas os modelos MM capturam o sinal errado, mas eles são negativamente correlacionados com o modelo hedônico. A razão para isto é clara, em anos quando uma nova geração de chips entra no mercado é quando vários chips antigos são descontinuados. Os chips que permanecem são principalmente aqueles que entram mais cedo no mercado (da nova geração de chips). A especificação com variáveis dummies fica com resultados intermediários entre MM e hedônica.

Nos EUA, o BLS (Bureau of Labor Statistics) está em processo de atualizar a forma de cálculo e amostragem dos dados de tal forma que eles possam utilizar índices hedônicos baseados em regressões que são estimadas separadamente em cada período (em nível de

Tabela 7.4: Características de Computadores Comercializados

Características	1995	1996	1997	1998	1999
Velocidade (Mhz)					
min	25	25	33	140	180
média	65	102	153	245	370
max	133	200	240	450	550
RAM (MB)					
min	2	4	4	8	16
média	7	12	18	42	73
max	32	64	64	128	128
HD (GB)					
min	.1	.1	.2	.9	2
média	.5	1	1.8	4.5	8.5
max	1.6	4.3	16.8	25.5	
Observações					
Observações	264	237	199	252	154
matched para $t + 1$	44	54	16	29	n.d.

Fonte: Pakes (2003).

escala). Para ir em frente de forma consistente seria necessário um sistema de demanda, mas o BLS e outras agências são hesitantes para realizar isto em modo experimental.

Tabela 7.5: Índices de Preços Alternativos para PC

Ano	1995/1996	1996/1997	1997/1998	1998/1999	média	
1. Hedônico	base	-0.097 (.040)	-.108 (.063)	-.295 (.045)	-.155 (.099)	-.164 (.091)
	altern.	-0.94 (.039)	-.111 (.052)	-.270 (.044)	-.155 (.054)	-.156 (.079)
2. Matched Model (MM)	Tornqvist	.12	.002	.09	.011	.028
	Laspeyres	-0.13	-.002	-.08	-.011	-.027
	matched (%)	16.6	22.8	8.0	11.5	14.7
3. Variáveis Dummy	base	-.135 (.038)	-.098 (.035)	-.160 (.027)	-.170 (.040)	-.141 (.032)
	altern.	-.152 (.04)	-.122 (.032)	-.213 (.041)	-.143 (.028)	-.158 (.039)

Fonte: Pakes (2003).

**Características Não-Observadas e Seleção (Erikson e Pakes, 2011).** Os autores utilizaram dados amostrais do BLS que são utilizados para formar os índices de preços. A correção hedônica assume que a seleção é baseada nas características observáveis.

Um exemplo é dado para TVs. Existe 20% de turnover ao longo do intervalo amostral. Este turnover é o número de modelos de TVs que se deixa de acompanhar no processo de amostragem (esse número é similar a PCs). Existe ampla evidência indicando que os bens que deixam a amostra possuem preços que estão caindo desproporcionalmente.

Quando computamos um índice hedônico baseado em um conjunto de características comparável ao que o BLS faz, obtemos um índice que é mesmo valor de um índice MM. Na base de dados de TVs, a grande maioria das características é medida por variável dummy: o modelo possui ou não determinada funcionalidade. Em geral não há medida cardinal das características.

A saída de produtos da base é desproporcionalmente de produtos com preços elevados que possuem a maioria das funcionalidades captadas pelas variáveis dummies. Estes produtos deixam o mercado porque se tornam obsoletos.

O problema desta medida baseado em dummies é que a seleção de TVs é parcialmente baseada em não-observáveis, para usar uma terminologia de econometria. Os controles hedônicos para mudanças nos valores das características observadas de todos os bens,



mas perde a mudança no valor das características não-observadas. Os modelos MM controlam por mudanças no valor das características observadas e não-observadas dos bens que são sempre ofertados, mas não controlam por todos os bens que deixam a amostra.

Observe as regressões que somente usam características não tão distintas dos índices MM (Tabela 7.6).<sup>25</sup> Por outro lado, Erickson e Pakes (2011) comparam no modelo hedônico os produtos que continuam no mercado frente aqueles que deixam a base de dados. Os produtos que vão deixar a base um período à frente são bens cujo valor das características estão diminuindo. Isto é mais evidente quando se compara pela característica não-observada destes produtos.

Tabela 7.6: Índices para TVs

Cálculo do Índice	MM	Hedônico
Regressão Ln-Preço Hedônico		
Hedônico S24	-10.11	-10.21
E-P (entre meses)	(5.35)	(7.53)
S24 % l.t. mm		.50
Regressão Ln-Preço Hedônico		
Hedônico S10	-10.11	-8.82
E-P (entre meses)	(5.35)	(7.05)
S10 % l.t. mm		.40

Fonte: Erickson e Pakes (2011). Índices de preços calculados entre junho de 2000 e dezembro de 2002. A variável dependente nos modelos hedônicos é o log do preço. As variáveis do lado direito são as especificações S10 e S24. As taxas na tabela são as taxas de inflação anual calculadas de cada modelo. A última linha fornece a fração dos 30 meses onde o índice hedônico é menor do que o índice do *matched-model* (% l.t. mm).

## 7.7 Anexo: Detalhes do Logit

O fundamental é que *alguns destes fatores que causam as escolhas são observados e outros não*. Faça:

- $x$  fatores observados (características);
- $\varepsilon$  fatores não-observados;

<sup>25</sup>Na Tabela S10 representa as variáveis dependentes do modelo que são: log do tamanho da tela (polegadas), uma dummy indicadora para projeção, a interação entre estas duas variáveis, o quadrado do log do tamanho da tela, uma dummy da variável do quanto a observação vem da sub-amostra mensal do BLS, dummies indicadoras para picture-in-picture, TV de tela plana, conexão HDTV, marca com reputação de alta-qualidade, outra para reputação de baixa qualidade. S24 são as variáveis do modelo S10 adicionando mais variáveis dummies, incluindo localização de LA, NYC, Chicago, tipo de loja, tipo de som, numero de inputs, segmento de TV antiga, etc. (veja Tabela 13 de Erickson e Pakes, 2011).

- $y = h(x, \varepsilon)$  a escolha dos agentes é relacionada aos fatores por meio de uma função (processo comportamental).

Como temos um fator não observado,  $\varepsilon$ , então a escolha dos agentes não é determinística.

- $f(\varepsilon)$  função de densidade de  $\varepsilon$ , que é considerado aleatório.

A probabilidade de que um agente escolha um resultado do conjunto de possibilidades é simplesmente a probabilidade de que os fatores não observados são tais que o processo comportamental resulta na escolha:

$$P(y | x) = \text{Prob}(\varepsilon \text{ s.a. } h(x, \varepsilon) = y)$$

Defina a função indicadora  $I[h(x, \varepsilon) = y]$ : ela assume 1 quando for verdade e 0 caso contrário. Então a probabilidade de que o agente escolha  $y$  é o valor esperado da função indicadora sobre todos os valores possíveis dos fatores não observados:

$$P(y | x) = \text{Prob}(I[h(x, \varepsilon) = y] = 1) = \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon \quad (7.73)$$

Para calcular a probabilidade, esta integral deve ser solucionada.

Agora suponha que os termos aleatórios possam ser decompostos em duas partes:  $\varepsilon_1$  e  $\varepsilon_2$ . Faça a densidade conjunta ser  $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2)$ . A densidade conjunta pode ser expressa como o produto de uma densidade marginal com e uma condicional:  $f(\varepsilon_1, \varepsilon_2) = f(\varepsilon_2 | \varepsilon_1) f(\varepsilon_1)$ . Com esta decomposição a probabilidade na equação (7.73) pode ser expressa como:

$$= \int_{\varepsilon_1} \left[ \int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2 | \varepsilon_1) d\varepsilon_2 \right] f(\varepsilon_1) d\varepsilon_1 \quad (7.74)$$

Suponha que existe uma forma-fechada para a integral entre colchetes.

O problema de escolha do agente  $n$  pela escolha  $j$  precisa ser especificado pela modelagem do termo de erro.

O modelo logit é obtido assumindo que cada  $\varepsilon_{nj}$  é distribuído como valor extremo iid (ou valor extremo tipo I / duplo exponencial). A densidade para cada componente não-observável da utilidade é:

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \quad (7.75)$$

e a distribuição cumulativa é

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}} \quad (7.76)$$

A variância desta distribuição é  $\pi^2/6$ . Assumindo que a variância é  $\pi^2/6$  estamos implicitamente normalizando a escala da utilidade.

A diferença entre duas variáveis valor extremo é distribuída como logística. I.e., se  $\varepsilon_{nj}$  e  $\varepsilon_{ni}$  são valor extremo iid, então  $\varepsilon_{nji}^* = \varepsilon_{nj} - \varepsilon_{ni}$  segue a distribuição logística:

$$F(\varepsilon_{nji}^*) = \frac{e^{\varepsilon_{nji}^*}}{1 + e^{\varepsilon_{nji}^*}} \quad (7.77)$$

A distribuição de valor extremo possui caudas um pouco mais grossas do que uma normal (pode suportar mais casos extremos). A distinção entre as duas não costuma ser perceptível quando aplicadas aos dados. A hipótese ou característica crucial da valor extremo é que os erros são independentes uns dos outros. A independência significa que a porção não-observada da utilidade de uma alternativa não é relacionado com o termo não-observado de outra alternativa. Essa hipótese não é tão restritiva e pode ser interpretada como resultado natural de um modelo bem especificado.<sup>26</sup>

A probabilidade das escolhas é uma reorganização de (??):

$$P_{ij} = \text{Prob}(\varepsilon_{ij} < \varepsilon_{ik} + V_{ij} - V_{ik}, \forall j \neq k)$$

Se esse é considerado como dado, esta expressão é a distribuição cumulativa para cada  $\varepsilon_{ij}$  solucionado em  $\varepsilon_{ij} + V_{ij} - V_{ik}$ . Em termos da distribuição valor extremo será:  $\exp(-\exp(-(\varepsilon_{ij} + V_{ij} - V_{ik})))$ . Como por hipótese a distribuição cumulativa sobre todo  $j \neq i$  é o produto das distribuições cumulativa individuais:

$$P_{ij} | \varepsilon_{ij} = \prod_{j \neq k} e^{-e^{-(\varepsilon_{ij} + V_{ij} - V_{ik})}} \quad (7.78)$$

Como  $\varepsilon_{ij}$  não é dado, a probabilidade de escolha é a integral de  $P_{ij} | \varepsilon_{ij}$  sobre todos os valores de  $\varepsilon_{ij}$  ponderada pela densidade (7.75):

$$P_{ij} = \int \left( \prod_{j \neq k} e^{-e^{-(\varepsilon_{ij} + V_{ij} - V_{ik})}} \right) e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} d\varepsilon_{ij} \quad (7.79)$$

A manipulação desta equação e fazendo  $V_{ij} = \beta' x_{ij}$  resulta na expressão de forma fechada para a *probabilidade de escolha logit*:

$$P_{ij} = \frac{e^{\beta' x_{ij}}}{\sum_j e^{\beta' x_{ik}}} \quad (7.80)$$

Por exemplo, no caso de uma escolha binária a probabilidade pode ser representada em uma forma mais resumida:

$$P_1 = \frac{1}{1 + e^{\beta' x_{ij} - \beta' x_{i1}}} \quad (7.81)$$

#### Parâmetro de escala

Em geral a utilidade pode ser expressada como:  $U_{ij}^* = V_{nj} + \varepsilon_{ij}^*$ , onde a variância do erro é  $\sigma^2 \times (\pi^2/6)$ . Utilidade normalizada:  $U_{ij} = V_{ij}/\sigma + \varepsilon_{ij}$ , tal que  $\varepsilon_{ij} = \varepsilon_{ij}^*/\sigma$ .

<sup>26</sup>Sob a hipótese de independência, o termo de erro de uma alternativa não adiciona informação ao pesquisador sobre o termo de erro de outra alternativa. Explicado de outra forma, se especifica  $V_{nj}$  tal que o restante da função utilidade (não-observado) se essencialmente um “white noise.”

O termo não observado tem variância  $\pi^2/6$ :  $\text{Var}(\varepsilon_{ij}) = \text{Var}(\varepsilon_{ij}^*/\sigma) = (1/\sigma^2)\text{Var}(\varepsilon_{ij}^*) = (1/\sigma^2) \cdot \sigma^2 \cdot \pi^2/6 = \pi^2/6$ . Então a probabilidade de escolha é:

$$P_{ij} = \frac{e^{(\beta^*/\sigma)x_{ij}}}{\sum_k e^{(\beta^*/\sigma)x_{ik}}} \quad (7.82)$$

Cada coeficiente é re-escalado por  $1/\sigma$ . Observe que apenas a razão  $(\beta^*/\sigma)$  pode ser estimada. Usualmente o modelo é expresso na forma re-escalada:  $\beta = (\beta^*/\sigma)$ .

O parâmetro de escala não afeta a razão dos coeficientes. Disposição à pagar, valor do tempo, e outras medidas de taxas marginais de substituição não são afetadas pelo parâmetro de escala. Apenas a interpretação da magnitude é afetada.

## Capítulo 8

# Tecnologia, Entrada e Estrutura de Mercado

A estrutura de mercado são as empresas que operam em determinado mercado. Como vimos anteriormente, firmas são heterogêneas, sendo distintas em tamanho, produtos, experiência, etc. A questão central deste capítulo trata da determinação da estrutura de mercado, i.e. do número de firmas que operam em um mercado.

As empresas em determinado mercado podem ser classificadas como empresas que permanecem funcionando ao longo do tempo, as incumbentes, e as empresas que entram e saem do mercado. Este processo de entrada e saída de empresas determina o total de firmas funcionando, bem como a performance: produtividade, lucratividade, qualidade. Fusões e aquisições também mudam as unidades operantes em um mercado.

Economistas mostraram que é comum existir entrada e saída de empresas dos mercados. Importante é que a redução do número de firmas pode implicar na redução na concorrência. Barreiras à entrada de novos competidores também pode implicar em menor concorrência, afetando diretamente lucros, produtividade e preços.

De forma geral, o problema relacionado a entrada e saída de empresas em um mercado é geralmente associado a mobilidade de recursos entre mercados. Algo que inibe o movimento de recursos entre mercado é denominado de barreira à entrada, ou, parafraseando o título do famoso livro de Joe S. Bain, barreira à nova competição (Bain, 1956).

Barreiras à entrada são interpretadas como estruturais, institucionais e até comportamentais, mas a definição do termo é controversa em economia. Bain (1956) definiu barreira à entrada como conjunto de tecnologia ou condições do produto que permitem a firma incumbente obter lucro no longo prazo. As condições são economias de escala, diferenciação de produtos e vantagens absolutas de custo. Stigler (1968) criticou esta abordagem, principalmente a ideia de que a economia de escala poderia ser barreira à entrada, oferecendo a ideia de que a barreira seria o gasto para produzir que a entrante enfrenta, mas não a incumbente. Veremos neste capítulo o papel dos custos, tecnologias

e demais fatores que afetam o número de competidores e medidas de performance de mercado.

## 8.1 Função Custo: Tecnologias

No capítulo 3 vimos os princípios de funcionamento da firma teórica em economia, mas com estruturas de custo e tecnologia simplificadas. Nesta seção definimos alguns conceitos tradicionais em economia que são úteis para analisar a dinâmica de mercado em que os custos podem afetar a performance e o número de concorrentes.

### 8.1.1 Economias de Escala

Começamos com um exemplo simples de função de produção com economia de escala. Dizemos que uma função de produção é homogênea de grau  $\theta$ :

$$y = f(\lambda x) = \lambda^\theta f(x) \quad (8.1)$$

Nesta equação  $x$  representa a quantidade de insumo produtivo,  $y$  a produção total e  $\lambda > 0$  é uma variável de nível tecnológico. Dizemos que a função tem rendimentos constantes se  $\theta = 1$ , e rendimentos crescentes se  $\theta > 1$ . Com a existência de economia de escala, o custo de duplicar a produção é menos do que duplicado.

Por sua vez, podemos relacionar esta definição à função custo. Neste caso a elasticidade do custo em relação ao nível de produção é menor do que 1. Exemplo: se fizermos  $\lambda = x$  e  $x = 1$ , temos  $y = x^\theta f(1)$ . Escrevendo está em equação na forma log-linear teremos  $\ln y = \theta \ln x + \ln f(1)$ . Logo, podemos definir a elasticidade do insumo em relação à produção como:

$$\epsilon_{xy} = \frac{\partial \ln x}{\partial \ln y} = \frac{1}{\theta}.$$

Agora suponha que o custo do fator produtivo  $x$  seja o preço constante  $w$ , então a elasticidade do custo total,  $C = wx$ , também será dada por  $1/\theta$ .

$$\epsilon_{Cy} = \frac{\partial \ln wx}{\partial \ln y} = \frac{1}{\theta}. \quad (8.2)$$

Assim, teremos economia de escala (rendimentos crescentes de escala) se e só se a elasticidade custo em relação ao nível da produção for menor do 1.

Uma interpretação útil é escrever a elasticidade do custo total como a razão entre o custo marginal ( $CMg$ ) e o custo médio ( $CMe$ ):

$$\epsilon_{Cy} = \frac{\partial C}{\partial y} \frac{y}{C} = \frac{CMg}{CMe}.$$

Desta forma, quando o custo médio for decrescente teremos economias de escala, ou ainda, quando o custo marginal for inferior ao custo médio.

Em trabalhos aplicados se utiliza com frequência o conceito de *Escala Mínima Eficiente*. A escala mínima eficiente é atingida quando a realização do CMe se aproxima do ponto de mínimo.

[inserir figura de EME]

Como determinar a EME? Estimar uma função custo.

Uma alternativa ao uso da estimativa da função de produção é o chamado teste de sobrevivência de Stigler. O teste segue a racionalidade descrita abaixo.

Num mercado concorrencial, o preço será fixado próximo ao mínimo do CMe. As empresas com custo médio superior ao preço não conseguirão sobreviver ao mercado. Portanto, a EME é o limite inferior de sobrevivência das empresas no mercado.

### 8.1.2 Economias de Escopo

Temos economia de escopo quando o custo de produção conjunto de um ou mais bens é inferior ao custo de produzir estes bens separadamente.

$$C(q_1, q_2) < C(q_1, 0) + C(0, q_2) \quad (8.3)$$

Exemplo clássico são serviços de transporte e telecomunicações.

### 8.1.3 Economias de Experiência

Economia de experiência também pode ser chamada de economia de escala dinâmica. Quando temos este caso  $CMe <$  experiência da firma (aprendizado).

A economia de experiência em geral é identificada quando a quantidade produzida ao longo do tempo reduz muito o custo produtivo. Exemplos clássicos são: asas de avião e indústria de semicondutores.

## 8.2 Tamanho do Mercado e Custo de Entrada

A questão de fundo aqui é determinar a relação entre tecnologia, tamanho do mercado e concentração da indústria. Vamos considerar um modelo simples em que todas as firmas são do mesmo tamanho. Com esta hipótese, determinar a concentração equivale a determinação do número de firmas no mercado.

Suponha uma firma genérica,  $j$ , que possua uma função custo dada por  $C = F + cq_j$  e uma função demanda dada por  $Q = (a - P)S$ . Neste modelo,  $S$  é o tamanho do mercado,  $P$  o preço,  $F$  o custo fixo,  $c$  o custo marginal, e  $C$  o custo total. Neste exemplo podemos permitir dois mercados iguais, mas com número diferentes de firmas. Este poderia ser o caso do mesmo produto sendo comercializado em dois países diferentes.

A função lucro da firma genérica é:

$$\pi_j = q_j(a - Q/S) - F - cq_j \quad (8.4)$$

Como as empresas são simétricas sabemos que  $Q = n.q_j$ , sendo  $n$  o número total de firmas. Solucionando este modelo Cournot-Nash, a condição de primeira ordem é:

$$a - \frac{2n}{S}q_j - c = 0 \quad (8.5)$$

A quantidade ótima para a firma representativa é:

$$q_j^* = \frac{a - c}{2} \frac{S}{n}. \quad (8.6)$$

Substituindo a quantidade ótima na função lucro, obtemos o lucro para cada firma que opera no mercado:

$$\pi_j(n) = S \left( \frac{a - c}{n + 1} \right)^2 - F. \quad (8.7)$$

Um equilíbrio livre de entrada é caracterizado por um conjunto de firmas ativas tal que (i) nenhuma firma ativa deseja deixar o mercado e (ii) nenhuma firma deseja entrar neste mercado. Esta análise simples fornece a condição de atividade de firmas no mercado. Ela implica que se a firma  $n$  tem lucro ela não deseja deixar o mercado, enquanto que se a possível entrante não tiver lucro, não ocorrerá entrada. As condições são as seguintes:

$$[C.1] \pi(\hat{n}) \geq 0, \text{ e}$$

$$[C.2] \pi(\hat{n} + 1) \leq 0.$$

Solucionando (8.7) para  $n$  para a condição de lucro igual a zero, temos:

$$\hat{n} = (a - c) \sqrt{\frac{S}{F}} - 1. \quad (8.8)$$

Esta equação nos diz que o número de firmas é uma função crescente do tamanho de mercado. Nesta estrutura o tamanho de mercado é descrito por  $S$  e de forma indireta por  $a$ , que é constante associada a função demanda.

Observe que a relação entre o número de firmas  $n$  não é proporcional ao tamanho do mercado. Para valores altos de  $\hat{n}$ , a relação entre número de firmas e tamanho do mercado é aproximadamente quadrática. Grosso modo, para dobrar o número de firmas é preciso aumentar o tamanho do mercado em quatro vezes.<sup>1</sup> Isto ocorre porque a medida que aumenta o número de firmas a concorrência é maior e a margem,  $p - c$ , é menor. Isto resulta em lucro variável por unidade de tamanho de mercado menor, que por sua vez o mercado sustenta menos firmas. *Devido a maior concorrência, o número de empresas ativas varia menos proporcionalmente com o tamanho do mercado.*

**Número de Firmas e Escala Mínima Eficiente:** Como dito anteriormente uma forma de mensurar retornos crescentes de escala é pela escala mínima eficiente. Assumindo a

<sup>1</sup>A relação é dada por  $\sqrt{2} - 1$ . Veja Cabral (2017), p. 254.



função custo usual  $C = F + cq$ , o custo médio  $CM = F/q + c$ . O mínimo do custo médio é  $c$ . Vamos fazer a EME ser a escala mínima tal que o custo médio seja  $CM = c'$ , neste caso  $c' - c = F/q$ , portanto a EME pode ser estimada como:

$$EME(q) = \frac{F}{c' - c}. \quad (8.9)$$

Podemos interpretar mudanças na EME como mudanças no custo no custo fixo ( $F$ ). Qualquer aumento em  $F$  aumenta a EME na mesma proporção. Observe que no modelo simples de mercado, equação (8.8), o número de firmas depende de  $F$ . Se a escala mínima eficiente em 2, o número de firmas diminui  $\sqrt{2}$ , aproximadamente. A intuição aqui é o inverso do aumento do mercado em 2. Portanto, em estudos empíricos é comum utilizar o tamanho do mercado dividido pela EME.

Uma alternativa a EME é utilizar o coeficiente de economia de escala é o coeficiente de economia de escala. Como vimos anteriormente a economia de escala é:  $\epsilon_{Cy} = CM/c$ , tal que  $CM$  é o custo médio e  $c$  o custo marginal. Se o custo médio é maior do que o custo marginal, então temos economia de escala.

Podemos esperar que uma indústria seja mais concentrada quanto maior for o grau de economia de escala.<sup>2</sup> Para a função custo que estamos considerando aqui,  $C = F + cq$ , temos:

$$\epsilon_{Cy} = \frac{F/q + c}{c} = 1 + \frac{F}{cq} \quad (8.10)$$

Em duas indústrias com custos fixos diferentes, a que tiver maior economia de escala será mais concentrada. Em geral, a concentração será maior quanto maior for o grau de economia de escala e a escala mínima eficiente.

Economia de escala e EME são barreiras à entrada no sentido apontado por Bain (1956). Neste sentido, quanto maior a barreira à entrada de novos competidores, maior será a concentração de mercado.

### 8.2.1 Barreiras Endógenas

Barreiras à entrada podem ser exógenas ou endógenas. O tamanho do mercado ou o custo fixo como vimos anteriormente são barreiras exógenas. A barreira endógena é função da ação de incumbentes para dificultar a entrada de novos competidores.

Barreiras endógenas podem ser tecnológicas, físicas ou mesmo burocráticas. Exemplo clássico de barreira endógena são dispêndios com publicidade. [Incluir texto]

<sup>2</sup> Pode ser mostrado que o custo médio é maior do que o marginal, somente se o custo médio é decrescente (em  $q$ ).

### 8.3 Entrada e Dinâmica de Mercado

Vimos anteriormente que a configuração de mercado que gera o máximo de bem-estar social é a concorrência perfeita. Neste caso, entrada livre de firmas garantiria a melhor alocação de recursos possível.

O principal resultado do modelo de concorrência perfeita é o que diz que nesta configuração é atingido o máximo de bem-estar social. Isto implica que com o máximo de concorrência os preços são mais baixos e é possível maior consumo de bens e serviços. Uma das condições para o resultado do teorema de alocação eficiente é a *livre entrada* de firmas. Se todas as demais condições são satisfeitas, a ausência de livre entrada implica em ineficiência alocativa. Isto é, se forem criadas barreiras à entrada, a alocação eficiente proporcionado pela configuração de concorrência perfeita pode deixar de ser alcançada. Um resultado teórico importante mostrado por alguns autores diz que se outras hipóteses da concorrência perfeita falham então a entrada livre (sem custos) não é mais desejável do ponto de vista da eficiência econômica. Este é o resultado do artigo de Mankiw e Whinston (1986).

Para entender este argumento assumamos margem positiva,  $p - c > 0$ , custos fixos estão presentes, o total das vendas da firma que entra é maior do que o aumento da produção total após a entrada. Em outras palavras, quando uma firma entra ela toma parte das vendas das incumbentes. Isto é chamado de *efeito de roubo dos negócios* devido à entrada.

Qual o tamanho do efeito roubo de negócios? Com produtos homogêneos, o volume de venda da indústria é  $Q_n$ , tal que  $n$  representa o número de firmas ativas. Se entra uma nova firma com produção  $q_{n+1}$ , parte desta produção é volume novo e parte é “roubo” de demanda de concorrentes. Assim a produção total da indústria será maior, mas o total de vendas da nova empresa é maior do que a variação de  $Q$ :  $Q_{n+1} - Q_n < q_{n+1}$ .

O benefício social da entrada é a margem multiplicado por  $(Q_{n+1} - Q_n)$  enquanto que o benefício privado (da entrante) é a margem multiplicado pelas vendas. Neste caso, para a mesma margem o benefício privado é maior do que o total. Como ocorre lucro com a entrada, novas firmas continuam a entrar no mercado. Mas qual razão para divergência entre incentivo privado e total (social)? A justificativa é que parte dos lucros da firma que entra é “roubada” das incumbentes.

Existem diversas aplicações explorando o resultado desta análise de entrada: transmissão de rádio (operadoras; Berry e Waldfogel, 1999), agentes de imobiliárias (Hsei e Moretti, 2003), índices de fundos (Hortacsu e Syverson, 2004), entre outros. Considere o caso da transmissão de rádio nos EUA, onde grande parte das rádios são temáticas, isto é, elas tem uma programação clara. Por exemplo, programação com foco em rock, música clássica, jazz, etc. A evidência empírica sugere que o principal efeito de abrir uma nova estação de rádio é desvio de audiência para a nova rádio. Neste caso o efeito de “roubo dos negócios” domina e ocorre porque o preço de trocar de rádio é ter que escutar anúncios. Como resultado deste processo de entrada, existirão muito rádios de rock, por

exemplo, em detrimento à outro tipo de gênero.

**Produtos diferenciados.** Quando se muda a hipótese de produto homogêneo para diferenciado, então a entrada também implica em redução de preços e aumento da variedade de produtos.

Com produto diferenciado, cada firma que entra fornece algum excedente para o consumidor. Como cada firma que entra não consegue capturar totalmente o excedente do consumidor, elas podem terminar não entrando mesmo que o ganho social seja positivo. Sendo assim, a existência de produto diferenciado com variedade na entrada, muda as conclusões obtidas para produto homogêneo.

Considere um modelo simples para descrever esta situação. Vamos representar uma função utilidade agregada como  $U$ , que captura o benefício que os consumidores tem com novas firmas provendo mais variedade:

$$U_i = U(g(q_1) + \dots + g(q_n)). \quad (8.11)$$

$U$  e  $g(\cdot)$  são funções côncavas, tal que os consumidores tem utilidade decrescente em cada produto novo ofertado pelas firmas. No mercado as firmas ofertam produtos ao preço  $p_i$ .

Os consumidores maximizam utilidade quando o preço for igual ao benefício (utilidade) marginal:  $p_i = U'g'(q_i)$ . Mankiw e Whinston (1986) colocam importância no custo fixo para explicar o resultado social da entrada. Portanto, a função custo adotada por eles é  $c(q) + F$ .

A medida de bem-estar é o excedente total, que é a soma do lucro e do excedente do consumidor. Após alguma manipulação é possível mostrar que o excedente total de existir  $n$  firmas operando é:

$$W(n) = U(n.g(q_n)) - n.c(q_n) - n.F. \quad (8.12)$$

Como estamos interessados no efeito da entrada de firmas, gostaríamos de avaliar a mudança no bem-estar,  $W()$ , quando muda o número de firmas. Portanto, tome a derivada parcial de  $W(n)$  em relação a  $n^3$

$$\frac{dW(n)}{dn} = U' \left[ g(q_n) + n.g \left( \frac{dq_n}{dn} \right) \right] - \left[ c(q_n) + n.c' \frac{dq_n}{dn} \right] - F. \quad (8.13)$$

Rearranjado para escrever o lucro da firma marginal, se tem:

$$\begin{aligned} \frac{\partial W(n)}{\partial n} &= [U'g'(q_n) \cdot q_n - c(q_n) - F] + \left[ n.U'g' - c' \frac{dq_n}{dn} \right] + \\ &+ [U'(f - f'q_n)]. \end{aligned} \quad (8.14)$$

<sup>3</sup>Uma formulação geral é:  $\frac{\partial W(n)}{\partial n} = U' \left[ g(q_n) + n. \frac{\partial g(q_n)}{\partial n} \right] - \left[ c(q_n) + n. \frac{\partial c(q_n)}{\partial n} \right] - F$

O primeiro termo da equação (8.14) é o lucro que será zero com entrada livre. O termo seguinte que está entre colchetes é o efeito “roubo dos negócios,” e está presente porque os preços excedem o custo marginal ( $c'$ ). O terceiro termo é positivo e contabiliza pelo excedente do consumidor pelo produto comercializado pela firma  $n$ . O segundo e terceiro termo tem efeitos opostos sobre a eficiência social. O efeito “roubo dos negócios” tem efeito negativo e é importante quando tem entrada em excesso de firmas. O outro termo tem efeito positivo devido a incapacidade de capturar o excedente do consumidor. Se este termo foi elevado teremos entrada insuficiente de firmas.

O efeito sobre o bem estar-social pode ser resumido da seguinte forma: se a diferenciação de produtos é muito importante então a entrada livre implica em entrada insuficiente do ponto de vista social. Por outro lado, se a diferenciação de produtos não é muito importante e a concorrência é suave, então o efeito “roubo dos negócios” domina, fazendo com que ocorra excessiva entrada.

Cabral (2017, p. 270-1) apresenta o seguinte exemplo extremo. Existe externalidade positiva da entrada no caso de medicamentos para doenças raras ou para doenças concentradas em população de baixa renda. Por razões regulatórias ou outras quaisquer, pode ser difícil para a empresa que vende o remédio escolher preço elevado que seja suficiente para recuperar o investimento. Neste caso, pode ocorrer da empresa nem entrar na venda do medicamento. Chamamos isto de “remédio órfão.” Por exemplo, por volta de 2005, cientistas do Canadá e dos EUA desenvolveram uma vacina contra o vírus do Ebola. Entretanto, o projeto permaneceu na prateleira porque existiam poucos casos da doença registrados nos países ricos (até 2014).

## **8.4 Dinâmica da Estrutura de Mercado: Destruição Criadora e Alocação Eficiente**

Na economia de mercado ocorre o processo constante de nascimento e morte de firmas. Às vezes a mortalidade de firmas está associada a obsolescência de produtos e a incapacidade de se manter eficiente em produzir bens e serviços. No mercado, constantemente as firmas são desafiadas por outros produtores utilizando técnicas mais eficientes ou introduzindo novos produtos ou serviços.

O famoso professor da Universidade Harvard, Joseph Schumpeter (1942), denominou este processo de *destruição criadora*, que era para ele o fato essencial sobre o capitalismo. O processo de destruição criadora se refere ao mecanismo de incessante inovação de produtos e processos em que a nova produção substitui a antiga.

O processo de destruição criadora também pode ser interpretado como reestruturação dos negócios e permeia os maiores aspectos da performance micro e macroeconômica. Em se tratando de performance das firmas, este processo é caracterizado por incontáveis decisões de criar e destruir arranjos produtivos. A eficiência destas decisões não depende apenas do talento gerencial, mas também das instituições que criem ambiente econômico adequado a este processo. Falhas nas dimensões institucionais, gerenciais e de novas ideias

tem severas consequências uma vez que elas interagem com o processo de destruição criadora.<sup>4</sup>

Algumas dessas limitações são naturais e tem origem na tecnologia, como por exemplo, a necessidade custo fixo elevado na produção de aço. Outras restrições são devido à configuração concorrencial dos mercados e a ambiente econômico propiciado pelas instituições. Por exemplo, em doses moderadas, restrições institucionais do ambiente econômico fomentam o padrão de comportamento cíclico da maioria das economias avançadas. Eles também ajudam a explicar problemas econômicos perenes tais como comportamento cíclico do mercado de trabalho, do investimento e dos salários. Em doses elevadas, o ambiente econômico adverso pode limitar a adoção de novas tecnologias e oportunidades, podendo resultar em mercado de trabalho disfuncional, má-alocação de recursos e estagnação econômica (Caballero, 2006).

Concorrência internacional é uma fonte importante de destruição criadora. Por exemplo, Bernard, Jensen e Schott (2006) mostraram que o crescimento da produtividade nos EUA é mais rápido em indústrias em que os custos comerciais foram reduzidos ao longo do tempo.

### **8.4.1 Estrutura de Mercado e Produtividade**

O ambiente econômico de um mercado é caracterizado pelas firmas que entram, saem e continuam, bem como pela tecnologia produtiva e arcabouço institucional. Na parte institucional estão tributação, regulação (técnica) não-tributária e estrutura legal. Assim como tecnologia, o arcabouço institucional pode produzir barreiras à competição.

A evidência empírica sobre o tema mostra que entrada e saída são importantes para o crescimento da produtividade. Firms menos eficientes deixam o mercado e, eventualmente, firmas que entram e se tornam mais eficientes ganham tamanho. Um exemplo muito importante deste caso foi a desregulamentação do mercado doméstico norte-americano de tele-equipamentos. Este caso foi estudado por Olley e Pakes (1996) e é fundamental para entender a análise empírica moderna de dinâmica de mercados.

**Detalhes da Indústria.** Maior reestruturação desta indústria começou no final dos anos 60. Esta mudança é devido a dois fatores relacionados. Por um lado ocorreu mudança técnica nos equipamentos, com muitos produtos novos tanto para serviços telefônicos (fibra ótica, centrais digitais, etc.) e para o uso da linha (fax, modem, etc.). Por outro lado veio a desregulamentação da produção e comercialização de equipamentos, que naturalmente foi impulsionado pela mudança técnica.

A AT&T possuía monopólio nos serviços de telecomunicações e monopólio essencial em equipamentos por meio de sua subsidiária, Western Electric. O monopólio em equipamentos era mantido pelo controle do processo de licitação para a rede única da AT&T, fazendo com que a Western Electric fosse virtualmente a única ofertante de equipamentos

<sup>4</sup>Veja Caballero e Hammour (1996) e Caballero (2006).

para roteamento e chamadas telefônicas. Virtualmente, a única possibilidade de entrada seria possuir uma rede própria. Todavia, em muitos estados isso seria ilegal.

A desregulamentação começou com a decisão do FCC permitindo que outros dispositivos fossem conectados à rede da AT&T. Esta decisão de 1968 é chamada de decisão Carterphone, pois permitia que este tipo de aparelho fosse conectado a rede de telefonia. Isto abriu o mercado para novas empresas e novos produtos, tais como modems, fax, centrais PBX, etc. Entretanto, grandes equipamentos ainda continuavam com a Western Electric.

A situação apenas mudou com a quebra do monopólio da AT&T em 1982. Por meio de acordo judicial, ocorreu a separação do serviço de telefonia local da longa distância, que permaneceu com a AT&T.<sup>5</sup> O serviço de telefonia local passou a ser ofertado por nove empresas locais, as *Baby Bells*,<sup>6</sup> sendo que nenhuma delas poderia ser proprietária de manufatura de equipamentos.

A medida também obrigou a AT&T a abrir a concorrência na chamada de longa distância para outras operadoras. Este ponto foi muito importante na disputa judicial, pois as empresas que poderiam conectar cidades teriam que ter acesso a rede da incumbente. Grande parte do argumento da AT&T para não abrir o mercado local era devido ao fato dele ser tratado como um bem público, onde a empresa usava as altas margens da longa distância para sustentar as redes locais mais acessíveis.

O acordo judicial de 1982 teve impacto imediato sobre a indústria de tele-equipamentos. As compras de grandes equipamentos passaram a ser de empresas do exterior, que ou construíram plantas nos EUA ou exportavam produtos. Em três anos o total de equipamento importado aumentou de 2% para 14% do total. Entre 1963 e 1987 o número de firmas aumentou de 107 para 481, com volume grande de entrada e saída de firmas e plantas.

**Relação entre tamanho e produtividade.** O objetivo do trabalho de Olley e Pakes (1996) foi analisar as mudanças na produtividade nesta indústria em função da desregulamentação (desverticalização). Após estimar a função de produção com método inovador, eles calculam a produtividade total dos fatores (TFP, veja capítulos 3 e 12) e perguntam quais fatores podem explicar o crescimento da produtividade. Obviamente tal crescimento está relacionado com a entrada de novos competidores, mas como isso poderia ser endereçado empiricamente?

O grande volume de entrada e saída de plantas no mercado americano, contribuiu para a *realocação* de capacidade e produto das unidades menos eficientes para as mais eficientes.

<sup>5</sup>Este caso começou com um processo da divisão de antitruste do Departamento de Justiça (DoJ), em 1974, contra a AT&T, em grande parte motivada pela verticalização entre provedor de serviços e fabricante de equipamentos (Western Electric). O acordo passou a vigorar a partir de 1984, reduzindo o valor da AT&T em 70%, aproximadamente. Para mais detalhes veja Coll (1986).

<sup>6</sup>As empresas independentes eram as Regional Bell Operating Companies (RBOC): Ameritech, Bell Atlantic, Bell South, Cincinnati Bell, NYNEX, Pacific Telesis, Southern New England Telephone, Southwestern Bell Corporation, e US West.

Para entender se a realocação é importante para o crescimento de produtividade setor, Olley e Pakes decompõem a produtividade em uma forma inteligente.

A medida de produtividade da indústria é a soma da produtividade de cada planta ponderada pelo tamanho proporcional de cada uma no total:

$$\omega_t = \sum_{j=1}^{N_t} s_{jt} \omega_{jt},$$

aqui com  $\omega_t$  definido como a produtividade total dos fatores da indústria no período  $t$ ,  $\omega_{jt}$  a produtividade de cada planta  $j$  em  $t$ ,  $s_{jt}$  é o tamanho proporcional na produção (market share) em  $t$ . Portanto, a produtividade da indústria em  $t$  é a soma da produtividade individual das  $N_t$  plantas, ponderada por  $s_{jt}$ .

A produtividade do setor pode ser decomposta entre a produtividade média por planta e a covariância entre produtividade e market share:

$$\omega_t = \bar{\omega}_t + \text{Cov}(\omega_{jt}, s_{jt}), \quad (8.15)$$

com  $\bar{\omega}_t$  representando a produtividade média e  $\text{Cov}(\omega_{jt}, s_{jt})$  a covariância entre tamanho e produtividade.<sup>7</sup> Este termo de covariância às vezes é denominado de termo de covariância Olley-Pakes.

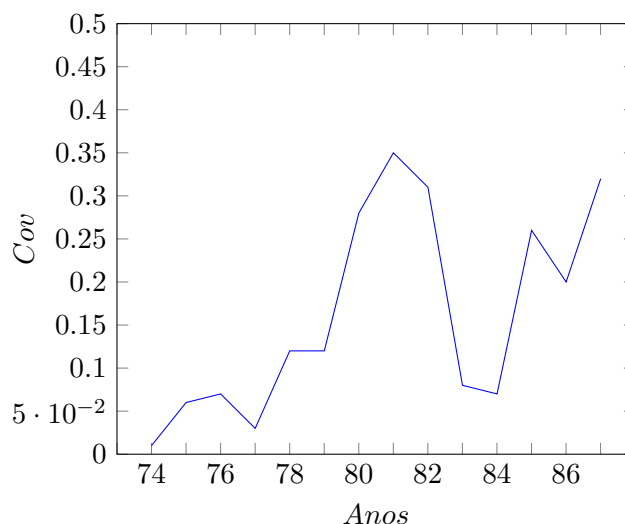
Na Figura 8.1 é apresentado o termo de covariância estimado por Olley e Pakes para a indústria de tele-equipamentos dos EUA.

**Importante: unidade de medida.** Muito importante na correta interpretação dos resultados é a unidade de medida da produtividade e da participação de mercado. Ambas as medidas devem ser livres de preços das firmas. A explicação é simples e direta: empresas com maior poder de mercado tem maior capacidade de determinar preços elevados. Neste sentido, uma empresa maior pode cobrar mais caro e o preço maior pode fazer parte da produtividade e do market share se ambos forem medidos sem eliminar o efeito-preço. Com a medida que não controla pelos preços, a firma que tem poder para colocar preço mais elevado pode erroneamente parecer mais produtiva e com participação de mercado, indicando melhor alocação eficiente. Todavia, aqui o interesse é saber se as firmas ganham mercado por eficiência e não pelo poder de mercado (precificação).

Como discutido anteriormente, esta é a distinção entre função geradora de vendas e função de produção. Olley e Pakes estimam o modelo utilizando uma função geradora de vendas, que não controla pelo efeito preço do poder de mercado. Um exemplo em que se estima uma função de produção controlado pelo efeito-preço é dada a seguir no caso da indústria de aço nos EUA.

<sup>7</sup> $\text{Cov}(\omega_{jt}, s_{jt}) = \sum_{j=1}^{N_t} (s_{jt} - \bar{s}_t)(\omega_{jt} - \bar{\omega}_t)$ , com  $\bar{s}_t$  como o market share médio do setor. Para Bartelsman, Haltiwanger e Scarpetta (2009) é mais simple calcular a covariância tamanho/produtividade fazendo  $\text{Cov}(\omega_{jt}, s_{jt}) = \omega_t - \bar{\omega}_t$ .

Figura 8.1: Decomposição da Produtividade: Covariância entre Tamanho e Produtividade (1974-1987)



**Indústria de Aço nos EUA.** Outro exemplo da relação entre entrada e crescimento da produtividade é a indústria de aço nos Estados Unidos (Collard-Wexler e De Loecker, 2015). Entre 1972 e 2002, a produtividade total dos fatores no setor aumentou em 28%, enquanto o crescimento médio da economia foi de 7%. Durante esse período, a produtividade do setor de aço foi a segunda maior da economia americana, ficando atrás apenas do setor de tecnologia da informação. Este caso é particularmente interessante porque a produção em 2002 possuía aproximadamente o mesmo nível de 1972, implicando que a performance de produtividade deve ser explicada pelo uso de fatores de produção. Neste caso, o emprego durante este período diminuiu 80%.

Collard-Wexler e De Loecker mostram que a explicação para grande parte do crescimento de produtividade foi a entrada de novas firmas com nova tecnologia. São empresas que operam *mini-mills*.

As *mini-mills* não apenas tomaram espaço no mercado das menos eficientes, mas elas também apresentaram ganhos de produtividade durante o período. Conjuntamente estes efeitos contabilizaram por, aproximadamente, metade de ganho de produtividade no período analisado. Na Tabela 8.1 estão os resultados estimados por Collard-Wexler e De Loecker (2015) para a covariância entre produtividade e tamanho. Eles mostram que a realocação para as plantas mais produtivas foi um processo muito importante de aumento da produtividade, contribuindo com 6,4 por cento do total de 22,1. Isto significa que um terço do aumento da produtividade foi devido a realocação da produção das plantas menos eficientes para as mais produtivas.

A decomposição da produtividade também é calculada para as plantas segmentadas pelo tipo de tecnologia: verticalmente integradas ou *mini-mills*. As plantas *mini-mills*



Tabela 8.1: Termo de Covariância Olley-Pakes para Setor de Aço (EUA): 1963-2002

	$\Delta$ Produtividade (TFP)	$\Delta\text{Cov}(\omega_{jt}, s_{jt})$
Todas plantas	22,1	6,4
Integradas	24,3	3,7
<i>Mini-Mills</i>	9,6	4,4

Fonte: Collard-Wexler e De Loecker (2015). Variação ( $\Delta$ ) da produtividade está em percentual.

aumentaram a produtividade em 10% e as integradas em 24,3. Para as ambas as tecnologias, o termo de covariância foi significativo, 3,7 e 4,4 para integradas e mini-mills, respectivamente. O que implica em bastante realocação de produção entre as mais e menos produtivas.

Como mostrado na definição da covariância Olley-Pakes, a produtividade pode ser decomposta entre o termo de covariância. Isto implica que maior parte do crescimento de produtividade das plantas integradas vem do crescimento médio e não da realocação. Embora grande parte do crescimento venha das plantas com a tecnologia antiga, as novas plantas forçaram a mudança das plantas que sobreviveram a esta mudança tecnológica. Neste sentido, a melhora da produtividade das incumbentes é uma *reação* à entrada de competidores com nova tecnologia. Pode ser interpretado que a entrada das mini-mills levaram ao aumento de 44 por cento da produtividade do setor. Deste número, 23 por cento é devido a realocação das plantas integradas para as mini-mills e 21 por cento devido a melhora de produtividade das plantas mini-mills.<sup>8</sup>

No geral, mecanismos de realocação entre produtores heterogêneos tem sido utilizados em diversos campos, além de economia industrial, comércio internacional e macroeconomia também lideram o uso deste tipo de análise. Estes modelos diferem na especificação, mas a intuição econômica é a mesma: devido a maior concorrência é mais difícil para as empresas de custos mais elevados conseguir lucros mais elevados, uma vez que o lado da demanda é sensível a diferenças de preços e custos entre firmas. A medida que a substitubilidade aumenta, compras são realocadas para fornecedores mais produtivos, aumentando a correlação entre produtividade e tamanho.<sup>9</sup>

<sup>8</sup>Este aumento da produtividade destas plantas pode ser devido a mudança técnica (novos equipamentos e processos) e a *learning by doing*.

<sup>9</sup>Veja Chandra, Filkstein, Sacarny e Syverson (2016).



## Capítulo 9

# Discriminação de Preços e Outras Estruturas de Demanda

### 9.1 Introdução

Tipos de discriminação de preços. Seleção por indicadores. Auto-seleção.

### 9.2 Seleção por Indicadores

Seleção por indicadores corresponde a situação em que o vendedor divide os compradores em grupos, determinando o preço diferente para cada grupo. Esta prática também é denominado de *segmentação de mercado*.

Uma segmentação clássica é a localização geográfica do consumidor. Neste caso a firma cobra preços diferentes de acordo com a localização dos consumidores. Por exemplo, preço diferente por país em produtos exportados: veja produtos da Apple, Microsoft, etc.

O modelo mais simples para este caso seria o do monopolista que vende o mesmo produto para dois mercados diferentes  $j = \{A, B\}$ . Este problema é similar ao caso do monopolista multi-produto, seção 4.1.2 (sem canibalização no mercado e com custo comum). A função lucro do monopolista é

$$\pi(p_A, p_B) = \{p_A \cdot q_A(p_A) + p_B \cdot q_B(p_A)\} - C(q_A(p_A) + q_B(p_A)) \quad (9.1)$$

tal que  $p_j$  é o preço no mercado  $j$ ,  $q_j$  a quantidade demandada, e  $C(\cdot)$  a função custo que depende da demanda.

Naturalmente a maximização implica que  $RM_A = RM_B = c$ , para  $RM_j$  como receita marginal em  $j$  e  $c$  o custo marginal que é comum aos mercados – uma vez que apenas se

muda o mercado geográfico. Isto implica na regra de elasticidade para os dois mercados:

$$p_A \left(1 + \frac{1}{\epsilon_A}\right) = p_B \left(1 + \frac{1}{\epsilon_B}\right) = c, \quad (9.2)$$

tal que  $\epsilon_j = \frac{\partial q_j}{\partial p_j} \frac{p_j}{q_j}$  é a elasticidade-preço da demanda.

Neste sentido, sob discriminação de preços com segmentação de mercado, um vendedor deve cobrar o preço mais baixo nos segmentos com maior elasticidade de preço. Por exemplo, vinhos chilenos podem custar mais caro em New York do que no Chile devido a elasticidade da demanda. Este caso é bem conhecido na literatura de comércio internacional, chamado de *viés doméstico*, pois o mercado local tende a ser menor do que o mercado de destino das exportações.

**Exemplo:** Um outro exemplo interessante é a discriminação de preços no mercado europeu de carros (Verboven, 1996). Na Tabela 9.1 são apresentadas estimativas dos markups (%) de alguns modelos em países selecionados da Europa.

Tabela 9.1: Markup Relativo Selecionado de Carros em Países Europeus (%)

Modelo	Bélgica	França	Alemanha	Itália	R. Unido
Fiat Uno	7,6	8,7	9,8	21,7	8,7
Nissan Micra	8,1	23,1	8,9	36,1	12,5
Ford Escort	8,5	9,5	8,9	8,9	11,5
Peugeot 405	9,9	13,4	10,2	9,9	11,6
Mercedes 190	14,3	14,4	17,2	15,6	12,3

Fonte: Verboven (1996).

As diferenças de margens podem ter algumas interpretações. A primeira poderia ser a discriminação espacial dos consumidores. A segunda, quotas de importação de veículos poderiam influenciar muito as margens, por fim, poderia existir diferença do nível de consumo entre países. Todavia, a evidência econométrica apresentada por Verboven (1996) sugere que a discriminação de preços seria a melhor explicação para a diferença de margens.<sup>1</sup> A discriminação de preços se manifesta pois mercados diferentes possuem elasticidades específicas e os produtores escolhem o preço de acordo com a demanda dos consumidores. No caso dos mercados serem países, um padrão fica muito claro: as margens são maiores no país produtor. Por exemplo, a maior margem do Fiat Uno é na Itália. Neste caso, como a maior margem é observada na Itália, a demanda pelo Fiat Uno é mais inelástica neste mercado do que nos demais. Portanto, o *viés doméstico* é refletido na demanda menos elástica (na ausência de outras explicações).

Por outro lado, observe que a margem do Nissan Micra, um carro de uma montadora japonesa, é muito mais elevada na França e na Itália. Qual poderia ser a explicação?

<sup>1</sup>Cabe ressaltar que várias configurações poderiam fornecer explicações da diferença de margens.

Verboven explica que isto se deve a quotas de importação elevadas que são aplicadas pela França e Itália, restringindo a oferta destes veículos.

### Preço ótimo por segmento

Como vimos anteriormente, a regra da elasticidade de determinação do markup é útil quando a elasticidade da demanda é constante em todos os pontos da curva de demanda. Se a elasticidade da demanda não é constante, então temos que solucionar “completamente” o problema de maximização de lucros para encontrar os preços ótimos. Neste caso, temos que encontrar o preço único ótimo bem como o preço ótimo por segmento de mercado. Para mostrar como isto pode ser feito seguimos o exemplo apresentado por Cabral (2017, p.127) sobre remédio relacionado a ansiedade.<sup>2</sup>

A startup fictícia BioGar desenvolveu Xamoff, um remédio “de marca” que reduz ansiedade, cuja droga permanece sob proteção de patente. BioGar planeja entrar no mercado europeu, mas pensa se deve cobrar os mesmo preço no mercado doméstico e no novo mercado. Vamos começar a análise assumindo que a curva de demanda tem a seguinte forma:

$$q_i = a_i - b_i p_i, \quad (9.3)$$

tal que  $i$  é o índice de mercado e  $a_i$  e  $b_i$  são os parâmetros de demanda que podem variar em  $i$ .

Por exemplo, considere que no mercado americano ( $i = 1$ ) os parâmetros são  $a_1 = 12$  e  $b_1 = 2$ . No mercado europeu ( $i = 2$ ), os parâmetros são  $a_2 = 4$  e  $b_2 = 1$  (demanda perfeitamente elástica). Assuma que o custo marginal (igual ao custo médio) é  $c = 1$ , assumindo que todas as unidades estão em milhões de dólares.

Quanto que a empresa pode ganhar operando preços diferentes nos dois mercados, EUA e Europa? Vamos considerar primeiro o caso de preço uniforme. A demanda total cobrando o mesmo preço,  $p_1 = p_2 = p$ , é:

$$Q = q_1 + q_2 = (a_1 + a_2) - (b_1 + b_2)p = A - Bp, \quad (9.4)$$

com  $A = a_1 + a_2$  e  $B = b_1 + b_2$  por simplicidade. A demanda inversa é:  $p = (A - Q)/B$ . A função lucro para o produto total  $Q$ , tratando os dois mercados como um, é

$$\pi(Q) = \frac{A - Q}{B} Q - cQ, \quad (9.5)$$

novamente adotamos  $c$  como custo marginal constante. O produto ótimo para  $Q$  é

$$Q = \frac{A - cB}{2}, \quad (9.6)$$

solucionando para os valores definidos anteriormente, a quantidade ótima é  $Q = 6,5$ . Substituindo o volume na função demanda inversa, temos  $p = (A - Q)/B = 3,17$ . Como

<sup>2</sup>Veja também em [https://pages.stern.nyu.edu/~dbackus/1303/notes\\_prdisc.pdf](https://pages.stern.nyu.edu/~dbackus/1303/notes_prdisc.pdf).

cada mercado tem uma função demanda específica, podemos retornar este preço para a função demanda (9.3) de cada mercado. Neste caso, os volumes serão  $q_1 = 5,67$  e  $q_2 = 0,83$  e o lucro da firma é 14,08.

Como os mercados possuem funções de demanda locais, podemos calcular os preços para os dois mercados. Nesta estratégia vamos supor que não é possível realizar transações do mercado com o produto mais barato para o mais caro. Para encontrar as quantidades ótimas, voltamos ao problema similar ao monopolista multi-produto, mas aqui cada mercado tem a sua própria função demanda. A função lucro é:

$$\pi(q_1, q_2) = \frac{a_1 - q_1}{b_1} q_1 + \frac{a_2 - q_2}{b_2} q_2 - c(q_1 + q_2). \quad (9.7)$$

Como no caso do monopolista multi-produto calculamos a quantidade ótima separadamente para cada produto. O resultado para cada produto  $i$  é:

$$q_i = \frac{a_i - cb_i}{2}. \quad (9.8)$$

Que implica em  $q_1 = 5$  e  $q_2 = 1.5$  no presente exemplo. Os preços são  $p_1 = 3.5$  e  $p_2 = 2.5$ , respectivamente resultado no lucro de 14.75. Isto implica que preços diferentes, para demandas diferentes, aumenta o lucro de 14.08 para 14.75, aumento de 4.76%.

A generalização deste procedimento funciona a medida que pode ser possível segmentar a demanda em mercados menores.

## 9.3 Auto-Seleção

### 9.3.1 Discriminação Intertemporal

## 9.4 Preço Não-Linear

O varejista escolhe  $p$  sob tarifa linear:  $T(q) = p^w \cdot q$ . Isto é, o problema de dupla margem surge quando a precificação é linear. O timing é o seguinte:

1. Manufatura escolhe a tarifa linear
2. Varejista escolhe o preço final para o consumidor

O lucro do varejista é:

$$\pi_R = (p - p^w)D(p) \quad (9.9)$$

A FOC é

$$\frac{\partial D(p)}{\partial p} (p - p^w) + D(p) = 0$$

se  $D(p) = 1 - p$  então  $\frac{\partial D(p)}{\partial p} = -1$ . Neste caso, podemos calcular  $p$  a partir da FOC:  $p = (p^w + 1)/2$ . No caso geral a solução é:

$$p = -\frac{D(p)}{[\partial D(p)/\partial p]} + p^w \quad (9.10)$$

## **9.5 Publicidade**

Publicidade é um componente importante da estratégia das firmas. Entre os tópicos de interesse estão a natureza dos efeitos da publicidade e a interação da publicidade com concorrência e estrutura de mercado.

## **9.6 Leilões**

## **9.7 Inteligência Artificial**

Atualmente muitas firmas tem utilizado algoritmos para precificação que exploram dados detalhados de preferências dos consumidores. Em algumas circunstâncias, as firmas usam processos computacionais de aprendizagem para desenvolver estratégias para atuação concorrencial em jogos de precificação em situação de oligopólio. Algoritmos de inteligência artificial (AIA) fornecem um arcabouço proeminente para a implementação de precificação. AIAs operam através do aprendizado do retorno obtido com ações possíveis (preços), para em seguida tomar as ações que geram o melhor retorno. Esse processo integrado é central no processo de aplicação de inteligência artificial para preços.





## Capítulo 10

# Fusões Horizontais

O propósito das leis antitruste é de estabelecer controle de como as firmas adquirem e mantêm sua posição de mercado. O tópico geral sobre antitruste também trata de comportamento colusivo, práticas de abuso de posição dominante, como discriminação de preços e barreira a acesso à mercados, relações verticais, além de fusões horizontais que vamos estudar neste capítulo.

### 10.1 História

Ao final dos anos de 1800 e começo dos 1900, os EUA viram o crescimento de empresas grandes por meio de fusões. Exemplos são General Electric, Standard Oil e US Steel. Em termos de proporção do PIB americano, esta foi a maior onda de fusão na história americana.

Em 1890 foi instituído o *Sherman Act*, que foi uma resposta a este comportamento, sendo a primeira legislação antitruste federal. Tanto pelos termos da lei quanto pela interpretação das cortes, o *Sherman Act* deixou muitas dúvidas sobre o que seria permitido. Podemos dizer que o desejo de deixar claro o dispositivo do *Sherman Act*, foram aprovados em 1914 o *Clayton Act* e o *Federal Trade Commission Act*.

Na seção 1 do *Sherman Act* é explicitamente proibido cartel. Na seção 2 parece proibir monopólio, entretanto as cortes interpretaram que monopólio não era proibido, desde que não tenha cometido nenhuma conduta.

O *Clayton Act* foi endereçado contra quatro práticas distintas e estabelece reparação de danos causado pelas condutas. A seção 2 previne discriminação de preços que suaviza a concorrência. Seção 3 previne venda-casada e exclusividades que reduzem concorrência. Seção 7, proibiu fusões cujo efeito podem ser “substantially to lessen competition, or tend to create a monopoly” (em tradução livre: cujos efeitos substancialmente restringem a concorrência ou tendem a criar um monopólio). Seção 8 trata de “interlocking directorates among firms.”

O *FTC Act* criou a agência do governo federal, cuja missão é aplicar as leis do anti-truste e lidar com disputas. A seção 5 proíbe “métodos injustos de concorrência.” Outras responsabilidades do FTC são proteção do consumidor e prevenção de propaganda enganosa.

**Aplicação.** Todas as leis antitruste são aplicadas tanto pelo FTC quanto pelo DOJ. Um caso conduzido pelo DOJ é submetido ao sistema americano de cortes federais.

As ações do FTC são decididos por um juiz administrativo do FTC e são revisados pelos conselheiros do FTC. O apelo das decisões do FTC são revisados por tribunais federais.

**Brasil.** O sistema brasileiro de defesa da concorrência é similar ao FTC, onde existe uma corte administrativa e cujos recursos são apresentados ao tribunais federais (principalmente ao TRF1).

### 10.1.1 Fusões e Aquisições

Sob o *Hart-Scott-Rodino (HSR) Act* de 1976, as firmas passaram a fazer a pré-notificação de fusão/aquisição. Empresa que está adquirindo e possui ativos superior a 100 milhões USD, devem informar todas as aquisições maiores do que 15 milhões ao FTC e ao DOJ. Uma fusão acima destes limites deve obter a permissão destas agências, que avaliam os efeitos sobre o ambiente concorrencial.

Tanto o FTC ou DOJ podem decidir pela injeção preliminar. Antes da decisão é possível ocorrer significativa barganha entre as partes envolvidas na fusão e a autoridades anti-truste. Por exemplo, se existir um problema com um produto específico a firma que está adquirindo poderia vender esta operação para reduzir as preocupações concorrenciais do mercado. Na jurisdição americana, se o FTC/DOJ decidem contestar a operação, as firmas precisam decidir se querem lutar na justiça para manter a fusão ou simplesmente desistir da operação. A última palavra sobre a manutenção da operação será dada pela justiça federal.<sup>1</sup>

No Brasil, a análise prévia de fusões somente foi instituída na lei 12.529 de 2011. No artigo 88, “os atos que se subsumirem ao dispositivo no caput desse artigo não podem ser consumados antes de apreciados [...] sob pena de nulidade.” Além disso, “ Até a decisão final sobre a operação, deverão ser preservadas as condições de concorrência entre as empresas envolvidas.”

## 10.2 Análise de Fusão

A maioria das análises de fusões/aquisições começa pela definição de mercado e o subseqüente cálculo das participações de mercado (*market shares*). A premissa é que devemos

<sup>1</sup>Um caso contestado pelo DOJ é enviado para o sistema federal americano. Por sua vez, as ações do FTC são julgado por um juiz do FTC e então revisao pelo tribunal do mesmo FTC. O apelo às decisões tomadas pelo FTC são feitas para as cortes federais.

nos preocupar apenas se os *market shares* indicam que a concentração é grande o suficiente para sustentar *markups* injustos. Geralmente, o nível de concentração depende da “definição” de mercado; e muito do debate ocorre em como definir o mercado. Após a definição do mercado é quando pode ser realizada uma investigação séria sobre os *markups* e sobre o que seria “lucratividade injusta.”

Existe uma literatura grande que mostra que alta concentração pode ocorrer sem um grande apelo para a diferença preço-custo marginal, e as autoridades em diversas jurisdições estão cientes disso. Portanto, na análise fusões/aquisições é importante analisar as concentrações de mercado, bem como preços, custos e margens. Se os preços são observados, então as autoridades naturalmente buscam dados sobre custos. Um problema que as agências encontram é a confiabilidade nas informações de custos. Portanto, muitas vezes são utilizados os métodos de organização industrial para inferir custos marginais e as margens implícitas. Por exemplo, assuma a fórmula de determinação de preços de equilíbrio de Nash, estime a elasticidade da demanda (inversa), e use ela para determinar o *markup* (assumindo algo sobre a relação entre custo marginal e produto). Este roteiro é simples, mas existem diversos problemas/questões que precisam ser tratados.

Entre as questões estão: em primeiro lugar todos os problemas associados a estimação de demanda, incluindo elasticidade de demanda correta e definição de mercado relevante. Em relação a esta última questão precisamos saber que produtos serão “colocados” no mesmo mercado. Se incluirmos todos os produtos substitutos possíveis iremos rapidamente para um problema de muitos parâmetros e com potencial de elasticidade mais elevada. Outro problema associado a definição de mercado é a de que área geográfica podemos incluir no mesmo mercado.

Em segundo lugar, a hipótese de equilíbrio estático “Nash-em-preços” ou em quantidades pode fazer sentido antes da fusão, mas pode deixar de fazer sentido se

- (a) preços são difíceis de ajustar, ou se for uma indústria de bens homogêneos. Neste último caso, podemos querer substituir uma solução Nash-em-quantidades e explicar como isto é diferente;
- (b) Existem dinâmicas tanto do lado da oferta quanto na demanda. Na lado da oferta podem ser dinâmicas associadas a learning by doing ou questões de compatibilidade (padrões, sistemas, etc.). No lado da demanda podem existir dinâmicas associadas a mercados com “bens de experiência” (remédios), e em mercados em que o consumidor aprende a usar o bem e deve reaprender para outro produto (caso do teclado Qwert). Observe que a fórmula de determinação de preços deve ter termos novos para a mudança de lucro futuro resultante da mudança de preço ou quantidade atual. Neste caso, devemos introduzir o conceito de maximizar o valor descontado do fluxo de caixa na análise:

$$V(\cdot) = q_i(\cdot)_t [p_t - CMg_t] + \sum_{\tau=0}^{\infty} [q_i(\cdot)_{t+\tau} (p_{t+\tau} - CMg_{t+\tau})]$$

No caso da produção com externalidades, podemos pensar que o *markup* deveria

ser mais baixo do que o previsto por uma solução estática.

- (c) a noção de equilíbrio prévio pode não fazer sentido se existia um equilíbrio colusivo (mesmo que implícito). É preciso explicar como acordos de conluio implícito podem ser suportados. Considere tanto o modelo sem entrada bem como o com entrada.

Esta forma de análise também supõe que nada mais muda em resposta a fusão além dos preços. A hipótese é a de que funções custo não mudam após a fusão, e muitas vezes o argumento para a fusão, apresentado pelas partes, é a de que a operação irá permitir que as firmas cortem custos de forma significativa. O termo usado no mercado é o ganho de *sinergias* proveniente da fusão. A outra hipótese é de que tanto entrada, bem como outros investimentos irão mudar. Ou seja, se a operação for ou não aprovada isto não tem impacto sobre investimentos ou incentivos de entrada.

A análise nas diversas jurisdições é norteadada pelos guias de fusões. No Brasil é o guia para análise de atos de concentração horizontal. De uma forma ou de outra, quase todas as jurisdições seguem o *merger guidelines* dos EUA, que começou a ser publicado em 1968.

Os guias apresentam um procedimento de cinco passos para determinar quando se deve “impugnar” uma fusão:

1. Definição de mercado relevante;
2. Medida e critério de concentração;
3. Efeitos adversos potenciais sobre preços;
4. Entrada;
5. Análise de eficiência e falência da firma.

### 10.2.1 Definição de Mercado Relevante

A análise começa pela definição de produtos relevantes para cada uma das firmas envolvidas na fusão/aquisição. Para cada produto ou serviço das firmas, a agência antitruste busca definir o mercado geográfico e de produto. Supostamente, as definições são baseadas apenas nos padrões de substituição. Uma vez definido, o mercado deve ser medido em termos de seus participantes e concentração.

Como participantes do mercado incluem firmas que negociam produtos na área geográfica relevante. Além destas, outras firmas que pode ofertar produtos devido a um aumento nos preços pequeno mas significativo e não-transitório. Isto é, se estas firmas são ofertantes que podem desviar produção rapidamente para atender o mercado. Entrantes em potencial que incorram em custos afundados ou tomem mais tempo para entrar, são considerados no estágio quatro. Portanto, firmas que podem desviar produção para o mercado em análise devem ser considerados.

Observe que na maioria dos casos é difícil determinar os custos afundados com precisão. Geralmente, estas são as variáveis mais difíceis de medir e, conseqüentemente, serão um problema para a análise da dinâmica.

### 10.2.2 Análise da Concentração

Existe o julgamento implícito de que não se pode existir preço ou lucratividade excessiva sem alta concentração no mercado. O modelo que justifica essa conclusão é um modelo de Cournot como visto no capítulo 5. Nesse modelo a função lucro é

$$\pi_i = p(Q)q_i - c_i q_i$$

da condição de primeira ordem se tem:

$$\frac{p_i - c_i}{p_i} = \frac{s_i}{\varepsilon}$$

tal que  $s_i = q_i/Q$  e  $\varepsilon = \frac{\partial p}{\partial Q} \frac{Q}{P}$ . Então

$$\sum_i s_i \frac{p_i - c_i}{p_i} = \frac{\sum_i s_i^2}{\varepsilon} = \frac{HHI}{\varepsilon} \quad (10.1)$$

Portanto, neste tipo de análise se olha para o HHI e possivelmente para a elasticidade da demanda. Um exercício poderia ser ver como uma fusão entre dois concorrentes poderia aumentar preços em função de valores de HHI.

Os mercados podem ser divididos em três grupos de concentração com base em valores de HHI. Normalmente as autoridades antitruste expressam o HHI calculado a partir da participação de mercado em percentual (multiplicado por 100). As divisões são as seguintes:

**Mercado não-concentrado:**  $HHI \leq 1000$ . Geralmente as fusões não são investigadas.

**Mercado moderadamente concentrado:**  $1000 \leq HHI \leq 1800$ . Quando fusões propostas nestes mercados produzem variação no  $HHI \leq 100$ , geralmente não são vetadas. Outras fusões podem ser questionadas e vetadas dependendo de outros fatores.

**Mercado concentrado:**  $HHI \geq 1800$ . Fusões nestes mercados causando variação no  $HHI \leq 50$  não geralmente vetadas. Outras fusões podem ser investigadas e vetadas dependendo de outros fatores.

Os guias reconhecem que nestas situações é importante entender a motivação para a mudança na estrutura de mercado. O argumento usual para fusões é de que uma nova tecnologia foi ou pode ser introduzida e será uma força que mudará a estrutura de mercado de forma independente da operação.

O uso do índice HHI para identificar problemas concorrenciais pode deixar de alertar para danos ao ambiente competitivo. O principal problema é que o HHI depende de

quais produtos entram no denominador. Quando se incluem mais produtos que “fazem parte” do mercado, naturalmente a concentração irá cair. Como dito anteriormente, um mercado é definido e avaliado pela grau de substituição entre produtos. Se o produto que é objeto de uma fusão possui muitos substitutos então o HHI será naturalmente baixo. Portanto, um método de “screening” baseado apenas na inclusão ou não de produtos em um mercado é inadequado e gera um debate pouco frutífero de quais produtos devem aparecer no denominador (sobre este debate veja seção 10.3).

### 10.2.3 Efeito Adverso da Fusão sobre Preços

Dados de market share e concentração são apenas o ponto inicial da análise do impacto concorrencial de uma fusão. Antes de determinar que uma fusão levanta sérias preocupações concorrenciais, a autoridade antitruste deve analisar o efeito potencial da operação sobre os preços de mercado, bem como os efeitos sobre entrada e eficiências. A “causalidade” da fusão sobre preços é denominada de *efeito unilateral*. O modelo padrão unilateral foi bem enunciado por Carl Shapiro quando foi economista-chefe do DOJ. Um modelo de análise de efeitos unilaterais é apresentado na seção 10.3.

### 10.2.4 Entrada

A questão básica nesta etapa de análise é: a entrada de novos agentes após a consumação da operação contrabalança os efeitos anticompetitivos identificados anteriormente? Importante observar que a análise da entrada geralmente ocorre quando identificamos efeitos negativos da operação.

Se um analista está trabalhando em um caso de monopólio e pode se pensar existe possibilidade de preço não-competitivo, uma questão importante para se perguntar é o quanto isto é devido ao poder de mercado. Como uma firma neste mercado consegue manter esta situação sem estimular entrada potencial? De forma relacionada, na discussão de fusão esta questão pode ser aplicada: como a firma pós-fusão poderá manter preço mais elevado sem estimular entrada? Isto não é impossível para eles manterem esta situação, ao menos por algum tempo.

Particularmente, se a entrada levar muito tempo os preços podem se manter elevados durante este período. A análise de Stigler deixa o analista com a impressão de que os participantes em uma fusão para monopólio estão cientes de que a operação induz entrada, que por sua vez, tende a erodir preços e market share. Por outro lado, eles também estão cientes que a entrada leva tempo e neste interim a nova empresa pode fazer uma fortuna. Sobre este tópico veja a análise de Stigler sobre a US Steel.

A empresa “fusionada” também pode manter preço elevado por um período se existir um pequeno número de atores na indústria e os preços ou quantidades podem ser facilmente mudados. Novamente, a entrante conhece os preços que são relevantes para ela e que não são os preços antes da entrada. Se os incumbentes puderem convencer os entrantes de que elas podem se comportar de forma diferente pós-entrada, pode ser que as firmas desistam de entrar no mercado. Este é o ponto principal da literatura de detecção de

entrada. Esta literatura considera o comportamento que a incumbente teria que fazer para tornar crível para a entrante que ela irá agir para manter o mercado como está.

Análise de entrada nos guias. Como observado a entrada lida com o impacto das possíveis entrantes como descrito acima. Além disso, os guias lidam com a entrada com compromisso (i.e. é a entrada que envolve custos afundados) que é baseada nas seguintes questões.

1. Pode uma entrada ter impacto em dois anos? (no caso de bens duráveis pode ser considerado processo de entrada mais longo se tem impacto sobre os consumidores).
2. A entrada efetiva será lucrativa? A lucratividade desta entrada deve ser determinada nos preços de mercado pré-fusão, se preocupando também com os custos fixos e com a escala mínima viável (as autoridades assumem que será mais difícil provar a entrada utilizando preços pré-fusão).
3. A entrada efetiva tempestiva e provável tem poder de devolver os preços ao nível pré-fusão?

Para avaliar estas questões, as autoridades antitruste demandam mais informações nas partes envolvidas na fusão. Ainda assim existem boas razões para pensar que será difícil obter as respostas. Quando ou a não entrada irá ocorrer dependem significativamente de objetos que são difíceis de medir, tais como custo afundado e distribuição de habilidades (tecnologia) do possível entrante. Entender a entrada também requer análise da dinâmica do mercado. A velocidade da entrada e, de forma mais geral, a capacidade de resposta a fusão por uma terceira parte que já está na indústria – mudança no custo de bem-estar e na lucratividade das partes envolvidas na fusão. Este tipo de cálculo é difícil e as hipóteses necessárias para avaliar dinâmica podem ser muito específicas.

### 10.2.5 Eficiências e Falência da Firma

Duas questões são importantes na análise de eficiências:

1. Existem ganhos de eficiência que não podem ser atingidos pelas partes por quaisquer outros meios que não pela fusão?
2. Se não for pela fusão uma das partes irá falir provocando a saída dos ativos da indústria?

Até esta etapa a autoridade antitruste não considera qualquer mudança nos custos que seja causada pela fusão. É possível que a sinergia de custos compense o aumento de preços esperado após a fusão?

Este passo na análise de atos de concentração é desenhado para se preocupar com ele. Uma vez que os preços caem com o custo marginal, o efeito líquido das fusões sobre os preços dependem do quanto este efeito pode contrabalançar o efeito “internalização” discutido anteriormente. O tom nesta parte dos guias é que os ganhos de eficiência

devem necessariamente resultar em redução do preços. Os ganhos de eficiência nos guias incluem:

- ganhos de economia de escala;
- melhor integração de estruturas produtivas;
- especialização das plantas;
- custos mais baixos de transporte;
- economias em “servicing”;
- distribuição, e;
- marketing.

É importante observar que muitos dos problemas associado ao estudo da eficiência é porque está relacionado à análise dinâmica.

Existe uma questão do quanto esta lista de eficiências capturam as mudanças que irão resultar em mudanças nas estratégias de investimento devido à fusão. Voltamos novamente a análise de dinâmica de mercado (veja Ericson e Pakes, 1995).

Desta lista faz parte ganhos de eficiência devido à P&D e como as fusões irão afetar este investimento. Muito pouco se sabe sobre os efeitos das fusões sobre investimento em P&D. Duas questões importantes sobre P&D são: (i) fusões tendem a reduzir investimentos em P&D? Os resultados apresentados por Hall (1988) e Auerbach encontram poucos efeitos. Noll (1987) observa que não ocorreu queda em P&D após o fatiamento da Bell. (ii) Qual a eficácia da atividade de P&D antes e depois da fusão? Poderíamos pensar que o fato de haver menos repetição e fluxo de ideias mais aberto poderia ajudar no uso dos resultados do esforço de P&D.<sup>2</sup>

Como toda análise, temos alguns problemas associado a análise de sinergias. Se existem retornos crescentes, então o custo marginal que temos da análise pré-fusão deve ter mudado para ambas sinergias possíveis, e para possíveis efeito-escala, e estes podem se cancelar. As únicas sinergias que contam para a análise estática da fusão são as sinergias que reduzem o custo marginal. Sinergias que afetam o custo fixo ou afundado não importam. Por outro lado, as sinergias sobre custo afundado que geralmente estão por trás de uma fusão somente irão aparecer na análise dinâmica da fusão. Principais sinergias estão tradicionalmente associadas com investimento em bens de capital e na atividade de P&D – isto é uma fonte de preocupação no cálculo de eficiências.

---

<sup>2</sup>Pakes (2004) sugere uma forma de analisar os impactos de eficiência S-R. Observe primeiro que podemos ter estimativas do sistema de demanda que sustentem estimativas de custo marginal pré-fusão. Isto é  $CMg = p_i(1 - 1/\eta_{ii})$ . Se desejamos especificar como as sinergias mudam e temos as razões de desvio, então em muitos casos podemos solucionar isto para os preços pós-fusão como função das sinergias e dos parâmetros de demanda – teríamos um sistema de J equações, um para o preço de cada bem, em J preços desconhecidos. A sugestão é fazer isto para um intervalo razoável de sinergias de custos e observar os preços resultantes destes cenários.



**Falência da firma.** A ideia deste tópico é que fusões/aquisições de uma firma que pode falir evita que os ativos existem deixem o mercado. No caso de veto à fusão, os ativos deixam o mercado, incluindo aqui ativos intangíveis. Este último ponto está implicitamente baseado na ideia de que a firma não consegue se reorganizar nos termos da lei de falências.

Grande parte deste argumento costuma ser usado quando uma indústria está em declínio. Quando o declínio da indústria é explicado por tecnologia obsoleta o argumento é mais forte.<sup>3</sup>

### 10.2.6 Questões e Problemas na Análise de Guias

O foco da análise dos guias é sobre as implicações das fusões sobre os preços, mas as fusões também afetam incentivos dinâmicos de mercado, i.e. controles dinâmicos. Os controles dinâmicos também determinam custos, qualidades, e variedade de produtos vendidos. É claramente possível que estes efeitos dinâmicos tenham efeitos grandes sobre o bem-estar maiores do que os efeitos da análise estática descrita acima. Abaixo discutimos alguns deles – o problema relacionado a reposicionamento de produtos será analisado na seção 10.4.

#### Incentivos ao Investimento

Usualmente esperamos maior investimento em oligopólios do que em monopólios.<sup>4</sup> A razão é que em um oligopólio iremos investir até que o retorno marginal do próximo real seja igual ao custo marginal e parte do retorno marginal virá do efeito “business stealing” – que significa tomar lucros do competidor. O monopólio internaliza este efeito.

O efeito business stealing fez os economistas pensarem que geralmente existirá menos investimentos após uma fusão. A questão então passa a ser: existia investimento demais antes da fusão em relação a um possível nível ótimo de bem-estar social? O investimento ótimo poderia ser calculado como resultado de um problema de *planejador social*. Aqui cabe observar que o planejador social maximiza os excedentes do produtor e do consumidor, enquanto que o monopolista maximiza o excedente do produtor. Então, se a inclusão do excedente do consumidor pode levar a volume diferente de investimento do caso do monopolista. Este efeito é mais provável a favorecer investimentos em indústrias de produtos diferenciados. Em resumo, não podemos detalhar o sinal destes efeitos sem um modelo detalhado, e o que se deseja é comparar as magnitudes do modelo estático com o modelo de investimento. Em muitos casos, o efeito sobre o canal de investimento é maior do que os efeitos de preços sobre o bem-estar.

#### Incentivos sobre Conhecimento

A relação entre apropriabilidade, geração e uso ótimo de tecnologias é uma questão antiga na análise de mercados (Arrow, 1962). O desenvolvimento e a difusão de tecnologia

<sup>3</sup>Um exemplo é um caso de fusão de empresas fabricantes de CDs que foi analisado pelo CADE.

<sup>4</sup>Veja Mankiw e Whinston (1986) e o capítulo 7 de Tirole (1988).

foi uma parte integral de casos famosos do antitruste: IBM, AT&T, Intel, Microsoft. O sentimento geral entre analistas é que por causa dos problemas de apropriabilidade existe sub-investimento em pesquisa. A questão tem dois lados da mesma moeda: (i) as despesas com P&D após a fusão serão mais efetivas?; (ii) as fusões estimulam maiores despesas com P&D?

Pesquisas sobre esta última pergunta, mostram que as despesas totais não são muito afetadas pelas fusões, embora as respostas sejam diferentes em indústrias particulares (e.g. farmacêuticas). Em relação à primeira questão, deveria haver menos espaço para despesa duplicada quando ocorre fusão, mas este resultado depende das trajetória de pesquisa das firmas envolvidas na operação. É bem claro que existe correlação de resultados, mas não em relação às trajetórias. Quando as trajetórias são diferentes, pode ocorrer externalidades não esperadas e efeito transbordamento sobre a economia.

Existe uma discussão sobre eficiência inovativa no guia americano, mas de forma pouco clara. É considerado uma “taxa de desvio de inovações” da firma  $A$  para a  $B$ .  $I_{A,B}$  seria a fração extra de lucro bruto que a firma  $A$  obtém quando destina mais recursos para inovação às custas da firma  $B$ . O guia então discute ajuste para elevações na eficiência do P&D e internalizam as externalidades entre as firmas. Entretanto, não há uma palavra sobre externalidades para os consumidores, e muito do trabalho empírico atual indica que esta pode ser a maior externalidade da atividade de P&D. Isto tipicamente requer firmas concorrentes para reduzir o preço de mercado.

### Efeitos Coordenados

Existem poucos espaço para análise de efeitos coordenados em decorrência de uma fusão. Se existir coordenação entre as firmas, o resultado da análise de UPP pode ser viesado, dado que os preços pós-fusão podem não ser maiores - além da coordenação ser um problema que já demandaria investigação. A discussão básica aqui é que a redução de competidores tornaria mais fácil ou provável algum tipo de comportamento colusivo. Quando as firmas tem tamanho e oportunidades parecidas seria mais fácil suportar colusão. Observe neste contexto que uma medida HHI nem sempre levará a markup mais elevado. Sobre este assunto veja a seção 10.5.

### Leilões

Leilões são considerados brevemente nos guias. Eles discutem sobre leilões de valor privado, e não há menção a eficiência relativa de terceiros. Este ponto é concomitante com a não definição de uma indústria, mas em leilões de valor privado, o problema com uma fusão tende a surgir principalmente entre as partes envolvidas na fusão e terceiros. Se um outro player tiver custos próximas das partes envolvidas na fusão a tendência é ocorrer poucas mudança no resultado dos leilões. Por outro lado, se os terceiros possuem custos muito diferentes então é possível que o resultado mude bastante.

Em segundo lugar, muitos dos leilões de licitações um componente de valor comum, e alguns casos podemos pensar nele como o componente dominante. Então poderíamos

considerar externalidades de informação. Nos leilões de direito de poços de petróleo na região do Golfo é permitido oferta conjunta, dado o aspecto de custo comum em obter informações (ver Porter).

Por último, não há discussão sobre conluio em leilões. Historicamente existem muitos casos de coordenação que lidam com leilões, especialmente em leilões licitatórios dos governos.

### 10.3 UPP

Como comentado anteriormente, efeitos unilaterais surgem se a fusão aumenta a capacidade unilateral da nova firma (entidade) de aumentar os preços ou provocar algum outro dano os consumidores (como afetar qualidade dos produtos).

A avaliação de efeitos anticompetitivos em atos de concentração que envolvem produtos diferenciados necessita de uma estrutura analítica diferenciada da avaliação do grau de concentração, devido a diferenças no grau de substituição entre os produtos das empresas envolvidas, como destacado nos Guias de Análise de Atos de Concentração Horizontal de várias jurisdições. No caso de fusões de empresas que operam em mercados com produtos diferenciados, a análise de efeitos unilaterais toma precedência lógica sobre a análise de efeitos coordenados (tácitos). Nas jurisdições dos EUA e Reino Unido, os guias de análise recomendam a análise direta destes efeitos anticompetitivos unilaterais, na forma de aumentos de preços, realizando o cálculo de 'pressão para aumento de preços' (UPP em inglês) como medida de filtro e inferência para presunção de efeitos anti-competitivos advindos da concentração.

Intuitivamente, se duas empresas ofertam produtos que são vistos pelos consumidores como substitutos, uma fusão gera um incentivo unilateral de aumento de preços. Tal aumento unilateral de preços de um dos produtos, apesar de reduzir a sua própria demanda (efeito canibalização), irá aumentar as vendas do outro produto, que é controlado pela empresa pós-fusão. Esta internalização do desvio dos consumidores, gera um incentivo a aumento de preços, dependendo do grau da taxa de desvio do consumo entre os produtos das empresas e as margens de lucro das mesmas.

#### 10.3.1 Metodologia UPP

Comparamos os preços de dois produtos diferentes, cada um destes produtos controlado por uma única firma em um duopólio de Bertrand com produtos diferenciados (veja capítulo ). Inicialmente apresentamos os cálculos de efeitos unilaterais brutos (sem considerar eficiências) como sugeridos por Shapiro (2010), aplicando tanto para produtos com custos marginais constantes bem como supondo heterogeneidade. Também apresentamos cálculos para demanda linear e sem seguida para demanda com elasticidade constante e simetria entre os dois produtos.

Este tipo de comparação pode ser útil para o estudo de fusões horizontais, tanto para os propósitos de definição de mercado ou para se atacar os efeitos unilaterais da competição.

A ideia básica dos cálculos dos efeitos unilaterais é construir um ponto de partida para uma análise mais profunda. Como destacado por Farrell e Shapiro (2010), a análise dos efeitos unilaterais pode ser mais valiosa do que observar índices de concentração, como é o caso de produtos diferenciados.

A firma  $i$  determina o preço para  $p_i$  para  $i = 1, 2$ . Demanda para o bem da firma  $i$  é dada por  $q_i = D_i(p_1, p_2)$ . Os custos da firma  $i$  são dados por  $C_i(x_i)$ . Seguindo as hipóteses tradicionais de convexidade e diferenciabilidade (veja Tirole, 1988, cap. 1), os lucros da firma  $i$  são dados por:

$$Lucro_i = p_i q_i - C_i(q_i) \quad (10.2)$$

Assumimos que as firmas determinam independentemente os preços de seus produtos antes da fusão. Além disso os preços pré-fusão formam um equilíbrio Bertrand-Nash. Portanto, os preços de equilíbrio pré-fusão  $p_1^{pre}, p_2^{pre}$  são a solução das duas equações de primeira-ordem ( $dLucro_i/dp_i = 0$ ). A condição de primeira ordem para o produto 1 é:  $q_1 + p_1(dq_1/dp_1) - c_1$ . Tal que  $c_1$  é o custo marginal da firma 1. A solução para a firma 2 é análoga.

Em geral os lucros combinados da firma pós-fusão são dados por (veja seção 4.1.2):

$$Lucro_{i,pós} = [p_1 q_1 - C_1(q_1)] + [p_2 q_2 - C_2(q_2)] + S[q_1, q_2] \quad (10.3)$$

Além da soma dos lucros das firmas individuais também temos as sinergias que resultam da operação, representada por  $S[q_1, q_2]$ . As sinergias são poupadoras de custo, pois as firmas podem combinar a produção dos produtos 1 e 2.

Os preços ótimos pós-fusão ( $p_1^*, p_2^*$ ) são a solução para as duas equações de primeira ordem para ( $dLucro_i/dp_i = 0$ ), para  $i = 1, 2$ . Por simplicidade, assumiremos que o custo marginal é constante, ao menos no intervalo de produção relevante. Em relação às sinergias, assumimos que elas se manifestam como redução destes custos marginais. Isto é:  $S[q_1, q_2] = E_1 q_1 + E_2 q_2$ . Por fim, o percentual de aumento de preços pós-fusão para o produto  $i$  é dado por

$$z_i = \frac{p_i^* - p_i^{pre}}{p_i^{pre}} \quad (10.4)$$

A quantificação da comparação dos preços pós-fusão com o pré-fusão é difícil de ser calculada. Em geral sabemos que deve haver algum aumento de preços se os produtos são substitutos e não existem sinergias na operação.

Uma abordagem simplificada da UPP para o produto 1 é

$$UPP_1 = D_{12}(p_2^* - c_2) - E_1 c_1 \quad (10.5)$$

tal que  $p_2^*$  é o preço do produto 2 pós-fusão,  $c$  o custo marginal do produto 1 e 2, respectivamente,  $E_1$  a eficiência decorrente da fusão do produto 1, e  $D_{12}$  é razão de desvio entre os produtos 1 e 2:

$$D_{12} = \frac{\partial q_2(p_2, p_1)/\partial p_1}{\partial q_1(p_2, p_1)/\partial p_1}.$$

A fórmula UPP em (10.5) é simplificada e não conta toda a história. Mas é útil para agências antitruste realizarem filtros iniciais em operação que tratam de produtos diferenciados.

#### *Demanda Linear*

Vamos considerar o caso da demanda linear para o modelo de duopólio. Neste caso, podemos definir as unidades de cada produto tal que a inclinação da curva de demanda seja -1. Estas podem não ser as unidades que os produtos são normalmente medidos, portanto deve-se ter cuidado com a aplicação das fórmulas de cálculo.

Com a normalização das unidades, podemos escrever a curva de demanda linear como

$$q_1 = A_1 - p_1 + D_{21}p_2 \text{ e } q_2 = A_2 - p_2 + D_{12}p_1 \quad (10.6)$$

$A_i$  é o parâmetro de intercepto para cada produto,  $D_{12}$  mede a taxa de desvio do produto 1 para o produto 2. Isto representa a fração de vendas perdidas da firma 1, quando ela aumenta o preço do produto 1, que são capturados pelo produto 2.  $D_{21}$  é claramente análogo.

#### Equilíbrio de Bertrand Pré-fusão

Resolvendo para a equação de lucro pré-fusão (10.2) e utilizando a equação de demanda (10.6) apropriada, o lucro total da firma é  $(p_1 - c_1)(A_1 - p_1 + D_{21}p_2)$ . Diferenciando com respeito a  $p_1$  e aplicando a condição de primeira ordem temos a função melhor-resposta da firma 1,  $2p_1 = A_1 + c_1 + D_{21}p_2$ . Da mesma forma, temos a função-melhor resposta para a firma 2:  $2p_2 = A_2 + c_2 + D_{12}p_1$ . A solução do sistema destas duas equações fornece a solução Bertrand-Nash do problema do duopólio (veja Tirole, 1988, cap. 5 e Cabral, 1994, cap. 3). A solução implica que o par de preços  $(p_1^{pre}, p_2^{pre})$  satisfaz a solução do sistema das funções-melhor resposta:

$$p_1^{pre} = [A_1 + c_1 + D_{21}p_2^{pre}] \frac{1}{2} \quad (10.7)$$

$$p_2^{pre} = [A_2 + c_2 + D_{12}p_1^{pre}] \frac{1}{2} \quad (10.8)$$

Solucionando explicitamente para os preços de equilíbrio, temos:

$$p_1^{pre} = \frac{2(A_1 + c_1) + D_{21}(A_2 + c_2)}{4 - D_{12}D_{21}} \quad (10.9)$$

$$p_2^{pre} = \frac{2(A_2 + c_2) + D_{12}(A_1 + c_1)}{4 - D_{12}D_{21}} \quad (10.10)$$

Estes preços são a base de comparação da equação (10.4).

*Equilíbrio de Bertrand Pós-fusão: Caso de Ausência de Sinergias*

Nesta subseção descrevemos os preços pós-fusão e comparamos com o preços pré-fusão, supondo que não existam sinergias. No caso pós-fusão ambos os produtos são propriedade de uma única firma. Supondo que as condições de mercado não mudem drasticamente, podemos assumir que a nova firma maximiza o lucro composto (equação (10.3)) sem sinergias:

$$Lucro_{pós} = (p_1 - c_1)(A_1 - p_1 + D_{21}p_2) + (p_2 - c_2)(A_2 - p_2 + D_{12}p_1) \quad (10.11)$$

Diferenciando com respeito a  $p_1$  e fazendo a CPO igual a zero, temos:

$$2p_1 = A_1 + c_1 + (D_{12} + D_{21})p_2 - D_{12}c_2 \quad (10.12)$$

A análoga para  $p_2$  é:

$$2p_2 = A_2 + c_2 + (D_{21} + D_{12})p_1 - D_{21}c_1 \quad (10.13)$$

Solucionando o sistema das equações (10.12) e (10.13), encontramos o preço pós-fusão:  $(p_1^*, p_2^*)$ .

Como o objetivo da análise é a comparação de preços descritos pela equação (10.4), vamos focar na diferença entre os preços pré e pós-fusão. Começando com o produto 1, por exemplo, partimos da equação (10.12) e subtraímos dela o preço pré-fusão como descrito pela equação (10.7). Desta forma temos:

$$2(p_1^* - p_1^{pre}) = D_{21}(p_2^* - p_2^{pre}) + D_{12}(p_2^* - c_2) \quad (10.14)$$

Para o produto 2:

$$2(p_2^* - p_2^{pre}) = D_{12}(p_1^* - p_1^{pre}) + D_{21}(p_1^* - c_1) \quad (10.15)$$

Substituindo a equação (10.15) na (10.14), temos:

$$2(p_1^* - p_1^{pre}) = \frac{D_{21}[D_{21}(p_1^* - p_1^{pre}) + D_{12}(p_1^* - c_1)]}{2 + D_{12}(p_2^* - c_2)} \quad (10.16)$$

Reorganizando os termos e expandindo temos:  $(p_1^* - p_1^{pre})(4 - D_{21}D_{12}) = D_{21}^2[D_{21}(p_1^* - p_1^{pre}) + p_1^{pre} + p_1^{pre} - c_1] + 2D_{12}(p_2^* - p_2^{pre} + p_2^{pre} - c_2)$ . Substituindo novamente a equação (10.15) para eliminar  $p_2$ , temos:  $(p_1^* - p_1^{pre})(4 - D_{21}D_{12} - D_{21}^2) = D_{21}^2(p_1^* - c_1) + 2D_{12}(p_2^* - c_2) + D_{12}[D_{12}(p_1^* - p_1^{pre}) + D_{21}(p_1^* - c_1)]$ . Movendo e isolando  $(p_1^* - p_1^{pre})$  no lado esquerdo

e expandindo  $(p_1^* - c_1)$  teremos:  $(p_1^* - p_1^{pre})(4 - (D_{21} + D_{12})^2) = 2D_{12}(p_2^{pre} - c_2) + D_{12}(D_{21} + D_{12})(p_1^{pre} - c_1)$ . Dividindo esta equação pelo preço pré-fusão para chegar ao termo análogo de aumento de preços  $z$  (10.4), teremos:

$$\frac{p_1^* - p_1^{pre}}{p_1^{pre}}(4 - (D_{21} + D_{12})^2) = 2D_{12} \left( \frac{p_2^{pre} - c_2}{p_2^{pre}} \right) \left( \frac{p_2^{pre}}{p_1^{pre}} \right) + D_{12}(D_{21} + D_{12}) \left( \frac{p_1^{pre} - c_1}{p_1^{pre}} \right) \quad (10.17)$$

Vamos usar as seguinte definição para simplificar a notação:

**Índice de Lerner pré-fusão:**  $M_i^{pre} = \frac{p_i^{pre} - c_i}{p_i^{pre}}$  para os produtos  $i = \{1, 2\}$ .

Portanto a mudança de preços após a fusão ser consumada será:

$$\frac{p_1^* - p_1^{pre}}{p_1^{pre}} = \frac{2D_{12}M_2^{pre} \frac{p_2^{pre}}{p_1^{pre}} + D_{12}(D_{21} + D_{12})M_1^{pre}}{4 - (D_{21} + D_{12})^2} \quad (10.18)$$

Como destacado por Shapiro (2010), esta formula é relacionada a chamada GUPPI

$$\text{GUPPI} = D_{12} \frac{(p_2^{pre} - c_2)}{p_1^{pre}}.$$

A GUPPI mede o custo de oportunidade de vender uma unidade do produto 1, devido a ter a propriedade do produto 2, medida como fração do preço do produto 1, tomando como dado o preço do produto 2.

De forma mais especifica o primeiro termo da equação (10.18) é  $2 \times \text{GUPPI}$ , se ignorarmos os demais termos da equação (10.18). Neste caso, o aumento de preços seria de  $\text{GUPPI}/2$ . Isto corresponde a taxa de passagem (pass-through) de 50% para uma única firma que se depara com uma função de demanda linear aplicada ao termo do custo de oportunidade de ser proprietária do produto 2 (mantendo todos os outros preços fixos). Como a equação (10.18) mostra, o aumento do preço de equilíbrio para o produto 1 com demanda linear é maior do que o previsto pela GUPPI pois o preço do produto 2 irá também aumentar bem como o termo de “feedback” entre os dois preços.

#### *Demanda linear sem normalização*

Shapiro (2010) descreve o aumento de preços do produtos normalizando a função demanda. Hausman, Moresi e Rainey (2011) apresentam fórmula similar mas sem a necessidade de normalização dos dois produtos para curva de demanda com inclinação -1.

O equilíbrio pré-fusão é caracterizado de forma similar as equações (10.7) e (10.8). Por exemplo, para o produto 1 a equação de primeira ordem seria:

$$\frac{-q_1(p_1^{pre}, P_{-1}^{pre})}{p_1^{pre} - c_1} = \frac{-q_1^{pre}}{p_1^{pre} - c_1} \quad (10.19)$$

Por sua vez, a equação de primeira ordem para o produto 1 pós-fusão seria dada por

$$\frac{p_1^* - p_1^{pre}}{p_1^{pre}} = \frac{2D_{12}M_2^{pre} \frac{p_2^{pre}}{p_1^{pre}} + D_{12}D_{21}M_1^{pre} + \frac{(p_1 - c_1)^2}{(p_2 - c_2)p_1} \frac{q_2}{q_1} (D_{21})^2}{4 - 2(D_{21} + D_{12}) - \frac{(p_2 - c_2)}{(p_1 - c_1)} \frac{p_1}{p_2} \frac{q_1}{q_2} (D_{12})^2 - \frac{(p_1 - c_1)}{(p_2 - c_2)} \frac{p_2}{p_1} \frac{q_2}{q_1} (D_{21})^2} \quad (10.20)$$

Esta equação é essencialmente igual a (10.18), mas sem a necessidade de normalização da função demanda. Quando realizamos este procedimento podemos observar que a fórmula é substancialmente parcimoniosa em sua notação.

Sem ganhos de eficiências ambas as fórmulas devem necessariamente prever aumento de preços em uma fusão. Esta é uma razão para permitir eficiências  $E$ . Observa-se que o tratamento nos guias é assimétrico para eficiências.

#### *Demanda linear com sinergias*

Vamos agora adicionar sinergias ao caso de demanda linear normalizada. Observamos anteriormente que os preços pós-fusão são solução das equações (10.12) e (10.13). Neste sistema, o custo marginal,  $c_1$  e  $c_2$ , é tratado como dado. Agora vamos permitir mudanças nos custos marginais, i.e., ganhos de eficiências resultante da fusão. Tomando o diferencial total de (10.17) com respeito a  $c_1$  e  $c_2$ , temos:

$$\frac{dp_1^*}{dc_1} [4 - (D_{21} + D_{12})^2] = 2 - D_{21}(D_{21} + D_{12}) \quad (10.21)$$

e

$$\frac{dp_1^*}{dc_2} [4 - (D_{21} + D_{12})^2] = D_{21} - D_{12}. \quad (10.22)$$

Com as sinergias de  $E_1c_1$  para o produto 1 e sinergias  $E_2c_2$  para o produto 2, o preço de equilíbrio para o produto 1 é o nível calculado acima, sem as sinergias, menos o montante estabelecido como:  $\frac{dp_1^*}{dc_1} E_1c_1 + \frac{dp_1^*}{dc_2} E_2c_2$ . Com a devida substituição, o decréscimo devido à sinergias é para o produto 1 é

$$\frac{1}{[4 - (D_{21} + D_{12})^2]} \times \{ [2 - D_{21}(D_{21} + D_{12})] E_1c_1 + (D_{21} - D_{12}) E_2c_2 \} \quad (10.23)$$

Esta queda de pressão sobre os preços pode ser expressa como fração do preço pré-fusão do produto 1, assumindo a regra da elasticidade  $c_1 = p_1(1 - M_1)$  e  $c_2 = p_2(1 - M_2)$ , como:

$$\frac{1}{[4 - (D_{21} + D_{12})^2]} \left\{ [2 - D_{21}(D_{21} + D_{12})] E_1(1 - M_1^{pre}) + (D_{21} - D_{12}) E_2(1 - M_2^{pre}) \frac{p_2}{p_1} \right\}.$$

Portanto, a equação completa UPP com sinergias no caso de demanda linear é:

$$\frac{p_1^* - p_1^{pre}}{p_1^{pre}} = \frac{X + Y}{4 - (D_{21} + D_{12})^2} \quad (10.24)$$



tal que

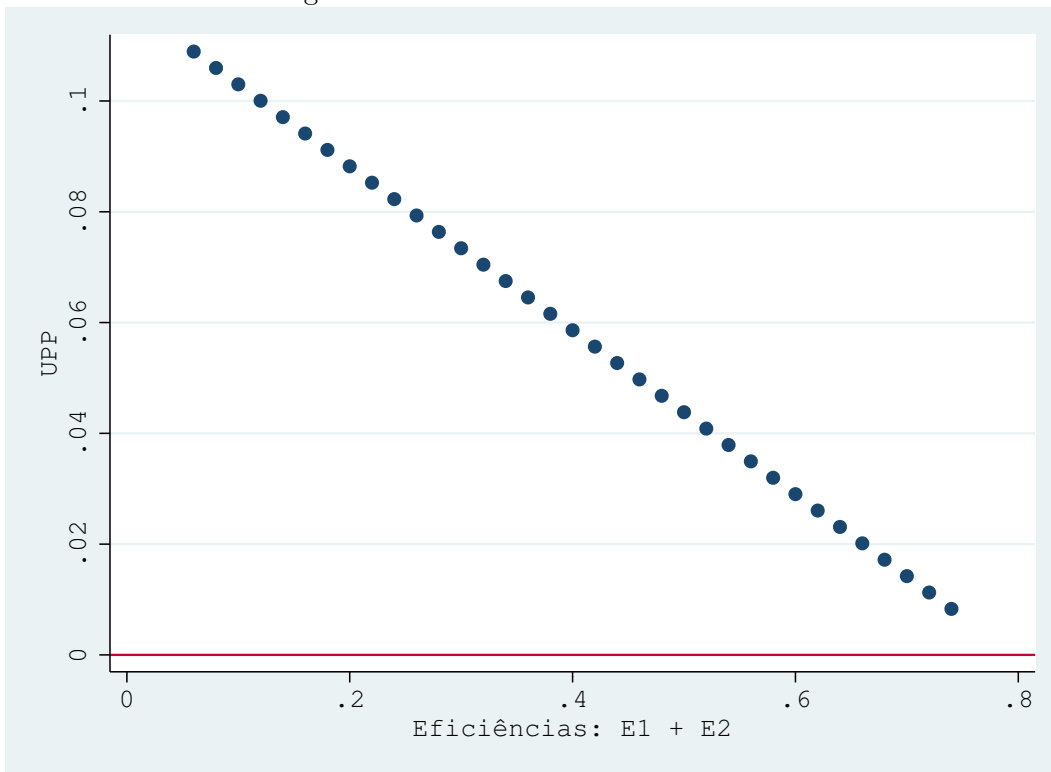
$$X = (2D_{12}M_2^{pre} - E_2(1 - M_2^{pre})(D_{21} - D_{12}))\frac{p_2^{pre}}{p_1^{pre}}$$

e

$$Y = D_{12}(D_{12} + D_{21})M_1^{pre} - E_1(1 - M_1^{pre})(2 - D_{21}(D_{12} + D_{21}))$$

Na Figura 10.1 é apresentado a relação entre a pressão por aumento de preços (UPP) e o efeito da eficiência sobre a redução do repasse de preços expresso na equação (10.24). A equação é parametrizada como  $D_{12} = 0.32$ ,  $D_{21} = 0.28$ ,  $M_2^{pre} = 0.5$  e  $M_1^{pre} = 0.4$ . Assumimos o preço relativo entre os produtos 2 e 1 como 1.1. As eficiências são simétricas,  $E_1 = E_2$ , e elas variam de 0.02 até 0.37. No eixo x da Figura 10.1 é apresentada a soma das eficiências,  $E_1 + E_2$ , e no eixo y o resultado final da UPP. Observe que para pouca eficiências a pressão por aumento de preços é de mais de 10%. A medida que a eficiência aumenta a redução é linear mas é lenta. Ou seja, para esta parametrização da equação (10.24), o repasse de eficiências é pequeno. A pressão sobre o preços passa a ser inferior a 5% quando a soma das eficiências é superior a 0.4.

Figura 10.1: UPP e Ganho de Eficiência



### 10.3.2 Problemas Conceituais e de Medida

Existem muitas questões relativa ao uso das fórmulas UPP. O maior problema conceitual é que na análise de UPP não há resposta de terceiros à operação de fusão. Isto é mais

do que dizer que as funções de reação não mudam, pois uma terceira firma não irá responder a nenhuma mudança. Isto significa que o preço de uma terceira firma não será um preço ótimo. O argumento é de que preços são complementos estratégicos. Todavia, não sabemos a priori se os preços são complementos estratégicos.

O argumento para se aplicar UPP é de que é uma método prático quando se ignora outras empresas na indústria. Quando permitimos a análise de terceira partes temos que analisar todo o mercado e isto pode ser não-trivial e trazendo novamente a discussão do que é o mercado.

## 10.4 Reposicionamento de Produtos

Reposicionamento de produto é uma mudança nas características do produto comercializado por uma firma incumbente. O ponto aqui aqui é que as características podem mudar em virtude de uma fusão.

A análise de respostas a mudança no ambiente econômico (principalmente concorrencial) separa as respostas em efeitos de controles estáticos (preços e quantidades) e dinâmicos (entrada, saída e várias formas de investimento incluindo novos produtos). Geralmente a análise de reposicionamento é tratada na análise dinâmica de mercados. A análise dinâmica é muito complicada de estudar e muitas vezes deixamos para uma análise menos formal. Como destacado por Pakes (2021), trabalhos recentes mostraram que um conjunto de indústrias, com firmas ativas, podem mudar as características dos seus produtos da mesma forma que mudam os preços. Quando isto ocorre, a análise estática aplicada para estudar fusões pode levar a conclusões errôneas.

Esta análise de reposicionamento de produto utiliza técnicas empíricas distinta do estudo de dinâmica, permitindo a análise estática compatível com a determinação de preços. Este tipo de análise também evita a questão de equilíbrios múltiplos típico da abordagem de dinâmica. Tipicamente esta abordagem usa duas etapas: (i) as características dos produtos são escolhidas; (ii) os preços são determinados.

Nosko (2014) analisou a relação Intel/AMD quando a Intel introduziu o chip *Core 2 Duo* no mercado em julho de 2006. Um mês antes da introdução do novo chip existia intensa competição no mercado de produtos de alta performance. AMD comercializava o produto de valor mais alto, com preço pouco acima de 1000 dólares (USD), e outros sete chips com valores entre 600 e 1000 USD. Em outubro, apenas três meses após a introdução do *Core 2 Duo*, a AMD não mais comercializava nenhum chip de alto valor, não existiam chips no intervalo 600-1000 USD, e a Intel vendia todos os seus chips acima de 1000 USD. O autor explicou que os retornos da inovação do *Core 2 Duo* foram primariamente dos markups que a Intel conseguiu sustentar como resultado de tirar a AMD do mercado de chips intermediários (deixando um espaço neste segmento) e de alto valor. Após a introdução deste novo produto a Intel passou a dominar o mercado de chips de alta performance.

Como destacado por Pakes (2021), a abordagem de reposicionamento de produtos é

relativamente recente mas tem sido utilizada em alguns mercados. Wollman (2018) analisou os impactos do “bailout” na indústria de caminhões durante a grande recessão nos EUA. Neste mercado, a cavalo (a unidade propulsora) pode ser adaptada facilmente ao tipo de unidade de transporte (o trailer).<sup>5</sup>

Nas indústrias onde o reposicionamento é relativamente fácil, as agências reguladoras deveriam considerar o papel central desta característica sobre o impacto nos preços quando ocorre uma fusão. Dada a habilidade de se analisar demanda, preços e lucros, esta análise de reposicionamento apenas adiciona o estudo dos custos fixos associados com a introdução/retirada de um produto ou contrato do mercado.

Para realizar esta análise é necessário um método para estimar um equilíbrio contrafactual. É preciso calcular lucros que seriam ganhos com um produto que não foi comercializado. O aumento nos lucros de introduzir um novo produto deve ser maior do que o custo fixo associado com a introdução condicional sobre o gerenciamento da informação que se tinha a disposição quando a decisão de introdução foi tomada. Similarmente se um produto que poderia ser comercializado não foi, se espera que o aumento nos lucros que poderiam ser ganhos deveria ser menor do que o custo fixo. Portanto, a média destes incrementos deve fornecer limites inferiores e superiores ao custo fixo médio.

**Um Rascunho de Modelo de Custo Fixo.** Vamos apresentar um modelo conceitual de estimação de limites para identificar o custo fixo. Suponha o caso de custo fixo constante  $F$ .<sup>6</sup> Faça  $x_j$  ser um vetor de 1 e 0: 1 quando o produto é oferecido e 0 caso contrário. Faça um vetor  $e_z$  com um 1 no lugar do  $z$  e zero em todos os outros lugares. Assuma que  $z$  foi adicionado. Calcule os lucros resultantes se o produto  $z$  não tivesse sido adicionado (desvio unilateral em um jogo de movimento simultâneo).

Faça

$$\Delta\pi_j(x_j, x_j - e_z, x_{-j}) \equiv \pi_j(x_j, x_{-j}) - \pi_j(x_j - e_z, x_{-j}) \quad (10.25)$$

Se o produto (ou características novas)  $z$  foi adicionado é porque o lucro esperado é maior do que o custo fixo:

$$E[\Delta\pi_j(x_j, x_j - e_z, x_{-j})|\mathcal{I}_j] \geq F, \quad (10.26)$$

tal que  $\mathcal{I}_j$  é o conjunto informação do agente  $j$ . Tome a média sobre todos os produtos introduzidos e assuma que a expectativa dos agentes não possuem viés. Isto leva a uma estimativa consistente do valor inferior de  $F$ .

Se  $z$  é uma adição possível ao portfólio e não foi oferecido e

$$\Delta\pi_j(x_j, x_j + e_z, x_{-j}) \equiv \pi_j(x_j, x_{-j}) - \pi_j(x_j + e_z, x_{-j}) \quad (10.27)$$

<sup>5</sup>Outras aplicações incluem a análise dos impactos de retirar do mercado computadores de baixa performance (Eisenberg, 2014) e o custo de negociar contratos bilaterais quando se muda as características do produtos que estão sob contrato (Ho, 2009, Crawford and Yurukoglu, 2012).

<sup>6</sup>Quando o custo fixo não é constante então é necessário um modelo mais complexo.

então

$$E[\Delta\pi_j(x_j, x_j - e_z, x_{-j})|\mathcal{I}_j] \leq F, \quad (10.28)$$

que fornece uma estimativa análoga do limite superior de  $F$ . Este tipo de estimativa é realizada empiricamente com o modelo de limites – que é complicado para uma abordagem de graduação em economia.<sup>7</sup> Entretanto, como agências antitruste podem acessar informação sobre custos, mesmo que sempre seja complicado navegar nestes dados, é possível ter uma ideia clara de qual o custo fixo associado a comercialização de produtos com características diferentes.

## 10.5 Efeitos Coordenados

Após uma fusão pode ocorrer maior facilidade para coordenação em determinado mercado. Por exemplo, Miller e Weinberg (2017) mostram que após a criação de uma *joint-venture* entre as cervejarias Miller e Coors nos EUA, a concorrente ABI e a nova MillerCoors passaram a praticar algum tipo de coordenação na determinação dos preços.

## 10.6 Fusão com Produtos Complementares

Nesta seção apresentamos um modelo simples de fusão/aquisição quando se trata de produtos complementares (veja Akgun, Caffarra, Etro, e Stillman, 2020). Suponha que dois bens compostos 1 e 2 são comercializados ao preço final  $P_i$ , tal que  $i = \{1, 2\}$ . Estes bens possuem função demanda  $Q_i = Q_i(P_i, P_j)$ , em que a demanda é decrescente em relação ao preço do próprio produto  $i$  e crescente em relação ao outro produto  $j$ . Cada bem requer dois componentes que sejam *complementos perfeitos*.

Por simplicidade suponha três firmas sem custos. Duas firmas são competem e vendem dois componentes substitutos ao preço  $p_i$ ,  $i = \{1, 2\}$ . Outra firma,  $C$ , é monopolística e vende o componente complementar ao preço  $w$ . Portanto, o preço final dos dois bens  $i$  será de  $P_i = w + p_i$ .

As funções lucro das três firmas serão:

$$\Pi_A = Q_1(P_1, P_2) \times p_1 \quad (10.29)$$

$$\Pi_B = Q_2(P_1, P_2) \times p_2 \quad (10.30)$$

$$\Pi_C = [Q_1(P_1, P_2) + Q_2(P_1, P_2)] \times w \quad (10.31)$$

Todas as três firmas determinam os preços simultaneamente e analisamos o impacto de uma fusão entre as firmas  $A$  e  $B$  e a firma monopolista  $C$ .

<sup>7</sup>O modelo de estimativa de limites é apresentado em Pakes, Porter, Ho e Ishii (2015).

**Integração vertical do produtor  $B$ .** Suponha a situação em que a firma  $C$  comercializa um componente apenas para o bem 1, enquanto que a firma  $B$  produz os dois componentes para o bem 2. Dois casos mais simples emergem neste modelo: (i) uma fusão das firmas  $A$  e  $C$  pode criar um efeito EDM reduzindo os preços ao consumidor, trazendo benefícios para os consumidores; (ii) uma fusão das firmas  $B$  e  $C$  deve criar um efeito RRC que leva ao aumento de preços dos produtos 1 e 2, sendo uma situação pior para os consumidores.



# Capítulo 11

## Mercados Verticais

### 11.1 Relações Verticais

O acompanhamento de fusões e aquisições é uma das missões centrais de uma autoridade antitruste. Nos EUA isto é claro na Seção 7 do Clayton Act que proíbe fusões tal que o efeito seja “substantially lessen competition, or to tend to create a monopoly.” Isto regularmente utilizado por agências antitruste para vedar fusões horizontais que possam ser problemáticas. Por outro lado, as agências destacam complexidade em poucos casos de fusões verticais e vedam alguns poucos casos. Nos EUA, a fusão AT&T/Time Warner foi o primeiro grande casos que foi para litígio em mais de 40 anos. Embora existe maior preocupação com operações verticais, o tema é um dos mais controversos do antitruste.

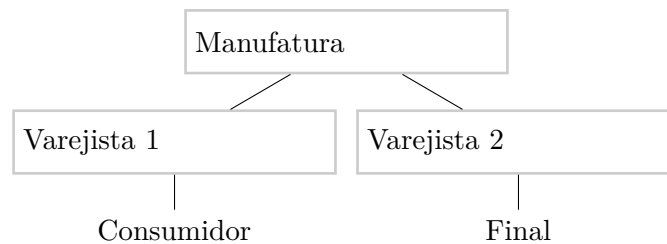
Embora tenham poucos casos de vedação da fusões verticais este tipo de operação entre empresas é muito comum. A maioria das empresas não vendem para os consumidores e sim para outras empresas. Exemplos:<sup>1</sup> produtores de cimento vendem para concreteiras, que por sua vez vendem para construtoras; Produtoras de TV vendem para varejistas, que por sua vez vendem para os consumidores; produtoras de conteúdo de TV vendem para *empacotadoras de canais*, que vendem para distribuidoras de programação, que vendem para o consumidor final.

Por que relações verticais são diferentes de horizontais (fusão e conluio)? As razões são as seguintes:

1. Uma firma que vende diretamente para consumidores normalmente controla grande parte das variáveis que determinam a demanda pelo produto: preço, qualidade, propaganda, etc. O mesmo não ocorre quando um produtor vende para um varejista. Vários determinantes da demanda final pelo produto estão fora do controle do produtor. Particularmente, o preço de varejo é controlado pelo varejista e não pelo produtor. Portanto, a demanda do produtor depende do preço (preço de atacado) e de diversos fatores que não são controlados por quem produz.

---

<sup>1</sup>Veja Cabral (2017, cap. 13).



2. Varejistas competem entre si. Em outras palavras, um varejista se preocupa com o preço que ele paga ao produtor, mas também se preocupa com o preço que os competidores pagam. Como o preço de atacado forma o custo marginal, ele é importante para a competitividade do varejista.
3. A terceira razão é que o número de empresas intermediárias é muito menor do que possíveis consumidores finais. Geralmente uma firma que vende para os consumidores finais possui maior poder de mercado do que uma firma que vende para outras firmas.

Geralmente usamos como exemplos de relação vertical um produtor que vende para um varejista, que por sua vez atende os consumidores.

Relação Vertical: Exemplo

Exemplo simples. Manufatura vende para 2 varejistas que podem vender para os consumidor final.

Essas razões justificam o estudo da relação vertical entre empresas. Por relações verticais é entendido a relação entre, pelo menos, duas empresas que estão na cadeia de valor de um mercado.

Dois dos principais tópicos são *integração vertical* e *restrição vertical*. Sobre restrições verticais uma excelente referência é o capítulo 4 do livro de Whinston (2006).

## 11.2 Dupla Margem e a Integração Vertical

Considere o simples exemplo (Cabral, 2017, seção 13.1). Num mercado com uma produtora  $M$  e uma firma varejista  $R$ . Um exemplo seria de  $M$  como uma refinaria de petróleo e  $R$  como um posto de combustível.

Suponha que existe demanda  $D(p)$  pelo produto ofertado por  $R$ . A suposição mais simples seria de que  $R$  precisa de pelo uma unidade de insumo para vender o produto (supondo apenas o insumo como custo). Portanto, o varejista tem custo  $p^w$ , que é o preço de atacado, e o produtor  $M$  tem custo marginal constante  $c$ .

Para fazer lucros, o produtor cobra  $p^w > c$ . O varejista determina a produção ótima fazendo a receita marginal igual ao custo  $p^w$ . O preço ótimo para o varejista é  $p$ .



Um exemplo de dupla margem pode ser solucionado com formas funcionais simples: demanda linear e custo marginal constante. *O problema da dupla margem emerge para a precificação linear.* Assumindo a função demanda  $D(p) = a - p$  e custo marginal sendo apenas o produto intermediário, o problema do varejista é

$$\pi_R = (p - p^w)D(p) = (p - p^w)(a - p) \quad (11.1)$$

O preço ótimo é

$$p = \frac{a + p^w}{2}$$

que resulta na quantidade:

$$q = \frac{a - p^w}{2}$$

Em seguida temos o problema da manufatura

$$\pi_M = (p^w - c)q = (p^w - c)\frac{a - p^w}{2} \quad (11.2)$$

A quantidade é o volume negociado pelo varejista e o preço da manufatura é o custo marginal do varejista. Aqui o problema da manufatura é o mesmo de um monopolista com custo marginal  $c$  e demanda  $q = \frac{1}{2}(a - w)$ . O preço ótimo será

$$p^w = \frac{a + c}{2}$$

Se a manufatura e o varejista forem integrados então temos o lucro da firma integrada:

$$\pi_{IV} = (p - c)D(p) = (p - c)(a - p) \quad (11.3)$$

O preço ótimo é  $p_{IV} = (a + c)/2$  e a quantidade é  $q_{IV} = (a - c)/2$ . A comparação é direta: o preço integrado é menor do que o não integrado:  $\frac{a+c}{2} < \frac{a+p^w}{2}$ .

Utilizando alguns parâmetros ( $a = 9$ ,  $c = 1$ ), o preço integrado é 5 e preço não integrado é 7 (dado que  $p^w = 5$ ). Em termos de lucros a variação resultante é de 33%. *Se a manufatura determina um preço de atacado para um varejista verticalmente separado, então os lucros conjuntos são mais baixos e o preço do varejo é mais elevado do que quando se tem integração vertical.*

**Concorrência *Downstream*:** As coisas ficam mais complicadas quando existe mais de um varejista. Este caso é o que está descrito na Figura acima. Faça  $p_i^w$  ser o preço do produto intermediário pago pelo varejista  $R_i$  e  $p_i$  é preço de varejo deste produto ( $i = 1, 2$ ). Suponha que a manufatura faz uma fusão com um dos varejistas, qual o impacto que podemos esperar sobre os preços?

O primeiro efeito que podemos esperar é a redução em  $p_1^w$ . Isto corresponde ao efeito da dupla marginalização. Uma vez que a manufatura e o varejista maximizam lucros conjuntamente o bem intermediário deixa de existir.

O segundo efeito que poderíamos esperar é um aumento em  $p_2^w$ . A firma  $R_2$  é agora rival da firma  $M - R_1$ . Um aumento em  $p_2^w$  tem um efeito sobre a receita da manufatura, um efeito que a firma  $M$  anteriormente não tinha. Um aumento em  $p_2^w$  é uma forma de *aumentar o custo dos rivais*.

Para analisar  $p_1$  o efeito é misto. A integração vertical elimina dupla margem e coloca pressão para redução do preço. Por outro lado, a integração gera o efeito de suavização da competição que tende a empurrar o preço para cima.

*Exemplo mais completo*

Apresentamos agora mais completo com a solução de Tirole (1988) para o problema da dupla margem. Inicialmente suponha uma manufatura (upstream) e um varejista (downstream).

O varejista escolhe  $p$  sob tarifa linear:  $T(q) = p^w \cdot q$ . Isto é, o problema de dupla margem surge quando a precificação é linear. O timing é o seguinte:

1. Manufatura escolhe a tarifa linear
2. Varejista escolhe o preço final para o consumidor

O lucro do varejista é:

$$\pi_R = (p - p^w)D(p) \quad (11.4)$$

A FOC é

$$\frac{\partial D(p)}{\partial p}(p - p^w) + D(p) = 0$$

se  $D(p) = 1 - p$  então  $\frac{\partial D(p)}{\partial p} = -1$ . Neste caso, podemos calcular  $p$  a partir da FOC:  $p = (p^w + 1)/2$ . No caso geral a solução é:

$$p = -\frac{D(p)}{[\partial D(p)/\partial p]} + p^w \quad (11.5)$$

No caso particular (-1) a demanda total é

$$q = \frac{1 - p^w}{2}$$

A manufatura soluciona

$$\max_{p^w} [(p^w - c)D_w(p^w)] \quad (11.6)$$

no caso particular o problema é

$$\max_{p^w} \left[ (p^w - c) \left( \frac{1 - p^w}{2} \right) \right] \quad (11.7)$$

Solucionando o problema do caso particular (11.7), se tem a FOC

$$\frac{1}{2} - p^w + \frac{1}{2}c = 0$$

$$p_w = \frac{1+c}{2} \quad (11.8)$$

A FOC para o caso geral é

$$\frac{\partial D_w(\cdot)}{\partial p^w}(p^w - c) + D_w(\cdot) = 0 \quad (11.9)$$

Solucionado para  $p^w$

$$p^w = -\frac{D_w(\cdot)}{\partial D_w(\cdot)/\partial p^w} + c \quad (11.10)$$

A dupla margem aqui é:

$$\frac{D(p)}{[\partial D(p)/\partial p]} - \frac{D_w(\cdot)}{\partial D_w(\cdot)/\partial p^w}$$

Como solucionar o problema da dupla margem? A solução para a dupla margem é a integração vertical (IV)?

**Solução contratual (restrição vertical):** Se o custo marginal da firma é  $c$ , então a firma a jusante deveria escolher a quantidade que maximiza o lucro total. A solução de Tirole é precificar usando *franchise fee* ou tarifa de duas partes:

$$T(q) = A + p_w q$$

$A$  é a parte fixa do preço a ser cobrado do varejista.

O problema simples do varejista seria:

$$(p - p^w)D(p) - A$$

Então a FOC não seria distorcida:

$$\frac{\partial D(p)}{\partial p}(p - p^w) + D(p) = 0$$

Nesse caso o lucro da manufatura é apenas  $A$ .

O problema aqui é que preços/contratos são em sua maioria linear... (puzzle?)

**Concorrência Downstream.** Se duas empresas são idênticas no mercado downstream e competem a la Bertrand, então  $A = 0$  e  $p^w =$  preço de monopólio. Isto permite que um monopolista acima na cadeia possa extrair todas as “taxas” com tarifa linear. Nesse caso não existe necessidade de IV para resolver o problema da dupla margem.

No geral, como a IV pode reduzir o tamanho da dupla margem (quando presente), pode existir melhora de Pareto para consumidores e firmas.

**Descrição completa da dupla margem.**<sup>2</sup> Incluindo custo marginal do varejista e diversas empresas acima e abaixo na cadeia de valor, a função de lucro é:

$$\pi_R = \sum_{j \in S_R} [p_{jt} - p_{jt}^w - c_{jt}^R] s_{jt}(p) \quad (11.11)$$

<sup>2</sup>Seguindo aqui as notas de aula de Sofia Villas-Boas.

tal que  $s_{jt}(p)$  é o market share. A FOC é

$$\frac{\partial s_{jt}(\cdot)}{\partial p_{jt}} [p_{jt} - p_{jt}^w - c_{jt}^R] + s_{jt}(\cdot) = 0 \quad (11.12)$$

Defina  $\Delta_R$ : a matriz de resposta do varejista contendo as primeiras derivadas de todos os shares com respeito a todos os preços ( $p$ ).  $\Delta_R \rightarrow (i, j) = \partial s_{j,t} / \partial p_{i,t}$ . Portanto:

$$p_{jt} - p_{jt}^w - c_{jt}^R = -s_{jt}(p) \Delta_R^{-1} \quad (11.13)$$

A função lucro de cada manufatura é:

$$\pi_w = \sum_{j \in S_w} [p_{jt}^w - c_{jt}^w] s_{jt}(p(p^w)) \quad (11.14)$$

A FOC é

$$\frac{\partial s_{jt}(\cdot)}{\partial p_{jt}^w} [p_{jt}^w - c_{jt}^w] + s_{jt}(\cdot) = 0 \quad (11.15)$$

Defina  $\Delta_w$ : matriz de resposta da manufatura  $\partial s_{jt}(\cdot) / \partial p_{jt}^w$

Podemos decompor em  $\Delta_w = \underbrace{\Delta_p'}_{p(p^w)} \Delta_R$

Simplificando se tem:

$$p_{jt}^w - c_{jt}^w = -\Delta_w^{-1} s_{jt}(\cdot) \quad (11.16)$$

Definindo  $MKP^w = p_{jt}^w - c_{jt}^w$  e analogamente  $MKP^R = p_{jt} - p_{jt}^w - c_{jt}^R$ , se tem que

$$MKP^w + MKP^R = -\Delta_w^{-1} s_{jt}(p) - \Delta_R^{-1} s_{jt}(p) \quad (11.17)$$

Esta é a dupla margem.

### *Oligopólio Bilateral*

Quando existem múltiplas firmas acima e abaixo da cadeia ( $m, j$ ), a relação contratual pode ser multilateral. Mas mesmo neste cenário o contrato continua a ser **bilateral**.

Questões chave: (i) externalidades existem *entre* firmas quando os contratos são formados e os termos são negociados. (ii) Concorrência entre firmas em cada nível da cadeia vertical podem dissipar os lucros da indústria.

Como consequência, os resultados de equilíbrio de séries de contratos bilaterais não precisam maximizar os lucros da indústria.

### 11.2.1 Mercado Vertical: Externalidades

O papel das externalidades será central na análise. Relembrar o Teorema de Coase (1960): se pode ocorrer externalidade nas trocas, então barganha levará a um resultado eficiente (embora não se diga nada sobre distribuição).

Então para a intervenção ser garantida (na perspectiva de um economista), precisa existir algo que limite a aplicação do Teorema de Coase.

- ✓ E.g., restrições sobre contratos; contratação multilateral em oposição a bilateral; partes afetadas não estão presentes quando o contrato ocorrer; outras fricções.

Vender para firmas em oposição a consumidores finais é uma fonte crucial de externalidades em análise vertical

- ✓ Uma firma que vende diretamente para consumidores normalmente controla a maioria das variáveis que determinam a demanda (preço, qualidade, propaganda, serviços, investimento, etc).
- ✓ Isto não é mais verdade quando uma firma vende por meio de um varejista ou intermediário. Ex. quando os consumidores dos bens geralmente não são os consumidores finais dos produtos, eles podem competir com uma outra empresa por seus próprios consumidores.

*Resposta das firmas.* Quando as firmas entram em diversas atividades para mitigar ineficiências ou pressão competitiva por meio de atividades tais quais

- ✓ Integração
- ✓ Seletividade de contrato
- ✓ Preço não-linear (e.g. tarifa multi-parte, bundling, AUD, compartilhamento de receita, ...)
- ✓ Restrição vertical (e.g. RPM, acordos de exclusividade, MFNs)

Historicamente, muitas restrições são vistas com ceticismo pelos tribunais.<sup>3</sup> Mas recentemente tem usado regra da razão para lidar com estes casos.

Para uma revisão de análise teórica sobre estes casos veja Whinston (2006) cap. 4. Muitos trabalhos empíricos na década passada sobre restrições verticais. Ver a resenha de Lafontaine e Slade (2008), Hastings (2004) e Hortacsu e Syverson (2007).

### 11.2.2 Hold-Up

Um problema associado a relação vertical e também aos limites da firma é um problema chamado de *hold-up* ou problema de comprometimento. Investimentos às vezes ocorrem em algum tipo de parceria comercial, neste caso o retorno desta parceria excede outras alternativas caso não tivesse o acordo/contrato. Uma vez que um investimento é

---

<sup>3</sup>Sherman Act §1: “Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal.”

afundado, o investidor deve dividir o *retorno bruto* com o seu parceiro comercial. Este problema é inerente a muitas relações bilaterais.

Por exemplo, manufaturas e ofertante de peças geralmente desenham seus equipamentos e processos produtivos para as necessidades especiais de seus parceiros, sabendo que a futura negociação irá conferir parte do benefício para os seus parceiros. Aqui o risco do investidor está no fato de ceder ganhos para o parceiro, isto desencoraja investimentos.

Um exemplo sobre *hold-up* é a indústria de carvão nos EUA, em que 80% do carvão é utilizado para a geração de energia elétrica.<sup>4</sup> Uma vez que os custos de transporte de uma mina para uma geradora são elevados, existe um bom *payoff* em localizar plantas geradoras perto das minas. Considere agora a decisão de construir uma planta de energia perto de uma mina de carvão. Uma vez que não existe outra mina perto, este investimento é denominado de *ativo específico*. Isto implica que a planta de energia dissociada da mina de carvão possui muito menor valor. Por esta razão, se o proprietário da planta for diferente de que controla a mina, a relação entre as duas partes é sujeito ao problema de *hold-up*. Uma vez que o comprador paga por um ativo específico para uma relação, o comprador pode cobrar um preço elevado.

Uma solução para este problema é a integração vertical. Tipicamente plantas geradoras localizadas próximas a minas de carvão possuem o mesmo proprietário no mercado americano. Vale observar que a integração vertical não é capaz de solucionar todo o problema de *hold-up*. Outras soluções seriam contratos longos ou regulação da relação entre as partes. Por exemplo, planta de energia e mina próximas que não são verticalizadas no mercado americano, operam com contratos de longo prazo.

**Hold-up e investimento.** <sup>5</sup> Um comprador ( $R$ ) e um vendedor ( $M$ ) podem comercializar o volume  $q$ , assumindo que a quantidade é positiva e inferior a um limite superior  $\bar{q}$ . A comercialização ocorre ou pelo menos se beneficia do investimento realizado pelo vendedor  $M$ . Por simplicidade, representamos a decisão de investimento como uma variável binária, isto é,  $M$  investe, então  $I = 1$ , ou não investe, logo  $I = 0$ . O custo de investimento para o vendedor  $M$  é  $C.I$ , com  $C > 0$ . O custo de entrega de  $q$  para  $R$  é  $c_I(q)$ . Pelo lado do comprador  $R$ , temos um excedente bruto ao consumir  $q$  de  $v_I(q)$ . Ambas as funções,  $c_I(q)$  e  $v_I(q)$  são estritamente crescentes em  $q$  e  $c_I(0) = v_I(0) = 0$ .

O problema de maximização é

$$\phi_I = \max_{q \in Q} \{v_I(q) - c_I(q)\}, \quad (11.18)$$

sendo que  $\phi_I$  representa o excedente social eficiente dado o investimento realizado pelo vendedor e  $q_I^*$  é o volume comercializado associado ao excedente socialmente eficiente, i.e. o volume comercial quando não se tem o problema do *hold-up*. Portanto, o excedente social líquida da despesa de investimento será  $W(I) = \phi_I - kI$ . Agora vamos supor a

<sup>4</sup>Joskow (1985) e Cabral (2017), p. 55.

<sup>5</sup>Seguimos uma versão estilizada de Che e Sákovics (2004).

seguinte condição

$$\phi_I - C > \phi_N, \quad (11.19)$$

então será socialmente desejável para o vendedor investir. Aqui estamos dizendo que o excedente da transação menos o custo do investimento será maior do que um valor mínimo que a firma  $M$  recebe se não investir.

Uma hipótese crucial neste modelo é que a decisão de investimento do vendedor, embora observada pelas partes, não é verificável, e portanto não pode ser garantida por um contrato. Por enquanto, também vamos assumir que a natureza da relação comercial é suficientemente incipiente tal que as partes podem contratar  $q$  somente depois que a decisão de investir de  $M$  foi feita. A modelagem deste contrato é no estilo de barganha de Nash, resultando em decisões eficientes  $q_I$  e dividindo o excedente bruto igualmente entre as partes. Portanto, o vendedor se apropria apenas de uma fração do retorno do seu próprio investimento enquanto ele arca com todo o custo de investir,  $C$ . Neste caso, dado que o vendedor investe o *payoff* será:

$$U_M(I) = \frac{1}{2}\phi_I - CI. \quad (11.20)$$

Nesta função o retorno do comércio,  $\phi_I$ , é dividido de forma simétrica entre  $M$  e  $R$ . Suponha

$$\frac{1}{2}\phi_I - C < \frac{1}{2}\phi_N. \quad (11.21)$$

Neste caso, mesmo que o investimento seja socialmente desejável, o vendedor não irá investir. Ele não investe porque a condição (11.19) não é satisfeita. Nesta situação o sub-investimento aparece – i.e. o problema *hold-up*.

Uma interpretação desta ineficiência é a falha do *Teorema de Coase*. As partes não podem alcançar o resultado eficiente uma vez que o resultado da não-contratação da decisão de investimento de  $M$  não permite que as partes possam negociar previamente. Aqui poderia ser interpretado como se ainda não existisse objeto sobre o qual negociar. A partir desta perspectiva, o problema de *hold-up* implica em um custo de transação e como Coase sugere, o custo de transação pode ser evitado ou reduzido por meio de outra estrutura organizacional.

**Soluções e remédios: direito de propriedade.** Como mencionado anteriormente, algumas soluções são possíveis. Integração vertical seria uma e outra seria contrato de longo-prazo. A integração vertical é outra estrutura organizacional como sugerido por Williamson (1979). Grossman e Hart (1986) e Hart e Moore (1990) perguntam como que o problema do *hold-up* some ou ao menos é reduzido com a integração vertical. Para eles esta solução não é clara e passa pela estrutura de propriedade que afeta as partes expostas a este tipo de problema. De acordo com estes autores, a propriedade de um ativo fornece ao proprietário o direito de determinar o uso do ativo que é contratualmente não especificado. As partes ainda irão negociar os termos comerciais, mas este direito residual que vem da propriedade importa, uma vez que ele determina os *payoffs* de status quo da partes em uma negociação.

Suponha agora que tanto a firma vendedora  $M$  como a compradora  $R$  podem ser proprietárias de todos os ativos necessários para a operação vertical. Aqui podemos ter dois tipos de integração: (i) integração-R, integração vertical pelo lado do comprador; (ii) integração-M, integração pelo lado do vendedor. O objetivo é corrigir a integração  $R$  ou  $M$  e a decisão de investir de  $M$ . Se as firmas falham em acordar a decisão comercial, a parte  $i$ , onde  $i = \{R, M\}$ , pode unilateralmente realizar o *payoff* de status quo que é  $\Phi_i^i(I)$  e, por sua vez, a parte  $j(\neq i)$  pode realizar o seu respectivo *payoff*  $\Phi_j^i(I)$ . Neste caso, dado que o investimento é 0 ou 1, Grossman, Hart, e Moore consideram ser razoável assumir que

- (i)  $\Phi_i^i(I) + \Phi_j^i(I) \leq \phi_I$ : o status quo é bem-estar dominado pelo comércio eficiente;
- (ii)  $\Phi_M^i(I = 1) - \Phi_M^i(I = 0) < \phi_I - \phi_N$ : o investimento de  $M$  é **investimento-específico**;
- (iii)  $\Phi_i^i(I = 1) > \Phi_i^i(I = 0)$ , e  $\Phi_i^j(I = 1) = \Phi_i^j(I = 0)$ : o investimento aumenta o payoff de status quo do proprietário mas não o payoff de quem não é proprietário.

Assumindo novamente que as partes dividem em simetria o ganho de comércio em qualquer situação, o *payoff* do vendedor  $M$  será

$$U_M^i(I) =$$

$$\Phi_M^i(I) + \frac{1}{2} [\phi_I - \Phi_R^i(I) - \Phi_M^i(I)] - C.I = \frac{1}{2}\phi_I + \frac{1}{2} [\Phi_M^i(I) - \Phi_R^i(I)] - C.I \quad (11.22)$$

Neste sentido o ganho de investimento do vendedor fazendo integração é

$$U_M^i(I = 1) - U_M^i(I = 0) = \frac{1}{2}(\phi_I - \phi_N) + \frac{1}{2}\Delta^i - C \quad (11.23)$$

$$\Delta^i = [\Phi_M^i(I = 1) - \Phi_M^i(I = 0)] - [\Phi_R^i(I = 1) - \Phi_R^i(I = 0)].$$

Dada as hipóteses (ii) e (iii) de Grossman-Hart-Moore,  $\phi_I - \phi_N \geq \Delta^M > 0 > \Delta^R$ . Então

$$W(I = 1) - W(I = 0) \geq U_M^M(I = 1) - U_M^M(I = 0) > \quad (11.24)$$

$$U_M(I = 1) - U_M(I = 0) > U_M^R(I = 1) - U_M^R(I = 0).$$

Isto mostra que integração-M é a estrutura de propriedade ótima, dominando a estrutura simétrica não-integrada, que por sua vez domina a estrutura com integração-R. Em particular, se vale a condição

$$U_M^M(I = 1) - U_M^M(I = 0) > 0 > U_M(I = 1) - U_M(I = 0), \quad (11.25)$$



então o investimento é sustentável quando o vendedor  $M$  é o proprietário dos ativos. Este resultado é a principal conclusão de Grossman-Hart-Moore de que a propriedade do ativo pode servir para reduzir a exposição do proprietário ao *hold-up*.<sup>6</sup>

**Soluções e remédios: contratos.** No modelo acima, a decisão comercial é contratável apenas depois de tomada a decisão de investimento. Se as partes podem calcular o ganho comercial esta hipótese pode não ser boa. Se as partes podem contratar baseado em  $q$  antes de decidir o investimento, o problema de sub-investimento pode ser solucionado, sem integração vertical ou similares.

Suponha que as partes assinam um contrato requerendo que eles negociem  $\hat{q}$  para o preço total  $\hat{t}$ . A menos que renegociado, este contrato irá fornecer ao vendedor  $M$  um *payoff* de  $\hat{t} - c_I(\hat{q}) - C.I$  se ele faz sua escolha de investimento  $I$ . Por outro lado, se  $\hat{q} \neq q_I^*$ , ambas as partes estarão em uma situação melhor se renegociarem para implementar  $q_I^*$ . Novamente assumindo que a renegociação divide o excedente simetricamente, o *payoff ex ante* do vendedor  $M$  será:

$$\hat{U}_M(I; \hat{q}) = \hat{t} - c_I(\hat{q}) + \frac{1}{2} [\phi_I - (v_I(\hat{q}) - c_I(\hat{q}))] - C.I. \quad (11.26)$$

Portanto, o benefício líquido de  $M$  investir com este contrato é

$$\begin{aligned} & \hat{U}_M(I = 1; \hat{q}) - \hat{U}_M(I = 0; \hat{q}) = \\ & = \frac{1}{2}(\phi_I - \phi_N) - \frac{1}{2}(v_I(\hat{q}) - v_N(\hat{q})) - \frac{1}{2}(c_I(\hat{q}) - c_N(\hat{q})) - C. \end{aligned} \quad (11.27)$$

O quanto um contrato como esse pode criar o incentivo suficiente para  $M$  investir depende da natureza do investimento. Primeiro suponha que o investimento da firma  $M$  é apenas para ela, um investimento “egoísta”, tal que ele apenas diminui o custo de  $M$  mas não afeta a valoração da firma compradora  $R$ , i.e.  $v_I(\cdot) = v_N(\cdot)$ . Neste caso, o contrato comercial pode ainda proteger o incentivo de  $M$  investir. é possível observar que

$$c_N(q_I^*) - c_I(q_I^*) = v_I(q_I^*) - c_I(q_I^*) - [v_N(q_I^*) - c_N(q_I^*)] > \phi_I - \phi_N. \quad (11.28)$$

Pela mesma lógica,

$$c_N(q_I^*) - c_I(q_I^*) < \phi_I - \phi_N.$$

Uma vez que  $c_I(\cdot)$  é contínua, então existe  $\hat{q}^*$  entre  $q_N^*$  e  $q_I^*$ , tal que

$$c_N(\hat{q}^*) - c_I(\hat{q}^*) = \phi_I - \phi_N.$$

Conseqüentemente,

$$\hat{U}_M(I = 1; \hat{q}^*) - \hat{U}_M(I = 0; \hat{q}^*) = W(I = 1) - W(I = 0),$$

<sup>6</sup>Os efeitos de estruturas de direito de propriedade podem depender da escolha de solução do problema de barganha. Por exemplo, a opção externa ou a solução bidding de Bertrand podem mudar os rankings de estruturas alternativas e podem eliminar as ineficiências.

então a vendedora  $M$  investirá até quando for eficiente. Alguns autores mostram que um contrato de preço fixo para  $q$  pode fornecer incentivos eficientes para um investimento egoísta de qualquer lado, com uma condição adicional, e para investimento “egoísta” dos dois lados, em ambientes mais gerais com investimento contínuo.

Todavia, os contratos podem não restaurar eficiência se os investimentos não são “egoístas.” Por exemplo, suponha investimento cooperativo, i.e.  $c_I(\cdot) = c_N(\cdot)$ .

### 11.3 Resenha da Evidência Empírica

Política antitruste sobre restrições verticais foi e continua a ser controverso. A teoria é ambígua, então a ênfase é em trabalho empírico cuidadoso.

Lafontaine e Slade (2008) resumem a evidência sobre restrições verticais (i.e., relações de exclusividade no varejo): tanto casos sobre os preços (RPM) e não-preço (exclusividade de território, tying, etc). Embora o artigo seja um resumo da literatura, ele é útil como introdução aos métodos usados para endereçar estas questões. O foco de Lafontaine e Slade é sobre as consequências e não sobre a incidência de restrições verticais. As autoras focam primariamente sobre os efeitos de preço.

#### *Rationale de restrições verticais*

Argumentos de eficiência: dupla margem (RPM máx., taxa fixa para franchise); serviço de revendedor (exclusividade) + “free riding” (RPM min); incentivos de investimento ex-ante ao revendedor (territórios exclusivos).

Razões anti-competitivas: cartéis e monopolização (RPM min, territórios exclusivos); fechamento e aumento do custo dos rivais (revenda exclusiva para dirigir/manter ausência de competição); suavização da concorrência através da delegação da precificação (territórios exclusivos para permitir poder de mercado com o varejista; Rey and Stiglitz, 1995).

Política pública: variável e heterogênea. Atualmente a maioria das restrições são avaliadas por regra-da-razão.

#### *Estimação por painel, cross-section e séries de tempo*

Estatísticas descritivas são úteis para apontar regularidades empíricas, mas podem existir muitas explicações possíveis para a ocorrência do padrão ou comportamento. A abordagem econométrica básica (cross-section de firmas ou estabelecimentos do varejo):

$$p_i = \alpha + \beta \times VR_i + \text{controles} + e_i$$

Questões: Primeiro, as restrições podem ser endógenas. Exemplos, exclusividade ocorre quando a firma upstream precisa fazer investimentos substanciais, mas isto também influencia preços. Outra questão é a falta de instrumentos apropriados: instrumentos deve ser correlacionado com  $VR$  mas não com preço.

Cross-section de regiões que bane e não-bane  $VR$ . Endogeneidade potencialmente reduzida, mas não eliminada: e.g., RPM pode não ser banido onde o poder de mercado upstream/downstream é maior (portanto, preços mais altos devido a dupla margem).

Séries de tempo: se os fatores que afetam tanto preço como a restrição vertical são invariantes no tempo, pode ser usado abordagem de dados em painel (efeito fixos):

$$p_{it} = \alpha_i + \beta \times VR_{it} + \text{controles}_{it} + e_{it} \quad (11.29)$$

remove marcas, características de lojas ou características regionais não-observadas que são invariantes no tempo. Para aplicar este método é preciso variação temporal nas restrições e nos preços. Além disso, não podem existir características não observáveis que mudam ao longo do tempo.

#### *Experimentos naturais*

Para aplicar o método empírico de experimentos naturais é preciso: (i) uma mudança exógena da política; (ii) um grupo que foi afetado (grupo de tratamento) e um não afetado (grupo de controle). Se estas condições estão presentes então se pode estimar um modelo DID típico entre regiões ( $j$ ) ao longo do tempo ( $t$ )

$$p_{it}^j = \alpha_i + \delta^j + \gamma^j D_{jt} + \beta^T X_{it} + u_{it}^j \quad (11.30)$$

tal que  $D_{ij} = 1$  se ocorreu a mudança de política ( $VR$ ).

Uma regressão usando primeiras diferenças resulta em estimativa consistente:

$$\Delta p_{it}^j = \gamma D_{jt} + \beta \Delta X_{it} + \Delta u_{it}^j \quad (11.31)$$

O importante aqui é que o tratamento precisa ser exógeno. Neste caso fica a pergunta: qual realista seria essa possibilidade em organização industrial?

#### *Modelos estruturais*

Muitas vezes não se tem mudanças na política para realizar estimativas de forma reduzida. Então como entender mudanças na política? A saída é utilizar modelos estruturais.

Vantagens:

- ✓ Permite a recuperação dos parâmetros da política que são invariantes. Como consequência permite análise contrafactual.
- ✓ Identifica mecanismos que ajudam a entender os efeitos de heterogeneidade, extrapolação dos resultados.
- ✓ Maior transparência das hipóteses.

Custos (alguns):

- ✓ Hipóteses mais fortes sobre o mercado (equilíbrio, especificação do jogo, comportamento das empresas, etc).
- ✓ Previsão de mudanças nos custos e outras características não-modeladas.

Em resumo, Lafontaine e Slade sugerem que restrições podem ser publicamente desejáveis quando privadamente desejáveis: implementação voluntária pode solucionar questões de incentivos e melhorar eficiência. Entretanto, restrições obrigatórias tendem a reduzir bem-estar.

Firmas podem ser melhor informadas quando restrições verticais são as mais úteis para resolver problemas de incentivos ou investimento; portanto, evidência inicial parece apoiar atitudes lenientes do antitruste em relação à restrições verticais.

### 11.3.1 Integração Vertical e Preços de Gasolina

Hastings (2004) analisa o impacto de integração vertical sobre o preço da gasolina.

Impacto do leasing de longo prazo e conversão de 260 postos de gasolina independentes (Thrifty's) pela ARCO sobre o comportamento de preços dos competidores.

Estratégia: foco no mercado de gasolina em Los Angeles e San Diego (1997), pré e pós operação, com substancial dispersão geográfica dos postos da Thrifty nos sub-mercados (localidades). Argumentação de que o evento é assumido ser não-estratégico e exógeno em relação aos sub-mercados locais, i.e. localizações da Thrifty e as características são pré determinadas.

Preço detalhado e características para os postos de gasolina. Utiliza DID por posto com efeitos fixos por cidade-tempo. Competidor é definido como aquele que está em um raio de 1 milha.

Modelo empírico:

$$p_{it} = \mu + \alpha_i + \delta\gamma.t + \phi c_{it} + \theta z_{it} + \epsilon_{it}, \quad (11.32)$$

aqui  $\alpha_i$  desvio específico a cada posto da média  $\mu$ ,  $\gamma$  é dummy de cidade,  $t$  dummy trimestral,  $\theta z_{it}$  variável indicadora se um posto compete com um posto independente (não verticalizado),  $c_{it}$  é indicador para o caso se um competidor muda para uma empresa integrada.

Findings

- ▶ Competidores da Thrifty tiveram aumento de preços; não dependente do quanto a operação foi para uma companhia ou um varejista (ver na Tabela 11.1 a perda de um competidor independente).
- ▶ Resultado com diferenciação por marca/lealdade por marca. Observada subsequente suavização da competição: postos independentes competem mais por preço; mudança para competidor com marca aumenta preços.
- ▶ Hastings e Gilbert (2005): estudam fusões usando variação intra e entre mercados e encontram que independentes pagam preços maiores no atacado se estão mais próximos de competidores.

Tabela 11.1: Estimativa de Efeitos Fixos: Efeito Tratamento, Preço de Varejo para Gasolina Comum

Variáveis	(1)	(2)	(3)
Empresas Oper.	0.1080 (.0107)	-0.0033 (.0178)	-.0033 (0.0122)
Independentes		-0.1013 (.0143)	-0.0500 (.0101)
LA × Fev.			.0180 (.0065)
LA × Jun.			.0243 (.0065)
LA × Oct.			.1390 (.0064)
SD × Fev.			-.0851 (.0036)
SD × Jun.			-.0304 (.0036)
SD × Oct.			.0545 (.0036)
Constante	1.3465 (0.0421)	1.3465 (0.0415)	1.3617 (0.0287)
F	70.86	623.59	
R2	0.3772	0.3955	0.7181

TABLE 4  
VERTICAL INTEGRATION AND READY-MIXED CONCRETE PRICES: PLANT-LEVEL RESULTS

	WITHIN-MARKET DIFFERENCE				CHANGE FOR CONTINUERS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Observations	12,553	12,553	8,555	8,555	4,025	4,025	2,439	2,439
$R^2$	.424	.425	.501	.501	.456	.460	.575	.576
Vertical integration indicator	-.022*	-.017*	-.006	-.006	.011	.011	.003	.005
	(.006)	(.006)	(.007)	(.007)	(.035)	(.035)	(.034)	(.034)
Multunit indicator		-.012*		-.001		-.037		-.033
		(.004)		(.005)		(.020)		(.028)
TFP			-.214*	-.215*			-.237*	-.237*
			(.015)	(.015)			(.028)	(.028)
	INTEGRATED VS. UNINTEGRATED ENTRANTS				INTEGRATED ENTRANTS VS. UNINTEGRATED INCUMBENTS			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Observations	2,771	2,771	2,025	2,025	7,490	7,490	6,104	6,104
$R^2$	.563	.566	.655	.655	.430	.430	.523	.523
Vertical integration indicator	-.037*	-.025	-.020	-.012	-.023	-.023	-.012	-.012
	(.018)	(.019)	(.020)	(.020)	(.012)	(.012)	(.012)	(.012)
Multunit indicator		-.032*		-.025		NA		NA
		(.012)		(.014)				
TFP			-.218*	-.215*			-.221*	-.221*
			(.031)	(.031)			(.019)	(.019)

### 11.3.2 IV e Ganhos de Eficiência

Hortacsu e Syverson (2007). Integração vertical leva a ganhos de eficiência? Ou ganhos de eficiência levam a integração vertical (seleção)? Além disso, existem efeitos de fechamento de mercado?

Analisa integração vertical entre produtoras de cimento e de concreto (ready-mix concrete) entre 1963 e 1997. Utilizam microdados do Censo da Manufatura (BoC).

- ✓ Amostra grande: aproximadamente 220 cimenteiras, 5200 plantas de concreto para cada ano de dados.
- ✓ Estratégia: examina variação no nível do mercado e das plantas em termos de preços, quantidades, produtividade e taxas de entrada.

Evidência para cada mercado: utilizando medidas em nível de planta com efeitos fixos por mercado-ano.

Resultados:

- ✓ IV não leva a provável fechamento: preços caem, quantidades aumentam e taxas de entrada permanecem imutáveis quando mercados se tornam mais integrados.
- ✓ Padrão consistente com razões de eficiência para se fazer IV: produtores com alta produtividade fazem mais integração, tem mais probabilidade sobrevivência e cobram preços mais baixos. Ainda existe a história de economias de escala e habilidade gerencial: IV em mercados corresponde a

expansão de firmas grandes e mais produtivas. Por outro lado pode ser apenas tamanho: grandes firmas que não fazem IV também apresentam este comportamento. Além disso, IV pode ser uma forma de firmas grandes e eficientes entrarem nos mercados.

## 11.4 Taxa de Passagem e Poder de Mercado

Usando o conceito de concorrência perfeita, de oferta positivamente inclinada e demanda negativamente inclinada, a teoria econômica prevê taxa de passagem de custos ou impostos para o consumidor entre 0 e 100%. Entretanto, a taxa de passagem varia substancialmente de acordo com a inclinação e forma das curvas de demanda e ofertam, bem como da concorrência no mercado.

Ambos produtores e consumidores tomam preços como dados e decidem quantidades que maximizam lucros ou utilidade, quando apropriado. Faça  $p$  ser o preço pago pelos consumidores, e por hora ignoramos a notação de consumidor ou firma. Firmas pagam um imposto por unidade  $t$  e portanto recebem  $p - t$ .

Vamos começar a análise com taxa de passagem no caso de mercado concorrencial.

### Questões de auxílio

Qual a importância da precificação nos problemas relacionados a relações verticais exclusionárias e com integração vertical?

Pense exemplos de tipo de restrições verticais: (i) RPM (resale price maintenance); (ii) territórios exclusivos, e (iii) revenda autorizada (exclusiva).

Um problema relacionado é a possibilidade de conluio com o tipo de relação vertical. Qual o tipo de contrato/tecnologia favoreceria conluio?

Quais tipos de contratos e ambiente econômico facilita *fechamento de mercado*?





## Capítulo 12

# Produtividade e Função de Produção (\*)

Neste capítulo tratamos de estimação de função de produção e algumas questões relativas às medidas de produtividade.

### 12.1 Função de Produção

No capítulo 3 vimos os fundamentos da função de produção, em seguida tratamos de análise de diversos situações de comportamento de mercado.

Considere uma função de produção Cobb-Douglas simples em log natural

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \varepsilon_{it}, \quad (12.1)$$

tal que  $y_{it}$  é log natural do produto (medida como valor-adicionado),  $k_{it}$  é o log do capital,  $l_{it}$  é o log natural do fator trabalho. Todas estas três variáveis são observadas. Dois termos da especificação econométrica não são observáveis:  $\omega_{it}$  e  $\varepsilon_{it}$ . O termo  $\varepsilon_{it}$  representam choques à produção ou produtividade que não são observáveis (são aleatórios) pela firma antes de sua decisão de contratação e usos de insumos em  $t$ .

O termo  $\omega_{it}$  representam os choques de produtividade que são potencialmente observados ou previstos pelas firmas quando fazem as suas decisões de insumos. De forma intuitiva,  $\omega_{it}$  representa variáveis tais como habilidade gerencial da firma, paradas programadas devido à manutenção, qualidade do solo, ou mesmo o índice pluviométrico em uma fazenda ou planta produtiva. Seguindo esta ideia intuitiva, o termo  $\varepsilon_{it}$  representa desvios da parada programada de maquinário ou do volume de chuva esperado. Além disso,  $\varepsilon_{it}$  poderia representar medidas de erro na variável de produção (que seriam seriamente correlacionadas). Nesta especificação, que é mais utilizada, a produtividade é Hicks-neutra, i.e., ela é aditiva na especificação log-linear.

O ponto-chave na identificação econométrica inerente ao conceito de função de produção é de que a escolha dos insumos produtivos  $k_{it}$  e  $l_{it}$  são determinados pelas firmas. Isto significa que se a variável econométrica não-observável  $\omega_{it}$  é conhecida (ou parcialmente observada) pela firma antes da escolha dos insumos  $k_{it}$  e  $l_{it}$ . Sendo assim, estas escolhas deverão depender da produtividade  $\omega_{it}$ , o que gera correlação entre  $\omega_{it}$  e  $(k_{it}, l_{it})$ . Se a equação (12.1) for estimada por MQO, o termo  $\omega_{it}$  faz parte do resíduo, levando a estimativas inconsistentes de  $\beta_k$  e  $\beta_l$ . Aqui voltamos no problema de identificação de equações simultâneas, tal como observado por Marschak e Andrews (1944).

Unidade de medida. Na especificação ideal,  $l$ ,  $k$ , o produto  $y$ , bem como o investimento  $i$  são medidos em unidades físicas comparáveis. Se por acaso, o produto  $y$  for receita, então o conceito de produtividade deve ter interpretação diferente. Seria como uma medida de produtividade baseada na receita (De Loecker, 2011).

### Modelo Comportamental

O modelo comportamental é construído para entender as decisões das firmas quando ocorrem mudanças no ambiente econômico. O modelo comportamental de Olley e Pakes (1996) é baseado nos trabalhos de Hoppenhayn e Rogerson (1993) e Ericson e Pakes (1995). A primeira definição das chamadas *variáveis de estado*. Elas resumem a história da firma e são a base para a tomada de decisões de mercado. As variáveis de estado da firma  $i$  são  $(k_{it}, a_{it}, \omega_{it})$ . A nova variável,  $a_{it}$ , é a idade da firma  $i$  (tempo de atividade no mercado).

A estrutura de mercado é representada pelo valor das variáveis de estado  $(k_{it}, a_{it}, \omega_{it})$  de todas as firmas ativas. Os lucros da empresa  $i$  relaciona as variáveis de estado mais a estrutura de mercado em moeda corrente (USD ou BRL).

*Cadeia de decisões e hipóteses:*

1. O conjunto de informação da firma ( $J_t$ ) é determinado em  $t$ . Neste conjunto é incluído os choques de produtividade do passado e atuais, i.e.  $\{\omega_\tau\}_{\tau=0}^t$ . Os choques futuros não são incluídos no conjunto de informação.
2. No começo de cada período a firma decide se continua ou sai do mercado.
3. Se sair, a empresa não retorna ao mercado;
4. Se continua, a firma escolhe os fatores variáveis e o investimento. Fator variável no caso mais simples (equação (12.1)) é o trabalho. Saída e investimento ocorrem dentro do período  $t$  e são função da produtividade  $\omega_t$ .

O modelo comportamental da firma é descrito por uma equação de Bellman, onde o subscrito  $t$  indexa uma estrutura de mercado (mesmo porque é tudo que ocorre em cada período do tempo – o indicador da firma é omitida porque o modelo descreve a

decisão de uma empresa):

$$V_t(k_t, a_t, \omega_t) = \max\{\Phi, \pi_t(k_t, a_t, \omega_t) + \sup_{i_t \geq 0} (-c(i_t) + \beta E[V_{t+1}(k_{t+1}, a_{t+1}, \omega_{t+1}) | J_t])\} \quad (12.2)$$

sujeito as leis de movimento das variáveis de estado  $(k_t, a_t, \omega_t)$ :

$$k_{t+1} = (1 - \delta)k_t + i_t \quad (12.3)$$

$$a_{t+1} = a_t + 1 \quad (12.4)$$

$$F_\omega \equiv \{F(\cdot | \omega), \omega \in \Omega; F(\cdot | \cdot) \text{ estocasticamente crescente na ordem natural de } \omega\}. \quad (12.5)$$

Das três leis de movimento, apenas o fator trabalho é afetada pela escolha da firma. Ele otimizado no curto prazo, i.e., durante (intra-período)  $t$  e pode reagir ao aumento em  $\omega_t$ . Se a firma investe mais ela aumenta seu estoque de capital. O estoque de capital apenas pode ser expandido pelo investimento. A lei de movimento para a idade da firma,  $a_t$ , é determinística, e a regra de movimento da produtividade é um processo de Markov. Isto significa que a produtividade expande de acordo com um distribuição governada por um processo de Markov e é conhecida pelas firmas e estocasticamente crescente em  $\omega_t$ . Vale observar que as firmas conhecem o processo de distribuição  $F$  mas não qual choque irão receber.

Na equação (12.2),  $V_t(\cdot)$  é a função valor,  $\Phi$  é a regra de saída da firma do mercado (descrita abaixo),  $\pi_t(\cdot)$  é a função lucro, que depende naturalmente das variáveis de estado.  $i_t$  é o investimento que possui custo corrente  $c(i_t)$ ,  $\beta$  é o fator de desconto da firma, e por fim  $J_t$  representa o conjunto de informação disponível no momento  $t$ .

O operador max na equação (12.2) indica que uma firma compara o valor de venda da empresa ( $\Phi$ ) para deixar o mercado com o valor esperado descontado dos retornos de continuar operando. Como dito anteriormente, se a empresa optar por continuar no mercado então ela decide o nível de investimento (que não pode ser negativo).

**Equilíbrio:** Este modelo utiliza o *Equilíbrio Perfeito de Markov* de Ericson e Pakes (1995). EPM e estratégias é função das variáveis aleatórias relevantes de resultado (payoff). O modelo usado gera processos de Markov de primeira ordem mas as estratégias de Markov podem ser de qualquer ordem finita. A importância do equilíbrio é que a ação das firmas neste “setup” produz uma regra de política (escolha) para investimento e outra para saída do mercado.

As duas regras ou funções-política são as seguintes:

**Política de saída:** a regra de saída é  $\chi_t = \chi_t(k_t, a_t, \omega_t)$ . Se a firma continua, então supomos que ela tem produtividade correte acima de nível mínimo que ela suporta. Se a produtividade cai abaixo de determinado valor, a empresa opta por deixar o mercado e ficar com o valor da venda da firma ( $\Phi$ ). No caso da firma continuar:  $\chi_t = 1 \Leftrightarrow \omega_t \geq \underline{\omega}_t(k_t, a_t)$ , tal que  $\underline{\omega}_t(\cdot)$  é o nível mínimo de produtividade para permanecer no mercado. Observe que a função é indexada por  $t$ , que representa a estrutura de mercado. Além disso,  $\underline{\omega}_t(\cdot)$  é decrescente em  $k$ , que implica em relação negativa entre  $k$  e  $\omega$ .

**Regra de investimento:**  $i_t = i_t(k_t, a_t, \omega_t)$ . A função investimento depende da idade da firma, do capital acumulado e do nível da produtividade. A função é estritamente crescente em  $\omega_t$  se o investimento é positivo ( $> 0$ ). Observe que está função também é indexada por  $t$ . Esta função é crucial pois relaciona diretamente *investimento e produtividade: firmas mais eficientes investem mais.*

O modelo pode ser construído ou adaptado a partir de algumas opções. Primeiro, o pesquisador deve ser hábil para fazer hipóteses suficientes para calcular a regra de saída,  $\chi_t$ , e a de investimento,  $i_t$ . Em segundo lugar, não é preciso obter forma funcional precisa, utiliza uma estimação semi-paramétrica para controlar os não-observáveis.

Nestas escolhos podemos nos deparar com alguns trade-offs. Assumir modelagem semi-paramétrica invoca modelagem robusta à hipóteses sobre função lucro, regras de entrada e saída, taxação, etc. Se as hipóteses estão corretas e o custo computacional é gerenciável, as regras de saída e investimento podem: (i) fornecer estimativas mais eficientes, (ii) responder questões importantes. A decisão de usar controles não-paramétricos prioriza a identificação da produtividade às expensas de respostas na produção ou investimento.

## 12.2 Estimação de Função de Produção

Importante decidir se a análise de função de produção será realizada com valor adicionado ou produção bruta. A diferença aqui é o uso de materiais na estrutura (veja capítulo). Inicialmente a discussão é realizada utilizando valor adicionado.

O objetivo da estimação é a função de produção apresentada na equação (12.1) e duas abordagens similares são utilizadas para a estimação. O modelo seminal Olley-Pakes (1996) utiliza mínimos quadrados não-linear (NLLS, abreviação do inglês) e os trabalhos subsequentes utilizam versões alternativas utilizando variáveis instrumentais (GMM). Primeiro apresentamos a estimação por NLLS e sem seguida por GMM.

### 12.2.1 NLLS

A estimativa por MQO dos parâmetros da função de produção possui viés na estimativa do coeficiente de trabalho - adicionalmente está é uma boa razão para usar valor adicionado e não vendas. Um dos principais problemas neste tipo de análise, tanto sobre técnica quanto em estrutura dos dados, é o viés de seleção na estimativa.

O problema de seleção é o seguinte:  $E[\omega_t \mid \omega_t > \underline{\omega}(k_t), \omega_{t-1}, \chi_t = 1]$ . O valor esperado da produtividade depende dela estar acima do mínimo de produtividade que a mantém no mercado, condicional a produtividade no período anterior. Observe que esta função é decrescente em  $k_{t-1}$ . Ou seja, a função valor é crescente em ambos, então a curva de “isovalor” possuem inclinação negativa, e se  $k$  é maior ela ira continuar operando com produtividade mais baixa. Portanto, em estimação MQO existe um viés negativo sobre o parâmetro do estoque de capital. Por tudo que foi apresentado, existem diferentes motivos para estimar a função de produção por métodos alternativos.

Importante sublinhar que a estimativa é baseada no modelo comportamental. Dois pontos teóricos são cruciais. (i) Firmas mais produtivas investem mais; e (ii) firmas mais produtivas permanecem no mercado, enquanto as firmas menos produtivas deixam o mercado.

A estimativa da função de produção no artigo de Olley e Pakes (1996) possui um fator de produção denominado idade da firma,  $a_t$ , que visa capturar o aprendizado da firma no mercado. Portanto, a função de produção é a seguinte:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_a a_{it} + \omega_{it} + \varepsilon_{it}, \quad (12.6)$$

A estimativa da função de produção como proposta por Olley e Pakes (1996) possui três etapas. Em uma etapa é modelada a saída das firmas do mercado. Esta abordagem permite atacar o problema do viés de seleção. Em seguida é estimada função de produção para controlar pela dinâmica “do que não é” escolha da firma, i.e., choques não-observados. Utilizando os resultados das duas etapas anteriores, os parâmetros de interesse são estimados no terceiro estágio.

### Estágio 1: Política de Saída

Primeiro podemos estimar a regra de saída como descrito no modelo comportamental, seção 12.1. A probabilidade de uma empresa continuar depende da produtividade:

$$\text{Prob} \{ \chi_{t+1} = 1 \mid \underline{\omega}_{t+1}(\cdot), J_t \} \quad (12.7)$$

tal que  $J_t$  é o conjunto de informação em  $t$ . Utilizando a regra de saída como função do nível mínimo de produtividade que permite a empresa continuar operando no mercado, podemos escrever (12.7) como:

$$\begin{aligned} \text{Prob} \{ \omega_{t+1}(\cdot) > \underline{\omega}_{t+1}(\cdot) \mid \underline{\omega}_{t+1}(\cdot), J_t \} \\ = 1 - F(\underline{\omega}_{t+1} \mid \omega_t) = P_t(i_t, a_t, k_t). \end{aligned} \quad (12.8)$$

$P_t$  é estimado como um polinômio com interações das variáveis de interesse e estado  $a_t$  e  $k_t$ , bem como em relação investimento. Como mostramos no modelo comportamental, firmas mais produtivas investem,  $i_t = i_t(\omega_t)$ .<sup>1</sup> Deste estágio obtemos o valor previsto  $\hat{P}_t$ .

<sup>1</sup>Olley e Pakes (1996) utilizam um polinômio de quarta ordem.

**Estágio 2: Função Controle e  $\beta_l$** 

Se o investimento no período  $t$  é positivo, um condição determinada por  $J_{t-1}$ , então é possível pensar que a produtividade é positivamente relacionada com  $i_t$ . Isto é, no modelo comportamental a regra de investimento é

$$i_t = i_t(k_t, a_t, \omega_t) \quad (12.9)$$

implicando que firmas mais produtivas investem mais. Como discutido anteriormente, a relação investimento e produtividade é fundamental para o modelo comportamental e para a estratégia empírica. A idéia de Olley e Pakes é que a regra de investimento pode ser invertida e escrita com a produtividade positivamente relacionada com o investimento, i.e.

$$\omega_t = h_t(i_t, a_t, k_t). \quad (12.10)$$

Substituindo a equação da produtividade (12.10) na função de produção (12.6), temos:

$$y_{it} = \beta_0 + \beta_l l_{it} + \phi_t(i_t, a_t, k_t) + \varepsilon_{it} \quad (12.11)$$

onde

$$\phi_t(i_t, a_t, k_t) \equiv \beta_k k_{it} + \beta_a a_{it} + h_t(i_t, a_t, k_t), \quad (12.12)$$

observe que em (12.12) estão todas as variáveis de estado do modelo comportamental (veja (12.2)).

Na terminologia econométrica, a relação um-para-um entre investimento e produtividade faz de  $h_t(\cdot)$  uma *função controle*. Econometristas sugerem o uso de função controle para eliminar preocupações com endogeneidade. Neste, caso a abordagem de função controle gera um parcialmente linear para o coeficiente de interesse na equação (12.11). Esta abordagem de função controle tem como objetivo eliminar endogeneidade e não gerar novas interpretações.

Olley e Pakes utilizam a função controle para obter o valor previsto  $\hat{\phi}(\cdot)$ . Este valor previsto é obtido com uma função controle que é um polinômio de quarta ordem entre os termos de  $\phi_t(\cdot)$ . Por exemplo, Olley e Pakes utilizam  $\sum_{i=1}^4 (a_t^{4-i} k_t^i + i_t^{4-i} k_t^i + a_t^{4-i} i_t^i)$  no lado direito de (12.12) – observe que este é um exemplo e a ordem do polinômio pode mudar em diferentes aplicações.

Deste segundo estágio temos os valores estimados de:  $\hat{\phi}(\cdot)$ ,  $\hat{\beta}_l$  e  $(y_{it} - \hat{\beta}_l l_{it})$ .

**Estágio 3: Estimativa Não-Linear**

No terceiro estágio são os estimados os parâmetros remanescentes:  $\beta_k$  e  $\beta_a$ . Das equações (12.10) e (12.12) sabemos que  $\omega_{it} = \phi_t(i_t, a_t, k_t) - \beta_k k_{it} - \beta_a a_{it}$ . Reescreva (12.11) um período à frente e tome expectativas condicional à informação em  $t$  e a sobrevivência da firma em  $t + 1$  :

$$\begin{aligned} E[y_{it+1} - \beta_l l_{it+1} \mid a_{it+1}, k_{it+1}, \chi_{it+1} = 1] &= \\ &= \beta_0 + \beta_a a_{it+1} + \beta_k k_{it+1} + E[\omega_{it+1} \mid \omega_{it}, \chi_{it+1} = 1]. \end{aligned} \quad (12.13)$$

O termo  $E[\omega_{it+1} \mid \omega_t, \chi_{it+1} = 1]$  captura a dinâmica da produtividade que é crucial para o sucesso das firmas (investimento) e para a decisão de deixar o mercado. Como a dinâmica pode ser complexa, a saída proposta por Olley e Pakes é utilizar novamente uma análise de função controle. Portanto, o termo de dinâmica da produtividade por ser escrito como

$$E[\omega_{it+1} \mid \omega_t, \chi_{it+1} = 1] = \int_{\underline{\omega}_{it+1}} \omega_{it+1} \frac{F(d\omega_{it+1} \mid \omega_t)}{\int_{\underline{\omega}_{it+1}} F(d\omega_{it+1} \mid \omega_t)} \quad (12.14)$$

$$\equiv g(\underline{\omega}_{it+1}, \omega_t) = g(P_{it}, \phi_{it} - \beta_a a_{it} - \beta_k k_{it})$$

O termo (12.14) pode ser interpretado como um viés presente na estimação dos parâmetros. A função  $g(\cdot)$  é formada por duas variáveis de estado das firmas:  $\underline{\omega}_t[k_{t+1}(k_t, i_t), a_{t+1}(a_t)]$  e  $\omega_t$ . Para controlar pelo viés de seleção precisamos de uma medida da produtividade, que é  $\omega_{it} = \phi_t(i_t, a_t, k_t) - \beta_k k_{it} - \beta_a a_{it}$ , e uma medida do valor da produtividade que faz com que a firma seja indiferente entre continuar operando e sair do mercado. Este valor é  $\underline{\omega}$ . Dado que a densidade de  $\omega_{t+1}$  condicional sobre  $\omega_t$  é positiva em uma região sobre  $\underline{\omega}_{t+1}$ , a equação de seleção pode ser invertida para expressar  $\underline{\omega}_{t+1}$  como uma função de  $P_t$  e  $\omega_t$ . Portanto, escrevemos  $g(\cdot)$  como função de  $P_t$  e  $\omega_t$ .

Para completar a dinâmica da produtividade (12.14) é preciso definir um termo de inovação de  $\omega_t$ , i.e. a primeira diferença da produtividade:

$$\xi_{t+1} = \omega_{t+1} - \omega_t. \quad (12.15)$$

Desta forma podemos escrever a equação (12.13) como

$$y_{it+1} - \beta_l l_{it+1} = \beta_0 + \beta_a a_{it+1} + \beta_k k_{it+1} + g(P_{it}, \phi_{it} - \beta_a a_{it} - \beta_k k_{it}) + \xi_{it+1} + \eta_{it+1} \quad (12.16)$$

Importante observar que a presença do termo  $\xi$  obriga a utilizar o primeiro estágio para estimar  $\beta_l$ . Uma vez que o trabalho pode responder a  $\xi$ ,  $l$  deve ser função de  $\xi$ , e não podemos deslocar o fator trabalho para o lado direito da equação para estimar  $\beta_l$ . I.e. o trabalho não é média-independente do termo de distúrbio em (12.16).

**Implementação.** O terceiro estágio (12.16) é implementado com mínimos quadrados não-linear utilizando um polinômio de quarta ordem para o termo  $g(\cdot)$ . Aqui é necessário utilizar método não-linear pois os parâmetros a serem estimados aparecem em  $t + 1$  e em  $t$ . Dada as estimavas de

- $\hat{\beta}_l$  do segundo estágio forme  $y_{t+1} - \hat{\beta}_l l_{t+1}$ ;
- $\hat{\phi}_t$  do segundo estágio forme  $\hat{h}_t \equiv \hat{\phi}_t - \beta_a a_t - \beta_k k_t$ , e;
- $\hat{P}_t$  do primeiro estágio.

obtenha a equação para ser estimada (omitindo o indicador das firmas  $i$ )

$$y_{t+1} - \hat{\beta}_l l_{t+1} = \beta_0 + \beta_a a_{t+1} + \beta_k k_{t+1} + \quad (12.17)$$

$$+ \sum_{j=0}^{4-m} \sum_{m=0}^4 \beta_{m,j} (\hat{P}_t^m \cdot \hat{h}_t^j(\beta_0, \beta_k)) + e_t.$$

obtendo as estimativas de  $\hat{\beta}_0, \hat{\beta}_a, \hat{\beta}_k$ .

### Análise de Robustez

A idéia para a análise de robustez é a seguinte. Primeiro compare ao que foi feito anteriormente, e então veja o quanto foi feito o suficiente. Provavelmente a hipótese mais questionável é a inversão da regra de investimento. Portanto, a pergunta seria: o investimento é suficiente para controlar pela heterogeneidade?

Se não, então existe um erro no terceiro estágio igual a  $(\hat{\beta}_l - \beta_l)l_{t+1}$ . Não se pode utilizar  $l_{t+1}$  pois é correlacionado com  $\eta_{t+1}$ , mas se pode utilizar  $l_t$ . Portanto estime o terceiro estágio com  $l_t$

$$y_{t+1} - \hat{\beta}_l l_{t+1} = \beta_0 + \beta_a a_{t+1} + \beta_k k_{t+1} + \gamma l_t \quad (12.18)$$

$$+ \sum_{j=0}^{4-m} \sum_{m=0}^4 \beta_{m,j} (\hat{P}_t^m \cdot \hat{h}_t^j(\beta_0, \beta_k)) + e_t.$$

#### 12.2.2 GMM

Levinsohn e Petrin (2003) e Wooldridge (2009) mostram que o método de estimação de Olley e Pakes (1996) pode ser implementado utilizando variáveis instrumentais, como consequência obtendo resultados mais precisos. Todavia, a alternativa proposta por Wooldridge (2009) pode ser um pouco custosa devido ao número de momentos necessários para a estimação. Uma alternativa é a versão GMM de Olley e Pakes (1996) expandida por Akerberg, Caves, e Frazer (2015) (doravante ACF).

A hipótese fundamental do modelo Olley-Pakes é a relação estritamente crescente entre produtividade e investimento (Equação (12.9)). Uma alternativa a regra de investimento é a ideia de Levinsohn e Petrin (2003) de relacionar maior produtividade com maior demanda por insumos – firmas mais produtivas, que ganham mercado, compram mais insumos produtivos. Embora o insumo intermediário,  $m$ , não seja uma variável de estado no modelo comportamental, ACF usam a ideia desta relação monotônica entre produtividade e demanda por insumos intermediários. Excluindo da análise o termo de produção  $a$ , a regra de demanda por insumos de Levinsohn e Petrin (2003) é

$$m_{it} = m_t(k_{it}, \omega_{it}). \quad (12.19)$$

Neste caso, a produtividade,  $\omega$ , pode ser escrita como função do capital e dos insumos:

$$\omega_{it} = \varrho_t(k_{it}, m_{it}). \quad (12.20)$$

ACF fazem uma discussão importante e argumentam que o coeficiente do fator trabalho não pode ser adequadamente identificado no primeiro estágio de Olley e Pakes (1996).



Portanto, na interpretação dos autores as regras de investimento e insumo intermediário que se relaciona com a produtividade e capital deveria incluir o fator trabalho  $l$ . Desta forma, a função inversa da produtividade deve ser escrita como

$$\omega_{it} = \varrho_t(k_{it}, m_{it}, l_{it}). \quad (12.21)$$

A estimação da função de produção proposta por ACF segue a proposta seminal de Olley e Pakes (1996) e ideias dos artigos de Levinsohn e Petrin (2003) e Wooldridge (2009). O primeiro estágio de Olley e Pakes não é apresentado pois se entende que ele é correto e pode ser incorporado ao procedimento utilizado. Portanto, a estimação utilizada por ACF começa no segundo estágio de Olley e Pakes.

### Estágio 2: Função Controle

A função de produção log-linear, utilizando valor adicionado ( $y_{jt}$ ), a ser estimada é a seguinte:<sup>2</sup>

$$y_{jt} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \varrho_t(k_{it}, l_{it}, m_{it}) + \varepsilon_{it} \quad (12.22)$$

entretanto é preciso controlar pelo processo de  $\omega$ . ACF também argumentam que  $\beta_l$  não pode ser identificado neste estágio. Portanto neste estágio se estima

$$y_{it} = \phi_t(k_{it}, l_{it}, m_{it}) + \varepsilon_{it} \quad (12.23)$$

tal que  $\omega_{it} \equiv \varrho_t(k_{it}, l_{it}, m_{it})$ . O termo  $\phi_t(k_{it}, l_{it}, m_{it}) = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \varrho_t(\cdot)$  é estimado como um polinômio de quarta ordem:

$$\begin{aligned} \phi_t(k_{it}, l_{it}, m_{it}) = & \\ = \sum_{j=0}^4 \beta_{k,l,j,4-j} k_{it}^j l_{it}^{4-j} & + \sum_{j=0}^4 \beta_{k,m,j,4-j} k_{it}^j m_{it}^{4-j} + \sum_{j=0}^4 \beta_{m,l,j,4-j} m_{it}^j l_{it}^{4-j} & (12.24) \\ & + \beta_0 + \beta_k k_{it} + \beta_l l_{it} \end{aligned}$$

A estimativa deste estágio permite calcular o valor previsto da produção controlando pelo termo aleatório  $\varepsilon_{it}$ . Choques não esperados entram no resíduo deste estágio. ACF propõem estimar  $\beta_l$  em conjunto com os demais parâmetros.

Após esta estimação, salve o valor previsto de  $\Phi_t(\cdot)$ :

$$\hat{\phi}_t(k_{it}, l_{it}, m_{it}). \quad (12.25)$$

<sup>2</sup>Observe que esta equação é escrita para valor adicionado e o uso de materiais está relacionado com  $\omega$ , mas não é um fator produtivo como em uma função com produção bruta.

**Estágio 3: Estimativa Não-Linear**

É importante para a estimativa proposta por ACF a equação de inovação (variação) da produtividade ( $\omega$ ) – um pouco diferente da utilizada em Olley e Pakes (ver equação 12.15):

$$\begin{aligned}\omega_{it} &= E[\omega_{it} | I_{it-1}] + \xi_{it} \\ \omega_{it} &= E[\omega_{it} | \omega_{it-1}] + \xi_{it}\end{aligned}\quad (12.26)$$

Aqui é importante o conjunto informação deste estágio. ACF assumem que a informação conhecida em  $t$  é apenas relativa a  $t - 1$  e  $E[\xi_{it} | I_{it-1}] = 0$ . ACF e Levinsohn e Petrin (2003) assumem que a produtividade corrente ( $t$ ) depende da passada por uma relação  $g(\cdot)$  em adição à inovação  $\xi$ :

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}. \quad (12.27)$$

Dada a hipótese sobre a dinâmica de  $\omega$  (equação 12.27), podemos fazer a substituição na função de produção (12.6) (sem a variável idade da firma,  $a$ ):

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + g(\omega_{it-1}) + \xi_{it} + \varepsilon_{it}, \quad (12.28)$$

sabendo que em  $t - 1$ ,  $\omega_{it-1} = \phi_{it-1}(\cdot) - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1}$ , então a função (12.28) pode ser escrita como

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + g(\phi_{it-1}(\cdot) - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1}) + \xi_{it} + \varepsilon_{it}. \quad (12.29)$$

Tomando expectativas em (12.29) e sabendo que  $E[\xi_{it} | I_{it-1}] = 0$  e  $E[\varepsilon_{it} | I_{it}] = 0$  (que implica em  $E[\varepsilon_{it} | I_{it-1}] = 0$ ), a equação de segundo estágio pode ser escrita como:

$$\begin{aligned}E[\xi_{it} + \varepsilon_{it} | I_{it-1}] &= \\ = E[y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} - g(\phi_{it-1}(k_{jt-1}, l_{jt-1}, m_{jt-1}) - \beta_0 - \beta_k k_{jt-1} - \beta_l l_{jt-1}) | I_{it-1}] \\ &= 0\end{aligned}\quad (12.30)$$

**Conjunto informação.** Como o estoque de capital segue a lei de movimento do estoque de capital (ver 12.3), o capital no período  $t$  é função do investimento em realizado em  $t - 1$ . Portanto  $k_{it} \in I_{t-1}$ . Então

$$k_{it} = (1 - \delta)k_{it-1} + i_{it-1}, \text{ e } E[\xi_{it} | I_{it-1}] = 0,$$

que implicam no momento

$$E[\xi_{it} k_{it}] = 0. \quad (12.31)$$

ACF argumentam que é necessário mais um momento, pois é preciso estimar o parâmetro  $\beta_l$ . Como  $l_{it-1} \in I_{it-1}$ , os autores sugerem

$$E[\xi_{it} l_{it-1}] = 0.$$

Além disso, se for preciso estimar  $\beta_0$  e  $g(\cdot)$  no terceiro estágio são necessários mais dois momentos.<sup>3</sup> ACF utilizam para  $\beta_0$   $E[\xi_{it}1] = 0$  e para  $g(\cdot)$  eles propõem  $E[\xi_{it}\phi_{it-1}] = 0$ .

Portanto, a condição de momentos para o segundo estágio de ACF pode ser escrita como:

$$E \left[ (\xi_{it} + \varepsilon_{it}) \otimes \begin{pmatrix} 1 \\ k_{it} \\ l_{it-1} \\ \phi_{it-1}(\cdot) \end{pmatrix} \right] = 0 \quad (12.32)$$

tal que

$$\xi_{it} + \varepsilon_{it} = y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} - g(h_{it-1})$$

e

$$h_{it-1} \equiv \phi_{t-1}(k_{it-1}, l_{it-1}, m_{it-1}) - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1}. \quad (12.33)$$

Observe que este modelo é exatamente identificado. Para uma generalização veja o anexo 12.3.1.

### Caso $\omega$ -AR(1)

ACF estimam a equação (12.30) utilizando o processo da produtividade, equação (12.27), tendo estrutura AR(1):

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it}.$$

Nesta equação,  $\rho$  é o parâmetro de persistência da produtividade (ou o termo AR(1)). Cabe ressaltar que Olley e Pakes estimam supondo estrutura de primeira diferença, quando  $\rho = 1$ . Portanto, a inovação,  $\xi$ , na produtividade é:

$$\xi_{it} = \omega_{it} - \rho\omega_{it-1} \quad (12.34)$$

No caso de AR(1),  $g(\cdot)$  é o parâmetro  $\rho$  e se pode utilizar o momento apresentado na equação (12.32). Neste caso a equação (12.32) passa a ser:

$$E \left[ (\xi_{it} + \varepsilon_{it}) \otimes \begin{pmatrix} 1 \\ k_{it} \\ l_{it-1} \\ \widehat{\phi}_{it-1}(k_{it-1}, l_{it-1}, m_{it-1}) \end{pmatrix} \right] = 0 \quad (12.35)$$

tal que

$$\xi_{it} + \varepsilon_{it} = y_{it} - \beta_0 - \beta_k k_{jt} - \beta_l l_{jt} - \rho(h_{it-1})$$

e  $h_{it-1}$  como definido em (12.33).

<sup>3</sup>Mais de dois momentos a depender do número de parâmetros em  $g(\cdot)$ .

**Implementação.** Para reduzir o número de momentos, ACF estimam  $\beta_0$  no primeiro estágio,  $\tilde{\beta}_0$ , e o termo de persistência da produtividade,  $\rho$ , é estimado em separado como uma regressão AR(1).

Para um guess hipotético de  $(\beta_k, \beta_l)$  eles constroem o seguinte termo:

$$\tilde{\beta}_0 + \widehat{\omega_{it}(\beta_k, \beta_l)} = \hat{\phi}_t(k_{it}, l_{it}, m_{it}) - \beta_k k_{it} + \beta_l l_{it},$$

o termo  $\hat{\phi}_t(k_{it}, l_{it}, m_{it})$  é estimado no primeiro estágio.  $\phi_t$  é usado sem o  $i$ ? média sobre empresas?

Uma vez formado o termo acima é feita a estimação de

$$\tilde{\beta}_0 + \widehat{\omega_{it}(\beta_k, \beta_l)} = \rho(\tilde{\beta}_0 + \widehat{\omega_{it}(\beta_k, \beta_{t-1})}) + \xi_{it}(\widehat{\beta_k, \beta_l})$$

observe que o resíduo desta regressão é  $\xi_{it}(\widehat{\beta_k, \beta_l})$ . Este resíduo é utilizado na equação principal

$$\mathbb{E} \left[ (\hat{\xi}_{it} + \varepsilon_{it}) \otimes \begin{pmatrix} k_{it} \\ l_{it-1} \end{pmatrix} \right] = 0 \quad (12.36)$$

A estimativa de  $\rho$  é dada por e utilizado de forma iterativa para  $(\hat{\beta}_k, \hat{\beta}_l)$

## Anexo

### 12.3 Momentos na Estimação

#### 12.3.1 ACF

**Segunda etapa:**  $\mathbb{E}[\varepsilon_{it} | I_{it}] = \mathbb{E}[y_{it} - \Phi_t(k_{it}, l_{it}, m_{it}) | I_{it}] = 0.$

**Terceira etapa:**  $\mathbb{E}[\xi_{it} + \varepsilon_{it} | I_{it-1}] = \mathbb{E}[y_{jt} - \beta_0 - \beta_k k_{jt} - \beta_l l_{jt} - g(\widehat{\Phi_{t-1}}(k_{jt-1}, l_{jt-1}, m_{jt-1})) - \beta_0 - \beta_k k_{jt-1} - \beta_l l_{jt-1} | I_{it-1}] = 0$

#### $l_{it}$ como variável de estado

ACF apresentam o caso em que  $l_{it}$  é uma variável de estado. Em algumas indústrias, o fator trabalho pode ser escolhido pela firma em  $t - 1$ , isto é,  $l_{it} \in I_{it-1}$ . Esta hipótese também pode significar que a produtividade não é conhecida pela firma até o período  $t + 1$ . Esta pode ser uma hipótese muito forte, mas pode ser plausível em situações onde existem custos significativos para contratação e demissão, ou outro tipo de rigidez no mercado do trabalho (caso do Brasil?). Neste caso, deveríamos adicionar  $l_{it}$  ao conjunto

## CAPÍTULO 12. PRODUTIVIDADE E FUNÇÃO DE PRODUÇÃO (\*)221

informação ou alternativamente substituir  $l_{it-1}$  por  $l_{it}$  como instrumento em (12.32). O modelo (12.32) sobre-identificado passa a ser:

$$E \left[ (\xi_{it} + \varepsilon_{it}) \otimes \begin{pmatrix} 1 \\ k_{it} \\ l_{it} \\ l_{it-1} \\ \phi_{it-1}(\cdot) \end{pmatrix} \right] = 0 \quad (12.37)$$

ACF (p. 2431) argumentam que esta hipótese mais forte geralmente leva a estimativas mais precisas.



## Capítulo 13

# Conluio

### 13.1 Introdução

Porque existe conluio?

Em qualquer estrutura de oligopólio o lucro total é menor do que o lucro de monopólio – assumindo equilíbrio de mercado. Este é o resultado da concorrência, onde a firma apenas maximiza seu próprio lucro. Portanto as empresas podem incorrer em práticas anti-concorrenciais para elevar o lucro total da indústria. Este tipo de comportamento é genericamente denominado de *conluio*.

Acordo de *Cartel* é um tipo de conluio. Um exemplo clássico de comportamento de cartel é o aumento de preços do barril de petróleo patrocinado pela OPEP em 1973.

O conluio pode ser explícito ou tácito. Em uma leitura simplista no conluio clássico as firmas realizam comunicação direta e coordenação clara para auferir lucro extraordinário, enquanto que no conluio tácito as firmas tem a capacidade de coordenar ações sem a ajuda de comunicação direta.

Desde o Sherman Act (1890) os cartéis são ilegais nos EUA. Isto implica que uma empresa não pode aderir legalmente a um contrato colusivo.<sup>1</sup> Como prescrito pela lei antitruste os cartéis são um crime “per se”, o que significa que não é necessário gerar malefícios a partir destes acordos.

Observar que o tratamento de informações diferente de preços e coordenação de investimentos (e.g. JV) pode não ser considerado ilegal. Isto é verdade para EUA e Brasil. Estas “coordenações” são tratadas sobre a “regra da razão” para saber se são ilegais ou não.

Contratos: Como cartéis são ilegais os contratos não podem ser “tradicionais”, mas existe algum tipo de contrato para a estabilidade do conluio. Em outras palavras, para

---

<sup>1</sup>No Reino Unido cartéis foram legais até os anos 50. No Brasil passam a ser ilegais nos anos 60, mas a primeira condenação é apenas nos anos 2000.

o funcionamento do cartel deve ocorrer “enforcement” para garantir operacionalidade. Desta forma deve haver um conjunto de incentivos para os seus membros permanecerem no acordo. O ponto central é que o valor do fluxo de caixa futuro do cartel deve ser maior quando o acordo está operacional. Importante que isso deve ser verdade em todos os estados ou pelo menos em todos os estados que observamos o cartel operando.

A ênfase inicial no estudo de cartel é no equilíbrio e estabilidade dos acordos.

Whiston (2006, pp. 15-19) faz uma interessante discussão sobre o uso da regra “per se” vs regra da razão. Quase toda empresa que entra em acordo de cartel possui alguma justificativa social para o conluio.

Vários acordos de cartel para determinar preço acima do nível competitivo são na verdade acordos que determinam produção (quantidade).

Não é raro que empresas que realizem algum tipo de acordo colusivo defendam do ponto de vista social os benefícios destas práticas. Whiston (2006, pp. XX) cita alguns exemplos desta defesa.

## 13.2 Jogos Repetidos

### Estabilidade dos Acordos

De forma introdutória explicitamos os fundamentos para definição de equilíbrio e estabilidade de um acordo colusivo. Como acordo de cartel é um tentativa de coordenação para melhorar conjuntamente o payoff das empresas então este é um problema dinâmico. Além disso, em todo período ou estado os acordos precisam ser validados tacitamente ou explicitamente pelos participantes. A abordagem mais simples, para evitar grandes complicações de análise dinâmica, é tratar dois estados possíveis:

1. Estado competitivo: sem acordo de conluio;
2. Estado conluio: os incentivos de conluio funcionam e existe cartel.

A questão colocada aqui é como modelar acordos colusivos em teoria dos jogos. Esta modelagem tem como objetivo reproduzir o comportamento real dos mercados, bem como entender como funcionam os mecanismos (a dinâmica) dos acordos colusivos.

### Jogos Repetidos

Um ponto inicial de pesquisa é utilizar idéias de jogos repetidos. Como empresas que realizam acordos de conluio estão conjuntamente operando no mercado um ponto de partida inicial é pensar em alguma ideia de jogos repetidos. A contribuição da análise dinâmica se resume em dois pontos (como em Pakes, 2019a):

1. considerando investimento projetado para melhorar (incrementar) as variáveis de estado que mudam as primitivas do problema (investimento para entrada e investimento projetado para mudar a demanda e as funções custo das firmas), e;



2. começar realizando hipóteses simplificadoras de que valores futuros das variáveis de estado são independentes das escolhas de preço condicionais sobre várias decisões de investimento. Este tipo de hipótese pode ser relaxada depois, mas apenas na forma que permita a escolha do preço e da quantidade de hoje ter impacto nas variáveis de estado de amanhã por meio de seus efeitos sobre as primitivas de demanda e custo.

Outro ponto de partida de análise teórica em jogos é baseada em jogos repetidos (como dissemos acima). As hipóteses simplificadoras deste *setup* poderiam ser:

- valores futuros das variáveis de estado que determinam a demanda e as condições de custo nunca mudam, então as primitivas são do problema são repetidas em todo período;
- embora os estados dos jogos nunca mudam, os preços podem mudar. Nesse caso seria permitido que as escolhas do preço corrente (ou quantidades) serem uma função das escolhas passadas de preços (ou quantidade).

As decisões repetidas são usualmente pensadas como centrais para suportar/permitir colusão por meio do arcabouço institucional apropriado, mas não seriam adequadamente capturadas por um jogo repetido. A teoria de modelos repetidos são exercícios planejados para fornecer intuição de como acordos colusivos funcionam.

### Fundamento simples de jogo repetido

Aqui é apresentado uma versão simples de modelo de jogo repetido onde é colocada a lógica de um mecanismo de funcionamento de cartel. A exposição segue Pakes (2019a), Cabral (2017, cap. 9) e Tirole (1988).

Hipóteses:

- Duas firmas idênticas,  $i = 1, 2$ , interagem repetidamente em um mercado de bens homogêneos com custo constante.
- Firmas competem em preços  $p$  – quando determinam os preços correntes elas conhecem TODOS os preços escolhidos previamente.

A história do jogo será todo o conjunto de informação relevante para ambas as firmas. É representado por:

$$\mathfrak{S}_t \equiv \{p_{1,\tau}, p_{2,\tau}\}_{\tau=0}^{t-1} \quad (13.1)$$

e o preço corrente:

$$p_{i,t} : \mathfrak{S}_t \rightarrow \mathfrak{R}_+ \quad (13.2)$$

tal que  $t$  é o tempo. O mais importante para se observar são as implicações da amplitude do espaço de estratégias. Nesse *setup* o equilíbrio de Nash  $p = mc$  é uma das possibilidades ( $mc$  é o custo marginal). Outros equilíbrios possíveis são:

- Considere um preço limite: uma firma sempre joga o preço igual ou acima do preço limite se a outra também o faz. A firma joga  $p = mc$  para sempre se a outra

empresa jogar alguma vez preço menor do que o preço limite. O quanto ou não isto será um equilíbrio depende se o ganho ao desviar do preço limite em determinado período será maior do que o resultado deste desvio nos anos seguintes.

- O jogo poderia começar com  $p = mc$  e permitir que uma firma determine preços um pouco maior em um período (um ano), no período seguinte a outra firma poderia responder cooperativamente aumentando o seu preço um pouco acima do aumento realizado pela outra firma. A questão seria saber se isto seria um equilíbrio e qual seria o preço limite para estas empresas.

### Jogo Repetido Finito

Quando o horizonte do jogo é finito não existe conluio. Pensando em solução retroativa, se pode ocorrer desvio e o resultado do desvio é  $p = mc$  então as duas firmas jogam  $p = mc$  desde o primeiro período.

### Jogo Infinitamente Repetido

Considere a seguinte estratégia:

1.  $p_{i,0} = p^m$  para  $i = 1, 2, \dots$  tal que  $p^m$  é o preço de monopólio, e
2.  $p_{i,t}(\mathfrak{S}_{t-1}) = \begin{cases} p^m & \text{se para todo } \tau < t, p_\tau = (p^m, p^m) \\ p_{i,t}(\mathfrak{S}_{t-1}) = mc & \text{caso contrário} \end{cases}$

Isto às vezes é chamada de estratégia “grim Nash reversion”. As firmas coludem com o preço elevado se ninguém desvia no passado. Se ocorrer desvio, a firma joga o preço mais baixo para sempre.

Aqui o jogo é repetido (novamente), então se o equilíbrio colusivo puder ser suportado em qualquer período, ele será suportado para sempre. Se ocorrer algum desvio, então veremos  $mc$  para sempre.

Observe que o quanto o preço de monopólio será suportado ou não depende do que ocorre quando se desvia. Se ele é suportado pode ser devido a uma situação que nunca ocorre.

Neste exemplo a sustentabilidade do preço  $p^m$  depende se o valor presente dos lucros é superior ao lucro de desviar do acordo colusivo. O lucro de cada firma com conluio é  $\pi^m = (p^m - mc)q(p^m)$ . Se ambas as firmas se mantêm em conluio, então a expectativa de payoff descontado da firma 1 é:

$$\frac{1}{2}\pi^m + \delta\frac{1}{2}\pi^m + \delta^2\frac{1}{2}\pi^m + \dots \quad (13.3)$$

tal que  $\delta$  é o fator de desconto, isto é, o valor de 1 dólar no futuro comparado com 1 dólar no presente. Simplificando (13.3), temos:

$$V = \frac{1}{2}\pi^m \frac{1}{1-\delta} = \frac{1}{2} \frac{1}{1-\delta} (p^m - mc)q(p^m) \quad (13.4)$$

No caso mais simples, se a firma 1 desvia do conluio  $p_1 \neq p^m$  em algum período  $t$ , então o seu payoff futuro será zero. Por hipótese, ambas as firmas revertem para  $p = mc$ . Neste caso, o melhor desvio para a firma 1 é o que maximiza o lucro de curto prazo. O preço que maximiza o lucro de curto prazo da firma 1 é  $p^m - \epsilon$ , tal que  $\epsilon$  é um número pequeno. Quando a firma 1 cobra menos do que a firma 2, ela captura toda a demanda (produto homogêneo) e coleta lucros totais de  $V' = (p^m - \epsilon - mc)q(p^m - \epsilon)$ .

Então a condição para existir conluio é de valor descontado suficientemente alto em comparação desvio do conluio. Neste caso existe um equilíbrio de Nash que sustenta o cartel quando  $V \geq V'$ , isto é:

$$\frac{1}{2} \left[ \frac{1}{1 - \delta} \right] (p^m - mc)q(p^m) \geq (p^m - \epsilon - mc)q(p^m - \epsilon) \quad (13.5)$$

Esta condição apenas será satisfeita para todo  $\epsilon > 0$  se somente se  $\delta > 1/2$ . Isto é, se as firmas são suficientemente pacientes.

### O fator de desconto

Neste exemplo, a estabilidade do conluio depende apenas do fator de desconto. Geralmente o fator de desconto está entre 0 e 1, e existem várias razões para  $\delta < 1$ . A condição (13.5) também pode garantir que pode existir conluio a qualquer preço. Isto é, poderíamos substituir  $p^m$  por qualquer preço  $p^m > p > mc$  e a colusão deveria ser sustentada.

Estas não são propriedades gerais da solução, mas são propriedades resultantes das hipóteses utilizadas neste exemplo. Isto significa que algumas vezes podemos sustentar colusão com preço mais baixo do que monopólio para certo  $\delta$ , mas não podemos sustentar o preço de monopólio com este  $\delta$ .

Para um exemplo simples assuma que pode ocorrer conluio entre a firma 1 e 2 ao preço  $p \in [mc, p^m]$ . Neste caso as firmas dividem os lucros, i.e.  $1/2\pi(p)$ . Se alguma firma desvia então ela recebe  $\pi^n$ ; aqui  $\pi^n$  não precisa ser zero (i.e. as firmas podem reverter para Nash em quantidades). Então a condição de não desvio do conluio é:

$$\frac{1}{2} \frac{\pi(p)}{1 - \delta} > \pi(p - \epsilon) + \frac{\delta}{1 - \delta} \pi^n \quad (13.6)$$

que será satisfeita para todo  $\epsilon > 0$  dado que

$$f(\delta) > \frac{\pi^n}{\pi(p)} \quad (13.7)$$

com a função  $f(\cdot)$  crescente em  $\delta$ . Aqui deve ser claro que diferentes combinações de  $(\delta, \pi^n)$  irão sustentar diferentes intervalos de preços colusivos. O argumento geral é que *quanto maior for  $\delta$  e maior for a punição (menor)  $\pi^n$  então maior é o preço colusivo que pode ser suportado*.

Observe que nesse modelo sempre vemos a mesma jogada (comportamento). Se existe cooperação inicialmente, os jogadores sempre cooperam. Se eles pagam para desviar, então o payoff é revertido para  $\pi^n$  para sempre. Se existir uma punição mais dura, então:

- permite sustentar um preço colusivo mais elevado e ninguém irá desviar da “trajetória de equilíbrio;”
- um *modo* de punição mais dura nunca custa nada para qualquer player.

Conseqüentemente se existe a escolha para lucros no modo de reversão (para maior de  $\pi^n$ ) então se quer a punição mais dura possível. Isto não precisa ser verdade se na realidade observamos desvios.

Alguns fatores são importantes para se relacionar ou influenciar a taxa de desconto intertemporal.

***J* firmas.** Se existem  $J$  firmas realizando conluio, então o retorno para conluio a qualquer preço são baixos:  $(1/J)\pi(p)$  ao invés de  $(1/2)\pi(p)$ , mas os retornos para desviar são os mesmos. Conseqüentemente é mais difícil manter conluio. Neste novo contexto é necessária taxa de desconto mais elevada para permitir conluios. No caso mais simples com  $\pi^n = 0$ , para manter o conluio a condição é  $\delta > (J - 1)/J$ .

**Crescimento do mercado.** Se o mercado está crescendo é mais fácil para colusão, uma vez que os ganhos do conluio (que estão no futuro) são maiores relativos ao ganho do desvio (que estão no presente). Ainda assim no modelo simples, onde nada é ganho durante a punição, se o crescimento é constante, ele multiplica a taxa de desconto. A taxa de desconto é calculada como função da taxa de juros, então quando se tem crescimento  $g$  a taxa de desconto pode ser  $\delta = \frac{1+g}{1+r}$ , tal que  $r$  é a taxa de juros.

**Colapso do mercado.** Se existe alguma probabilidade de colapso do mercado isto reduz o incentivo para colusão. Isto faz o desvio do conluio mais lucrativo em relação ao valor esperado da cooperação.

**Saída.** Se existe opção e custo de saída isto reduz o incentivo para conluio por duas razões: (i) é mais difícil punir um competidor que pode sair, (ii) se pode ter um incentivo “predatório” para reverter a punição, induzir o competidor a sair, e mantém uma única empresa monopolista a partir deste ponto. Entretanto, isto implica em um jogo dinâmico - o que é mais complicado.

**Contato multimercado.** Se as firmas interagem em diversos mercados, então se pode pensar em punição em todos os mercados se a firma desviar em algum deles. Às vezes somente é possível cobrar cooperação se existe alguma conexão multimercado. Particularmente o conjunto de alocações sustentáveis não pode ser menor em contato multimercado do que contato em um único mercado.

### 13.2.1 Firmas diferentes

O exemplo com firmas diferentes é mais complicado devido a diferentes capacidades, custos marginais ou produtos diferenciados. Mesmo em jogo repetido de acordo colusivo dois aspectos precisam ser tratados:

1. Os preços ou quantidades precisam ser determinados e;

## 2. A divisão subsequente dos lucros.

A divisão dos lucros deve ser tal que toda firma deseje manter o acordo colusivo (a não ser que se mantenha um acordo parcial). In particular se não existirem outros meios de transferir lucros de um agente para outro (o que será ilegal), os preços devem ser determinados para gerar lucros que compensem a todas as empresas participar do conluio. A participação nos lucros irá depender do quanto cada firma pode receber *ao deixar* o regime colusivo e da perda que o cartel pode impor a firma que desvia. Por exemplo, se uma firma pequena possui a capacidade de “inundar o mercado” devido a seu excesso de capacidade então essa firma pode ser perigosa a sustentabilidade do cartel, logo ela deveria receber uma fatia nos lucros proporcionalmente maior ao seu tamanho.

Uma idéia de modelo com firmas diferentes pode usar a seguinte ideia. Suponha uma solução “barganha de Nash”, com lucros do conluio vs lucros resultantes da punição:

$$\max_{\{p_1, \dots, p_n\}} \prod_{i=1, \dots, n} [\pi_i(p_i, P_{-i}) - \pi_i(p_i^p, p_{-i}^p)] \quad (13.8)$$

tal que  $p_i^p$  é o preço de *punição* da firma  $i$ .

Ideia alternativa de “tentação balanceada” de Friedman (1971). A ideia básica é que o acordo colusivo deve (ao menos implicitamente) resultar em possíveis três funções de lucro para cada firma:

- $\pi^d(\cdot)$ : lucro quando a firma desvia;
- $\pi^c(\cdot)$ : lucro quando a firma faz conluio;
- $\pi^p(\cdot)$ : lucro quando a indústria está punindo as firmas que desviam.

Pela igualdade de incentivos para desviar, a regra implica que

$$\frac{\pi^d(\cdot) - \pi^p(\cdot)}{\pi^d(\cdot) - \pi^c(\cdot)}$$

é a mesma entre firmas. Neste caso se garante que os lucros estão em uma *fronteira de Pareto*. Este conceito implica que o vetor de lucros sujeito a esta restrição (ativa) não pode haver aumento de lucros de uma firma sem descrever a de outra.

Observe que: (i) o lucro mais baixo é na situação de punição (equilíbrio sem conluio), mas não se sabe se equilíbrio colusivo é sustentável; (ii) em ambos esquemas existem múltiplos equilíbrios colusivos possíveis.

Duas características dos modelos simples. As leis antitruste não permitem que as empresas escrevam contratos que permitam preços colusivos. Portanto, a idéia geral é a de que se os preços são diferentes do equilíbrio estático, então eles devem ser “enforced” por um equilíbrio de Nash perfeito de subjogos. Por outro lado este tipo de solução possui dois problemas:

- Existe um conjunto de equilíbrio colusivo, mas não há forma de escolher entre eles.
- O modelo sempre gera uma de duas soluções e a escolhida permanece para sempre. Particularmente este tipo de modelo não permite guerra de preços e este acontecimento é frequentemente observado.

### 13.3 Demanda Flutuante

Questão frequente em modelos de conluio: é mais provável conluio em tempos de prosperidade ou em tempos ruins? Se as firmas fazem conluio, como os preços flutuam com a demanda?

Um equilíbrio de conluio deve ser tal que as firmas não possuem incentivo para cobrar preço menor do que o rival. Isto é, a diferença entre lucros futuros do conluio e lucros futuros em caso de guerra de preços deve ser suficientemente grande para evitar que uma empresa opte pelos lucros do desvio do conluio.

Uma das primeiras discussões formais é devido a Rotemberg e Saloner (1986). Vamos manter a hipótese de indústria de produtos homogêneos com concorrência em preços, mas assumindo que existem choques iid de demanda. Por simplicidade assumamos que assumamos dois valores:

$$D_2(p) > D_1(p) \forall p. \quad (13.9)$$

Em cada período o estado da demanda é conhecido antes de se escolher os preços.

Novamente a questão é se podemos sustentar preços de equilíbrio maior do que o custo marginal. Assim olhamos para um equilíbrio que não é Pareto dominado por outro payoff de equilíbrio (no sentido de que ambas as firmas preferam ele).

Uma vez que o choque de demanda é conhecido antes da determinação de preços de cada firma, os lucros descontados para cada firma para qualquer dois preços é dado por

$$\begin{aligned} V &= \sum_{\tau=0}^{\infty} \delta^{\tau} (1/4) [D_1(p_1)(p_1 - c) + D_2(p_2)(p_2 - c)] = & (13.10) \\ &= \frac{[D_1(p_1)(p_1 - c) + D_2(p_2)(p_2 - c)]}{4(1 - \delta)} \end{aligned}$$

Sabemos que se existe habilidade para cumprir qualquer acordo este deverá ser garantido com o máximo de punição - assumamos que é lucro zero para sempre. Então permanecemos presos às punições máximas e começamos com a questão do quanto podemos fazer cumprir preços de monopólio em cada período.

A distribuição do futuro é sempre a mesma, não importando o estado presente. Então o valor da punição é o mesmo para o estado atual. Então o valor da punição é o mesmo para o estado corrente. Uma vez que os lucros do desvio do conluio são maiores no estado de maior demanda, desvio pode ser esperado quando a demanda é mais elevada.

Comparando os lucros do desvio com a punição, o requerimento para um acordo de conluio ser cumprido é

$$\frac{\pi_2^m}{2} < \left[ \frac{\delta}{1-\delta} 4 \right] [\pi_1^m + \pi_2^m] \quad (13.11)$$

ou seja,

$$\delta > \delta_0 \equiv \frac{2\pi_2^m}{[\pi_1^m + 3\pi_2^m]} \quad (13.12)$$

Neste caso particular  $\delta_0$  está entre  $1/2(\pi_1 = \pi_2)$  e  $2/3(\pi_1 = 0)$ .

Assuma agora que  $\delta$  estava entre  $1/2$  e  $\delta_0$  (i.e. podemos não suportar preços de monopólio em períodos de demanda elevada). Então escolhemos  $(p_1, p_2)$  que

$$\max_{\{p_1, p_2\}} (1/4) \frac{\pi_1(p_1) + \pi_2(p_2)}{1-\delta} \quad (13.13)$$

s.a.

$$\pi_1(p_1)/2 \leq (1/4) \frac{\pi_1(p_1) + \pi_2(p_2)}{1-\delta}$$

e

$$\pi_2(p_2)/2 \leq (1/4) \frac{\pi_1(p_1) + \pi_2(p_2)}{1-\delta}$$

As duas condições podem ser rescritas como

$$\pi_1(p_1) \leq K\pi_2(p_2)$$

e

$$\pi_2(p_2) \leq K\pi_1(p_1)$$

tal que  $K \equiv \delta(2 - 3\delta)$ .

Aqui se está mais preocupado com as últimas restrições. Para qualquer  $p_1$  escolha  $p_2$  que maximize lucros. Mas então esta solução nos fornece funções objetivo que aumentam em  $p_1$  até atingir  $p_1^m$ . Então determine:

$$p_1 = p_1^m \text{ e } p_2 \text{ s.a. } \pi_2(p_2) = K\pi_1(p_1^m) \quad (13.14)$$

Então se cobra o preço de monopólio no trimestre em que a demanda está baixa, e alguma coisa abaixo do preço de monopólio quando a demanda está elevada. Isto não significa que um preço seja maior do que o outro. Tudo depende de qual o preço de monopólio do trimestre (o nível e a inclinação da curva de demanda podem ser diferentes).

Mesmo não sendo clara a diferença de preços entre os períodos, os autores interpretam este resultado como determinação de preços “contra-cíclico” e eles referem a esta mudança de preços como “price-wars.” Existe um bom número de artigos que mostram que é mais difícil manter acordo colusivo durante boom do que em baixa de demanda, embora este resultado não seja homogêneo entre indústrias.

Em termos gerais, mantendo tudo igual, é mais difícil manter conluio durante fase positiva do ciclo econômico do que na fase negativa, porque o incentivo para abandonar

o cartel é maior no boom. Isto é verdade se o período de crescimento não for seguido de um período ainda maior. Este tipo de situação não é prevista no modelo, pois o resultado depende fortemente da hipótese i.i.d. dos choques, que não é considerada ser realista.

Note também que bons tempos (boom) é em termos de boa realização da demanda. O mesmo argumento aqui poderia ser aplicado aos insumos da indústria. Isto é, não poderíamos garantir conluio a medida que o tamanho das margens oscila quando os preços dos insumos estão altos ou baixos (choques do petróleo afetando estabilidade de cartel no varejo dos combustíveis).

### 13.3.1 Demanda Cíclica

Haltiwanger e Harrington (1991) substituem a hipótese de choques de demanda i.i.d por movimentos previsíveis. Estes movimentos podem ser o comportamento cíclico da economia ou apenas flutuações sazonais. Os períodos não são diferentes apenas no retorno do desvio do conluio (como em Rotemberg-Saloner), mas porque os períodos diferentes podem possuir futuros distintos, a perda quando se deixa um acordo colusivo será diferente entre os estados do futuro.

Esse modelo é baseado nos ciclos de demanda determinísticos. As curvas de demanda são crescentes (em todo  $p$ ) até  $\hat{t}$  e então decrescentes até o ciclo ser completo em  $\bar{t}$  (não seria complicado adicionar choques i.i.d. a estas curvas de demanda). A abordagem permanece a mesma do jogo repetido, com as primitivas permanecendo sempre as mesmas (número de firmas, f. custo, etc.).

Como no jogo de bens homogêneos existe um oligopólio de firmas simétricas que vive para sempre com competição em preços. As punições são máximas – punição com lucro zero para sempre. O equilíbrio que as firmas escolhem é a trajetória de preços que é sustentável por um equilíbrio perfeito de subjogos. A condição de sustentação do equilíbrio em qualquer ponto é que o valor da punição ( $P(t)$ ) deve ser maior do que o valor do desvio ( $D(t)$ ) para todo  $t$  no ciclo, ou

$$P(t) \equiv \sum_{\tau=t+1} \delta^{\tau-t} [(p(\tau) - c)D(p(\tau); \tau)] / n \geq \quad (13.15)$$

$$(n - 1) [(p(t) - c)D(p(t); t)] \equiv D(t)$$

Observe aqui que  $t$  é indicadora do período do ciclo. A derivação do equilíbrio depende apenas do valor de  $\delta$ :

- Se  $\delta \in [\hat{\delta}, 1]$ , i.e. se  $\delta$  é grande o suficiente, eles mantêm o preço de monopólio para sempre. O quanto isto é pró ou contra-cíclico depende da forma da sequência de  $\{D(p, t)\}$ .
- Se  $\delta \in [0, (n - 1)/n]$ , i.e. se  $\delta$  é baixo o suficiente, eles mantêm o preço é  $mc$  para sempre, i.e. não existe conluio. Isto é mais provável quanto maior o valor de  $n$ . Se  $\exists \tilde{\delta} \in [(n - 1)/n, \hat{\delta}] \exists t^* \in [\hat{t}, \bar{t}]$  s.a.  $p(t^*) < p^m(t^*)$  e  $p(t) = p^m(t)$  para qualquer  $t \in [1, \dots, \bar{t} - t^*]$  para qualquer  $\delta \in [\tilde{\delta}, \hat{\delta}]$ .



No último caso  $\bar{t}$  é a tamanho do ciclo e  $\hat{t}$  é o pico. Então esta condição diz que existe um intervalo de valores de  $\delta$  onde iremos não apenas manter o monopólio após um ponto no ciclo e este ponto é após o pico (ponto mais elevado do ciclo econômico). Ou seja, o ponto a partir do qual não se pode sustentar o conluio é sempre quando a demanda está caindo.

Se o valor de  $\delta$  for reduzido existem mais pontos onde não se pode manter conluio, mas para demanda corrente igual, se perde habilidade de manter conluio quando a demanda começa a cair em relação ao ponto em que a demanda começa a aumentar (ver Figura).

Existem duas forças aqui: (i) a demanda mais elevada aumenta o retorno de desvio do conluio e (ii) a demanda declinante torna a punição do desvio menor. Então quando a demanda é alta e declinante é quando é mais difícil manter preços de monopólio.

Obs. Estes resultados são relativos ao preço de monopólio. Por sua vez o preço de monopólio depende de como as curvas de demanda se deslocam ao longo do ciclo econômico (preços são contra-cíclicos? Preços mais elásticos quando a demanda cresce?).

Quando os preços caem abaixo do valor de monopólio não se significa que os lucros devam cair. Neste caso não existe guerra de preços ou fase de punição. Haltiwanger e Harrington mostram que sob certas condições os lucros são pró-cíclicos. O que termina ocorrendo é que os lucros da indústria lideram fracamente o ciclo: *os lucros começam a cair antes da demanda decrescer*.

### 13.3.2 Preço Dinâmico no Varejo de Gasolina

Não existe muito trabalho empírico detalhado sobre conluio no sentido de se estimar as primitivas relevantes de um modelo e trabalhar os diferentes impactos do equilíbrio colusivo. Pakes (2019a, p.19) que isto deve ocorrer porque os modelos são muito estilizados para serem utilizados com os dados ou talvez porque o trabalho empírico esteja um pouco atrasado nesta área. As hipóteses assumidas nestes modelos não se adequam bem aos dados ou alternativamente se busca por implicações de modelos de forma reduzida que sejam consistentes com os dados.

O trabalho de Borenstein e Shepard (1996) sobre o varejo de gasolina é um exemplo de modelo de forma reduzida. Este é um mercado de produto diferenciado (em geral pela localização) com conhecidas mudanças sazonais na demanda e preços dos insumos (o insumo primária é a gasolina no atacado). Existem muitas firmas relacionadas em cada mercado, por isso os autores duvidam que a maximização conjunta de lucros seja sustentada.<sup>2</sup>

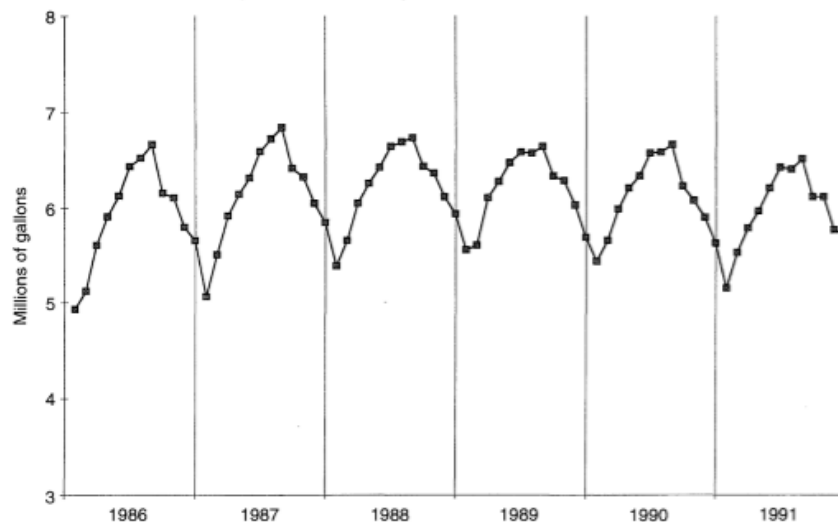
A estrutura dos dados é por cidade, 60 cidades por cinco anos, significando que não é modelada a competição dentro da cidade. Na Figura 3 eles apresentam o comportamento sazonal (mensal) das quantidades. Na Figura 1 é apresentada a evidência de sazonalidade para os preços relevantes (petróleo, preço no terminal e nos postos). O preço médio

<sup>2</sup>Ao menos sem pagamentos entre as empresas, o que seria ilegal.

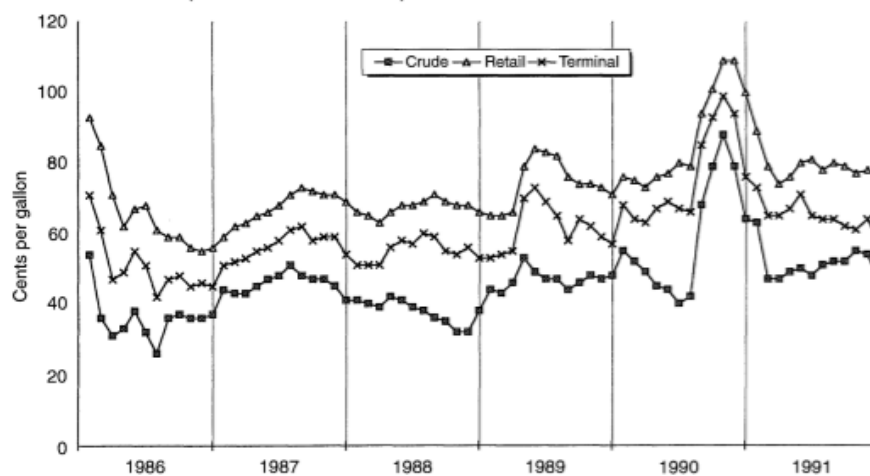
mensal da gasolina no terminal é o mais próximo que eles tem do preço do atacado, então as margens são aproximadas como a diferença entre o preço médio no terminal e o preço médio no varejo.

Esta aproximação do artigo é bem grossa. Na realidade existem diferentes tipos de contrato entre postos e as ditribuidoras. O modelo apropriado de para precificação depende da natureza dos contratos verticais entre os vários atores bem como da relação competitiva entre eles. Para análise de relação vertical veja Villas-Boas (2007) e Asker (2008). As séries de preços são muito mais erráticas do que as séries de volume. Portanto, é mais difícil observar comportamento sazonal nas margens. Este é um mercado com OPEP com várias considerações políticas e colusivas a serem feitas.

DAILY VOLUME 1986–1991 (STATE AVERAGES)



PRICES 1986–1991 (AVERAGE CITY PRICES)



A equação básica do artigo é:

$$M_{it} = \alpha_1 Q_{it} + \alpha_2 E(Q)_{it+1} + \alpha_3 TER_{it} + \alpha_4 E(TER)_{it+1} + \alpha_5 \Delta TER_{it} + \epsilon_{it}$$

além de time effects. Aqui  $Q$  é o volume médio no estado dividido pelo volume médio das vendas do varejo no período amostral.  $TER$  é o preço médio de terminal da gasolina.

A hipótese de que na falta de incentivos para colusão o preço do varejo deve ser uma defasagem distribuída dos preços passados do varejo e de terminal em torno de um nível de equilíbrio determinado pelo volume e pelos efeitos de cidade (dummies). Aqui eles fazem os preços de terminal e de varejo diferentes pois eles não possuem um modelo para atacado e varejo e existe a preocupação de que os preços de terminal possam ser estruturalmente diferentes dos preços de atacado.

A teoria de conluio de Haltiwanger e Harrington implicaria em  $\alpha_2 > 0$  enquanto que  $\alpha_3 < 0$  (uma vez que se a demanda está aumentando as punições são mais prováveis de serem efetivas e pode ser suportado preço mais elevado enquanto que uma vez que os preços de terminal aumentam as punições são menos efetivas e não se pode suportar preços mais elevados).

Eles preveem mudanças nos volumes com uma equação separada para cada cidade usando a seguinte equação:

$$Q_t = f(Q_{t-1}) + f(tempo) + \text{dummies-mês}$$

O *fit* é bem alto, entre .8 e .95. A boa capacidade de previsão é resultado do forte comportamento sazonal. Os autores preveem preços de terminal de forma similar: uma regressão por cidade como função do mês, defasagem do preço, e preços passados do barril de petróleo. O *fit* aqui é pior: entre .3 e .6. O preço de insumo varia de forma muito menos previsível do que as medidas de volume.

Borenstein e Shepard também se preocupam com com a medida de volume, uma vez que ele é função do preço e a variável do lado esquerdo é apenas preço menos o preço no terminal. Ou seja, a variável dependente contém preço e qualquer determinante não-observável que ficará contido em  $\epsilon$  ou nos efeitos das quantidades. O modelo para quantidade é

$$\ln(Q) = Z\beta + \eta \ln(p) + \nu$$

então para diferentes estimativas de  $\eta$  eles usam  $[\ln(Q) - \hat{\eta} \ln(p)]$  como um instrumento para  $Q$ .<sup>3</sup>

Evidência na Tabela 2 e 3 de Borenstein e Shepard (1996). Ambas as variáveis com expectativa possuem o sinal correto e são significantes: margem muda com o volume futuro ( $\alpha_2 > 0$ ), 3.22 para OLS e 3.92 para 2SLS, e a margem cai com o valor esperado positivo do preço de terminal (insumo) ( $\alpha_3 < 0$ ), -0.061 para OLS e -0.63 para 2SLS.

<sup>3</sup>Observer que eles estão assumindo que  $\nu$  é ortogonal a  $\epsilon$ . Isto é, os distúrbios nas equações de preço e quantidade não são correlacionados. Uma vez que se faz uma hipótese dessa,  $\eta$  pode ser identificado sem instrumentos adicionais.

Também é interessante notar a estimativa da estrutura autoregressiva. As estimativas mostram que a estrutura é assimétrica.

A margem média calculada simples como preço do varejo menos do terminal é em torno de 10.6 cents de dólar, relativo a média do preço de terminal de 73 cents por galão. O efeito de um desvio-padrão no volume esperado sobre a margem avaliada na média é de 0.26 cents, aproximadamente, e isso é similar ao número para o impacto da mudança de preços de terminal. Estes números não são muito grandes, mas ainda assim significantes. Este baixo valor pode significar pouco espaço para conluio, possivelmente pela existência de muitas empresas por cidade.

Ainda assim Borenstein e Shepard argumentam a favor de evidência de conluio tácito. As margens respondem como previsto por Haltiwanger e Harrington. Tudo o mais constante, eles encontram margens mais baixas quando a demanda é esperada declinar no período seguinte do que aumentar. As margens são mais elevadas quando o preço do terminal é esperado declinar no mês seguinte do que quando é prevista aumentar.

## 13.4 Informação Assimétrica em Jogos Repetidos

O interesse aqui é em jogos onde o preço ou quantidade de um concorrente não é diretamente observável e flutuações na demanda dificultam inferir o preço das quantidades observadas (ou o contrário, inferir quantidades de preços). A situação é que cada firma conhece suas escolhas mas não a do rival, este é o problema de informação assimétrica. Neste caso não é possível saber com certeza quando um competidor desviou de um acordo colusivo.

*Punição.* Quando as escolhas de preço são perfeitamente observáveis faz sentido a estratégia de punição extrema porque elas não serão observadas na trajetória do acordo colusivo (nesse caso eles seriam sem custos para as firmas). Quando existe incerteza, erros podem ser inevitáveis. Por exemplo, se poderia inferir que o preço jogado pelo rival foi baixo pela observação da demanda, enquanto que na realidade o preço era elevado. No caso deste engano, o cartel entraria em modo de punição. Isto implica que a punição máxima, que contém custo máximo, não é desejada. Punição máxima pode não ser desejada se isso destrói possibilidade de conluio.

Considere uma versão do modelo de jogos repetidos de Green e Porter (1984). Aqui é apresentado uma versão de competição em preços e não em quantidades (segundo Pakes, 2019a, pp. 23-5). Considere um mercado de bens homogêneos com custo marginal constante. Podemos permitir um estado com demanda baixa com probabilidade  $\alpha$  e um estado com demanda elevada, sendo as realizações i.i.d. ao longo do tempo.

Quando a demanda é baixa nenhuma das firmas recebe lucros. Quando a demanda é alta e existe um corte de preços apenas uma firma recebe lucro. Quando a demanda é alta e ambas as firmas coludem com lucro de monopólio, elas dividem os lucros. Quando não se recebe lucros não se sabe se este resultado foi por:

- demanda baixa, ou porque
- o rival jogou preços baixos enquanto a demanda era elevada.

Neste jogo os mercados se repetem indefinidamente. Considere um dado conjunto de estratégias e pergunte o que precisa ser satisfeito para estas estratégias serem de equilíbrio de Nash. As estratégias são as seguintes:

- Ambas as firmas cobram  $p^m$ , o preço de monopólio ou um preço alto, até uma das firmas receber lucro zero.
- A ocorrência de lucro zero desencadeia um período de punição. Nesta fase as firmas cobram  $mc$  por  $T$  períodos.
- Em  $T + 1$  as firmas reverterem para a fase de conluio.

Estas estratégias são um equilíbrio de Nash?

Na fase de punição elas claramente são. Dado que meu rival irá jogar  $mc$  por  $T$  anos, não faz sentido fazer algo diferente.

Período de conluio. Faça  $(V^+, V^-)$  serem valores nas fases colusiva e de punição, respectivamente, quando esta estratégia é jogada. Assumindo preços que são determinados conhecendo as realizações dos erros de demanda:

$$V^+ = (1 - \alpha)(\pi^m/2 + \delta V^+) + \alpha \delta V^- \quad (13.16)$$

e desde que

$$V^- = \delta^T V^+$$

temos que

$$V^+ = \frac{\pi^m(1 - \alpha)}{2[1 - (1 - \alpha)\delta - \alpha\delta^{T+1}]} \quad (13.17)$$

Para essa estratégia ser de equilíbrio é preciso satisfazer a restrição de incentivo que garante que a firma irá querer jogar o conluio, i.e. o requerimento é:

$$(1 - \alpha)\pi^m/2 + \delta[(1 - \alpha)V^+ + \alpha V^-] > (1 - \alpha)\pi^m + \delta V^- \quad (13.18)$$

que pode ser re-escrita como:

$$\delta(V^+ - V^-) > \pi^m/2 \Rightarrow V^+ > \frac{\pi^m}{\delta(1 - \delta^T)2} \quad (13.19)$$

Isto é:

- para impedir comportamento de corte de preços é preciso que  $V^+$  seja suficientemente grande do que  $V^-$ . Para isto é preciso ter  $T$  grande.
- Por outro lado, se a restrição de incentivos é satisfeita, quanto mais alto for  $V^-$ , melhor estão as empresas.  $V^-$  será maior quanto menor for  $T$ .

Então para encontrar a “punição ótima” na classe de esquemas de punição por  $T$  períodos, escolha um  $T$  mínimo que satisfaça esta última equação.<sup>4</sup>

Com um pouco de álgebra pode se mostrar que para alguma firma sustentar o conluio em  $T$ ,  $(1 - \delta^T) > (1 - \delta)/\delta(1 - \alpha)$ , então tomando  $T \rightarrow \infty$  é claro que apenas se sustenta conluio se

- $\delta$  é suficientemente elevado (as pessoas são suficientemente pacientes), e
- $\alpha$  é suficientemente baixo (existe uma probabilidade grande o suficiente de bons estados, tal que seja forçado a desistir dos lucros na fase de punição).

Se isso for verdade podemos solucionar para o  $T$  mínimo que satisfaz as condições e este será a punição ótima.

Comentários:

- O modelo original Green-Porter é um pouco mais complexo. Eles assumem um jogo em quantidades com reversão para Nash em  $T$  períodos de punição. Firms conhecem sua própria escolha de quantidade  $q$ , mas não a do competidor. Além disso as empresas não conhecem o choque de demanda. Dadas as escolhas  $q$  e o choque de demanda, as firmas observam o preço  $p$  do mercado. Se  $p < \bar{p}$  então a fase de punição começa. Caso contrário as firmas continuam em conluio. Este é o “gatilho de preço.”
- Observe que em ambos os casos limitamos o conjunto de punições possíveis a:
  - serem simétricos e
  - ser reversão para Nash por  $T$  períodos.

Se ambas restrições são ignoradas a punição muda.

- Abreu, Pearce and Stachetti (1985) mantém simetria mas permitem uma classe mais geral de estratégias simétricas, assumindo apenas uma propriedade razão de verossimilhança monótona para as realizações  $p_t$ . Isto é,  $p_t$  elevado são mais prováveis de serem observados com quantidade total comprada,  $Q_t$ , baixa. Eles mostram que podemos restringir atenção para uma fase colusiva e de punição, com quantidades únicas em cada, e com a fase de punição começando com um “gatinho de preço” de  $p < \bar{p}$ . Entretanto, a fase de punição não possui duração fixa; esta fase se parece com período de conluio no sentido de que um certo nível de patamar pode tirar a firma dela. I.e. se o preço de mercado permanece abaixo de algum outro  $p$ , digamos  $\underline{p}$ , eles continuam punindo, mas se o preço estiver abaixo de  $\underline{p}$  eles retornam ao conluio. A razão para precisar de algo abaixo de  $\underline{p}$  para manter o conluio é que uma punição dura requer quantidades muito elevadas, maior do que a reversão para Nash e portanto mais baixo do que é estatisticamente privadamente

<sup>4</sup>Para calcular isto, mude a equação para igualdade, substitua por  $V^+$  e tome o menor inteiro que seja maior do que  $T$  que garante a igualdade para os dois lados da equação.

ótimo. Portanto, apenas se realiza conluio de novo se o preço é bem baixo, o que sinaliza que as firmas jogaram a punição correta.

- Se permitirmos comportamento assimétrico então o resultado pode ser mais informativo sobre se alguém e quem desvia (e.g. se um ator faz alguma coisa que outro ator não faz, então pode ter implicações observadas que podem apenas ocorrer como resultado das ações de uma firma). Isto permite um esquema mais duro de punição e maior valores em geral – se uma firma pode descobrir como ser diferente, pode ser no próprio interesse.
- Observe que em todos estes modelos ninguém de fato desvia. Observe que se atinge a fase de punição pois é necessário manter o benefício (lucro) do conluio. Então as punições existem quando a trajetória de conluio deseja ser mantida. Porque as fases de punição são geradas observamos coisas que se parecem com *guerras de preço*.
- Finalmente tenha em mente que nenhum destes modelos analisados permitem entrada ou investimentos nas capacidades das empresas.

### 13.4.1 Conluio e Guerra de Preços

**The Joint Executive Committee** (Porter, 1983).<sup>5</sup> JEC foi um cartel de transporte ferroviário (principalmente de grãos) que operou a partir de Chicago para os portos da costa leste (foco em Baltimore, Boston e NY) entre 1880 e 1886. Este cartel funcionou antes do Shreman Act (1890) e da Interstate Commerce Commission (1887) e podia se declarar *publicamente* como cartel.

As alocações de mercado tomam a forma de *market shares* e as firmas determinam os preços consequentemente. O transporte fluvial nos Grandes Lagos era um concorrente, mas não afetou o acordo de conluio. Por serem previsíveis, as flutuações na demanda por transporte nos Grandes Lagos, devido ao fechamento no inverno, não afetaram a conduta. Mais do que isso, os preços praticados se adaptavam a concorrência náutica.

O cartel se baseava nos dados de quantidade por semana. A demanda foi altamente variável tal que os *market shares* realizados dependiam de todos os preços e da realização dos choques de demanda. Neste caso, o *enforcement* do cartel foi uma variante do gatilho de preços.

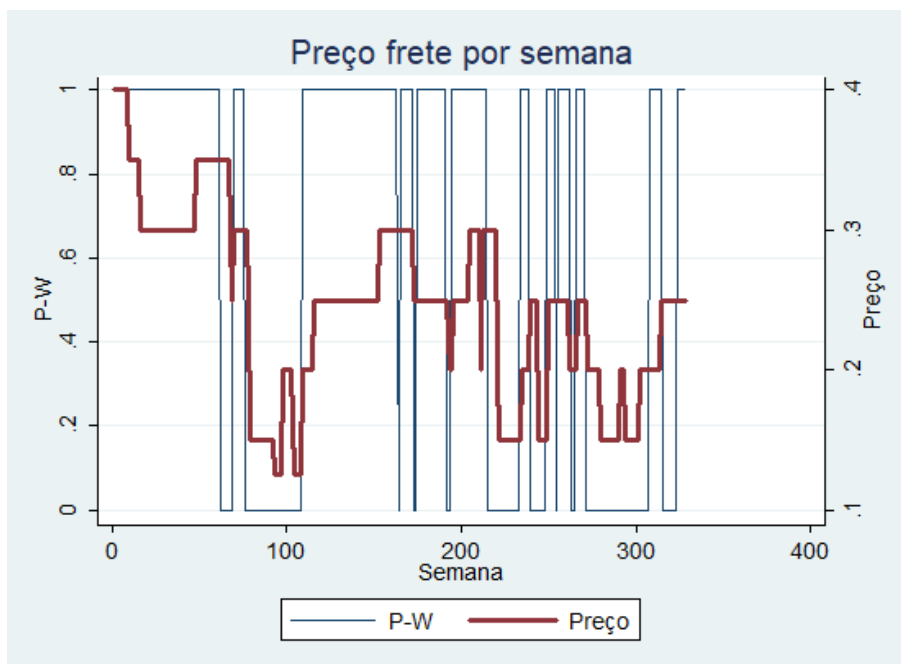
Uma função principal do JEC foi coletar informação e disseminar aos membros do cartel semanalmente. A quantidade total de grãos transportados (TQG) por semana variou bastante durante o período do cartel.<sup>6</sup>

A variável de preços (GR) foi informado pelas empresas ao JEC. Este preço pode ser suspeito pois seria pouco provável as firmas informarem quando elas cortam valores.

<sup>5</sup>Ver Box 9.2 de Cabral (2017).

<sup>6</sup>Entre parêntesis segue a notação do artigo de Porter.

O cartel publicava os dados na “Railway Review.” Guerra de preços era registrado pela revista. Porter utiliza uma variável *dummy* (PO) para indicar quando ocorria guerra de preços de acordo com a revista e realizou uma estimativa se de fato desta fase (PN). Observe o comportamento de preços durante o período de guerra de preços (P-W) e o de comportamento cooperativo (conluio).



Existiram várias mudanças na estrutura da indústria no período estudado. Incluindo a entrada de duas firmas, a saída de outra do cartel, o fechamento e abertura de serviços concorrentes (Grandes Lagos) e vários efeitos sazonais. Estas características ou acontecimentos faz com as hipóteses por trás do “jogo repetido” sejam suspeitas e a forma com que o artigo lida com elas é de que a mudança na estrutura provoca mudanças exógenas nos preços do cartel (observe que é apenas na fase de punição). A equação que eles estimam é:

$$\ln(p_t) = \beta_0 + \beta_1 \ln(Q_t) + \beta_2 S_t + \beta_3 I_t + U_t \tag{13.20}$$

aqui  $\beta_0 + \beta_1 \ln(Q_t) + \beta_2 S_t$  representa o preço na fase de punição, com  $S_t$  sendo um vetor de características do mercado (entrada, saída, etc).  $I_t$  é uma *dummy* igual a 1 quando se está no período de conluio e 0 na fase de punição.

Uma questão interessante não abordada é o quanto a existência e sucesso do cartel induz à mudança na estrutura da indústria (quando a indústria tem sucesso deveríamos esperar maior entrada e saída quando não se tem sucesso). Um ponto interessante observado por Porter é que duas entrantes no período foram absorvidas pelo cartel sem muita luta (essa é uma reação que pode incentivar entrada pois seria possível para todos entrantes).

As estimativas estão na Tabela 3 de Porter (1983). As *dummies* para conluio indicam que o preço era 40% acima do que na fase de punição. A estimação apresentada aqui é



por 2SLS e utilizando instrumentos para GR e TQG (Lagos é a dummy para abertura dos Grandes Lagos). Os preços cooperativos são claramente mais elevados do que do período de punição. Entretanto Porter relata que os preços parecem ser menor do que seria sugerido pela maximização conjunta de lucros. Este ponto leva a questão de se os custos de manutenção do conluio são muito elevados ou pelo menos muito elevados quando o ambiente é muito variado no tempo.

- A fase de punição corresponde a guerra de preços ( $P-W = 0$ ) na figura, mas a guerra de preços parece variar bastante em duração e em magnitude;
- Reversões para a guerra de preços acontecem mais regularmente nos últimos períodos após a entrada e portanto quando se tem mais membros no cartel (qual o  $\Delta$ ?).
- O modelo diz que a guerra de preços deve ocorrer quando há uma realização muito baixa da demanda que não era prevista. Porter não encontra evidência nos erros de demanda, mas a questão é que muitas variáveis não estão presentes no sistema de demanda que podem de fato dominar o comportamento do termo de erro. A questão é que estas variáveis não observadas podem ser conhecidas pelas empresas e não pelo economista.

Tabela 13.1: Estimativas de Demanda e Oferta (2SLS)

Variáveis	Demanda	Oferta
Preço (GR)	-0.742 (.121)	
Lagos (Dummy)	-0.437 (.120)	
Q (TQG)		0.251 (.171)
PO		0.382 (.059)
C	9.169 (.184)	-3.944 (1.76)

## 13.5 Detecção de Conluio

### 13.5.1 Licitações

Nesta seção vamos tratar de comportamento de cartéis em licitações, com foco na detecção.<sup>7</sup>

Um exemplo brasileiro que ilustra o comportamento colusivo em cartéis de licitações é o leilão de trechos da ferrovia norte-sul (EF-151) em 8 de maio de 1987.<sup>8</sup> A licitação para

<sup>7</sup>Importante observar que a solução eficiente do cartel é a maximização conjunta de lucros sujeita às restrições de participação e revelação de informação.

<sup>8</sup>A história da licitação é descrita em Gaspar (2020, p. 79).

a construção da ferrovia ligaria as cidades de Anápolis (GO) e Açailândia (MA), sendo a cidade goiana central para o agronegócio e Açailândia o entrocamento da ferrovia nortesul com a Estrada de Ferro Carajás (EF-315), provendo assim a conexão da produção agrícola do Centro-Oeste ao porto de Itaqui (MA). O projeto começou a ser discutido em 1985, no governo Sarney, e o primeiro trecho (Porto Franco (MA) – Açailândia (MA)) de 215 km foi entregue em 1996, no governo Fernando Henrique Cardoso. A primeira licitação, em 8 de maio de 1987, com o total de 2,5 bilhões de dólares em contratos, foi um exemplo de funcionamento de cartel em licitação. O jornalista Janio de Freitas publicou no dia 8 de maio um anúncio, na seção de classificados, antecipando o resultado do conluio. No dia 13 de maio ele denunciou a conspiração em sua coluna,<sup>9</sup> que era a seguinte: cada grande empreiteira havia sido contemplada com um lote.<sup>10</sup> A segunda proposta sempre foi com valor 10% inferior ao da vencedora. Tal prática evidencia a comportamento das empresas em conluio na licitação. Após a denúncia, a concorrência foi suspensa.

Detecção de conluio não possui estratégia geral. Como veremos, a circunstância da indústria e dos contratos são determinantes para encontrar evidência de atividade colusiva.

Uma estratégia comum é a detecção pelas autoridades antitruste ou por vítimas da atividade colusiva. É comum a detecção do cartel utilizar incentivos de leniência. Essa estratégia possui seus méritos, mas, como sublinhado por Porter (2005), a “reclamação” por parte de uma firma concorrente pode ser suspeito. Rivais tipicamente possuem ganhos com preços mais elevados e eles enfrentam dificuldades com mais concorrência.

Um conluio em licitação deseja evitar detecção pelas autoridades antitruste se as discussões são ilegais e também por um competidor que pode alterar o mecanismo em resposta. Para essa situação, Porter (2005) apresenta o seguinte exemplo: um vendedor pode elevar o lance mínimo em um leilão ou manter ele em segredo para os sócios do cartel para manter a disciplina do acordo. Se o cartel de licitação não inclui todas as firmas, os participantes podem decidir manter o acordo em segredo de outros potenciais ofertantes. Os lances são tipicamente complementos estratégicos, tal que os rivais que são certos de participarem irão realizar lances menos agressivos em resposta à um conluio, mas outros participantes em potencial podem entrar e realizar lances agressivos.<sup>11</sup>

A melhor prova – e talvez a mais usada – contra um conluio é a prova direta produzida por um dissidente. Para além deste tipo de prova, cartéis são difíceis de detectar.

<sup>9</sup>Janio de Freitas, Concorrência da ferrovia Norte-Sul foi uma farsa. Folha de São Paulo, 13 de maio de 1987. Para mais sobre conluio em licitações de estradas de ferro veja “<http://en.cade.gov.br/cade/noticias/cade-mpf-go-e-policia-federal-realizam-operacao-para-investigar-suposto-cartel-em-licitacoes-de-ferrovias/historico-da-conduta-versao-publica.pdf>.”

<sup>10</sup>As empreiteiras foram: Odebrecht, Queiroz Galvão, Mendes Jr, C.R. Almeida, Serveng, Egit, Cowan, Ceesa, CBPO, Camargo Correa, Andrade Gutierrez, Constran, Sultepa, Construtora Brasil, Aleino Vieira, Tratex, Paranapanema, Ferreira Guedes.

<sup>11</sup>Um argumento análogo é o caso de um mercado de produtos diferenciados, em que firmas determinam seus preços, onde a introdução de produtos novos introduzido por rivais possa dissipar os lucros de um cartel.

Operadores de um conluio podem criar a aparência de concorrência para tentar evitar detecção, por exemplo.

Em licitações, os cartéis podem submeter propostas que simulem competição entre empresas. Neste caso, apenas a proposta mais baixa seria “séria” – no sentido de ser uma proposta competitiva. Entretanto, uma oferta falsa, de forma diferente de uma “séria”, pode não estar relacionada com as propostas sérias em eventos que essa mesma firma ganha – pode existir não-aderência entre o tipo as propostas “sérias” e falsas.

Um caso que descreve este comportamento foi apresentado por Porter e Zona (1993, e Porter, 2005), relacionado a lances falsos, que versava sobre um cartel formado para participar de licitação de pavimentação de ruas em Long Island (New York). Neste caso, um subconjunto de firmas participaram de encontros antes da licitação para conseguir acordos de lances baixos (redução da concorrência). Como autores não tiveram acesso a informação contida nos contratos, eles analisaram a ordem dos lances submetidos. Eles mostram que a ordem das propostas submetidas pelas firmas que não faziam parte do conluio, eram relacionadas com componentes dos custos, tais como capacidade produtiva, uma medida de utilização da capacidade, e o *backlog* de contratos que foram vencedores em outras licitações. A proposta mais baixa apresenta por participante do cartel era mais provável ter sido realizada pela empresa com o menor custo. Por outro lado, a ordem dos lances mais elevados submetidos pelos participantes não foi correlacionada com as mesmas medidas de custos. Porter (2005) argumenta que cartéis mais sofisticados poderiam passar neste teste – as empresas poderiam inflar todas as propostas submetidas acima do custo no mesmo fator (percentual). Os lances falsos também podem servir para manipular a expectativa do comprador. Em um setor ou mercado, o histórico de lances falsos elevados, servem para manter a margem mais confortável para os operadores de um cartel.

De forma geral, a detecção é mais fácil se a estrutura de preço for relativamente simples.

Existe uma percepção que cartéis em licitação podem operar com maior facilidade se for leilão inglês ou leilão de segundo-preço com lance selado (SPSB da sigla em inglês). Em um SPSB, o maior lance ganha, mas o pagamento é realizado pelo valor do segundo maior lance (o lance inferior ao vencedor), ou o lance mínimo se ninguém mais submete uma proposta. Se o membro do cartel que foi designado para ofertar o lance vencedor tem a maior valoração e se esta mesma empresa faz um lance de valor verdadeiro, os outros membros não ganham desviando. Observe que ofertar valor verdadeiro é uma estratégia dominante em um leilão SPSB quando não existem elementos de valor comum nos payoffs, tal que a estratégia ótima da empresa escolhida para ganhar a licitação não depende da competição. O sucesso do cartel na licitação depende de como muitos participantes em potencial escolhem não enviar propostas, reduzindo assim o preço esperado a ser pago pelo licitante selecionado quando este ganhar (veja Porter, 2005, e Robinson, 1985).<sup>12</sup>

<sup>12</sup>De forma similar, em um leilão inglês ascendente, a proposta séria apenas precisa apresentar uma oferta mais competitiva do que as demais. Portanto, existe um problema de curto prazo no cartel apenas

Um método econométrico para detectar cartel entre um subconjunto de participantes em leilões inglês foi proposto por Baldwin, Marshall, e Richard (1997). Este método pode ser aplicado diretamente a leilões SPSB, assumindo que o cartel é eficiente em selecionar o vencedor que possui a maior valoração para a licitação da vez. Se os participantes possuem valoração privada, e o leilão é concorrencial, então o lance vencedor é a estatística de segunda ordem da distribuição dos valores, dado que o licitante de maior valoração ganha certame em relação ao segundo ofertante. Se um subconjunto de licitantes faz colusão, então o lance vencedor é afetado apenas se os dois licitantes que dão o maior valor aos lances são membros do cartel.

- Se não for o caso de conluio, o lance vencedor continua sendo uma estatística de segunda ordem na distribuição dos valores.
- Se for conluio, então o lance vencedor será de algum participante que não faz parte da conspiração. Ele será uma estatística de terceira ordem se o valor for o terceiro maior entre todos os licitantes.

Portanto, se existir conluio em um leilão SPSB, a proposta vencedora será uma mistura de estatística de segunda ordem com de ordens mais baixas.<sup>13</sup>

Por outro lado, colusão é mais difícil em leilão de primeiro preço com lances selados (FPSB, da sigla em inglês *first price sealed bid*). Neste caso, o participante sério pode oferecer lances mais baixos do que o valorado por outros membros do cartel. Neste ambiente, os demais membros podem ganhar a licitação a um preço lucrativo, existindo aqui o problema usual de evasão unilateral do conluio.

Neste tipo de leilão, para além de criar a aparência de leilão concorrencial, os lances complementares podem ter um papel se a ameaça de punição futura é insuficiente para prevenir a deserção. O participante do cartel designado a ganhar a licitação deve ter lance maior do que o ótimo para contrapor a concorrência fora do cartel – para remover o incentivo para outros membros do cartel desertarem e apresentar oferta mais elevada. Nesse caso, um lance complementar realizado por um membro do cartel deve ser inferior ao lance designado a ganhar a licitação, para dissuadir a deserção.

De forma parecida com o SPSB, em alguns desenhos de leilões multi-unidade, o recurso de punição para evitar deserção faz parte da própria licitação. Um exemplo fornecido por Porter (2005) é o leilão de espectro para serviços de comunicação pessoal (PCS) realizado pelo FCC (Federal Communications Commission), que funciona no formato de leilão ascendente de múltiplas unidades. Neste design, a punição aplicada pode ser ampla – a deserção no leilão de um objeto, pode provocar diversas retaliações em outras unidades. Quando isto é possível, a estrutura simultânea ascendente pode facilitar conluio tácito (Gertner, 1995). Um equilíbrio deste tipo de jogo, funciona com a divisão de licenças entre firmas, com cada firma submetendo um lance baixo para um conjunto de licenças de

---

se o ofertante sério não possua a maior valoração entre os membros do cartel.

<sup>13</sup>Neste modelo, a hipótese sobre forma funcional da distribuição da valoração (valuation) possui papel importante na distinção entre concorrência e conluio.

espectro. O equilíbrio é sustentado pela ameaça dos lances competitivos serem realizados em regiões que previamente seriam exclusivas para a empresa que desertou. Neste caso, a empresa que desvia pode ter que lidar com ondas de lances competitivos em sua área pré-estabelecida no acordo colusivo. O design do FCC, permitia novos lances por qualquer licença a qualquer momento em que leilão está ativo, e portanto, a retaliação em outro território era crível.

Outro caso é o leilão de preço uniforme para múltiplas unidades, onde o preço de mercado será aquele que faz a oferta ser igual à demanda. Exemplo clássico deste tipo de esquema de precificação foi utilizado pelo Reino Unido para contratar volume (carga, no jargão do setor) de energia.<sup>14</sup> Neste caso, o preço de mercado seria dado pelo maior lance perdedor. Os operadores do cartel que participam do leilão podem tornar custoso para os rivais roubarem market share por meio de lances mais baixos para a oferta inframarginal e um preço elevado para a oferta marginal. Porter (2005) diz que esta estratégia é as vezes denominada de lances de *hockey stick*. Uma unidade geradora é inframarginal se ela será provavelmente chamada para ofertar energia, mas provavelmente não é decisiva na determinação do preço de mercado. Se todos os participantes seguirem esta estratégia e as unidades de geração inframarginal contabilizam por uma grande parcela da oferta de energia, o preço de mercado será elevado, mas os ganhos de deserção serão baixos.<sup>1516</sup>

Porter (2005) chama atenção que um problema que é consequência do comportamento colusivo é a probabilidade de entrada de novos competidores quando as firmas conseguem obter lucros acima do competitivo. O problema da entrada é muito importante pois afeta tanto o desenho bem como a estabilidade dos acordos colusivos.

Restrições legais à entrada mantidas ou propostas por governos são efetivas como barreiras à entrada em mercados cartelizados. Sanções ilegais também podem ser utilizadas para evitar entrada de competidores – sanções ilegais aqui são métodos criminais.

Outra barreira possível é os incumbentes praticarem preços predatórios. Neste caso, as empresas que compõem o cartel praticam preços baixos para expulsar as empresas que entrem no mercado, aumentando estes depois que a entrada foi evitada.

Exemplo de leilões *drenagem*,<sup>17</sup> onde a informação assimétrica funciona como barreira à

<sup>14</sup>Sobre o mercado do Reino Unido veja por exemplo Wolfram (1999) e Doraszelski, Lewis e Pakes (2018). Essa configuração de preço uniforme não mais é utilizada.

<sup>15</sup>Neste caso, um leilão discriminatório, em que cada unidade fornecedora receba o valor de sua licitação, pode induzir a licitações mais competitivas.

<sup>16</sup>Em um cenário em que firmas não distinguem entre condições adversas na demanda e deserção, voltamos aos problemas de demanda flutuante da seção 13.4. Porter (2005) chama atenção para o seguinte exemplo: se as firmas não observam os preços efetivos dos rivais e apenas conhecem suas próprias vendas, então vendas mais baixas podem ocorrer porque uma firma foi vencida por um rival ou devido à demanda mais baixa. Então um cartel pode responder a vendas inesperadamente mais baixas, declarando guerra de preços aos rivais. Neste contexto, se a demanda for relativamente variável se pode esperar quebra do acordo de conluio, mesmo que não ocorra deserção. A guerra de preços em resposta à frustração de vendas seriam necessárias para manter os incentivos de cooperação intactos (Green e Porter, 1984).

<sup>17</sup>Tradução livre de *drainage auctions*.

entrada. Hendricks e Porter (1988, 2007) apresentam este caso para leilões de óleo e gás nos EUA, onde atribuem o leilão drenagem para áreas em que existia exploração prévia. Firms com experiência de exploração anterior ao local leiloado possuíam vantagem informacional em relação às empresas que possuem apenas as informações de sísmica (disponibilizada no processo de licitação das áreas). Neste leilões, a identidade das firms com direito exploratório de áreas adjacentes as definidas como objeto da outorga são conhecidas previamente. Além disso, o número destas empresas são limitados pelo quantitativo de outorgas anteriores. Neste caso, os vizinhos podem ter ganhos advindos de coordenação, e eles não precisam se preocupar com a entrada de não-vizinhos, que dissipariam todos os ganhos do cartel. Os autores encontram evidência de que os vizinhos recebem lucros altos, enquanto que os não-vizinhos apenas conseguem pagar os custos (com pouco lucros).

Apesar dos altos retornos para os participantes do cartel, ocorreu menos entrada (isto foi contabilizado por menos lances apresentados) do que no arrendamentos das áreas wildcat (exploração de área arriscadas, em que já ocorreu exploração ou não existe informação prévia), onde os participantes do leilão possuem aproximadamente a mesma informação. As taxas mais baixas de entrada nos leilões de drenagem são consistentes com as assimetrias de informação agindo como barreiras à entrada.

Se os vizinhos participam de leilões drenagem não-cooperativamente, então não deve ocorrer entrada de firms não-vizinhas, uma vez que esta última não tem informação sobre a performance de perfuração privada. No entanto, ainda assim ocorreu entrada de não-vizinhos. Além deste comportamento, as ofertas de não-vizinhos são independentes do número destas empresas, ao contrário de ser uma função decrescente, como previsto pela maldição do vencedor. Além disso, muitas vezes existem vários lances dos vizinhos em um único contrato de drenagem, mas seus retornos ex post são uma função crescente do número de lances apresentados. Por último, o lance mais elevado de vizinho é independente do número de firms deste tipo, e seu lance médio é uma função decrescente deste total. Este último fato é consistente com os vizinhos submetendo apenas um lance sério, e a probabilidade de submeter lances complementares sendo uma função crescente do número de arrendamento vizinhos, a fim de criar aparência de concorrência.

### 13.5.2 Quebra do Cartel

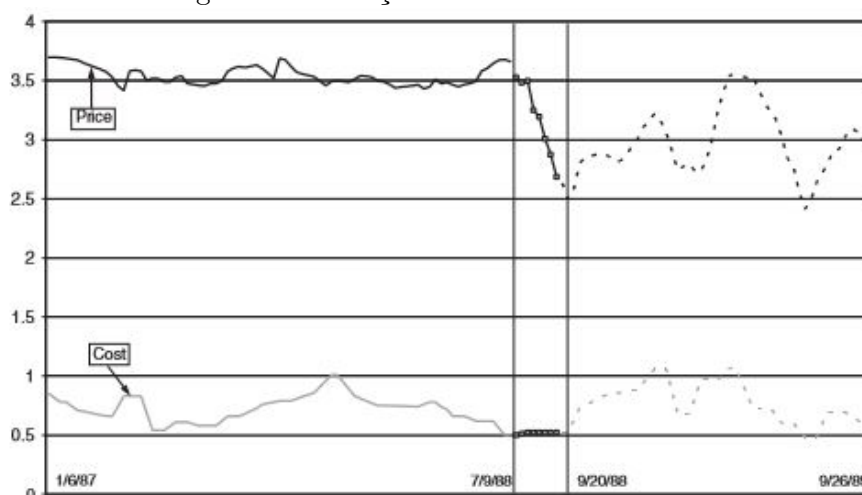
Uso de regularidades como filtros de comportamento de cartel é baseado no caso clássico de conluio na licitação de filé de perca (peixe) congelado para o Philadelphia Defense Personal Support Center entre 1987 e 1989. O caso de conluio na licitação de peixe durou até agosto de 1988. Na Figura é apresentado o preço e o custo do filé de perca congelado, com o período de quebra do cartel marcado entre as linhas verticais. O custo do filé de perca é o preço do produto sem processamento. Esta figura ilustra o colapso do conluio e se observa o colapso do preço em relação ao custo após o fim do ilícito. No período competitivo o preço começa a variar de acordo com o custo e com maior variância (quando existem muitos competidores). Este é o típico comportamento que se espera ao fim de atividades anti-competitivas.

Tabela 13.2: Média e Variância do Preço e Custo Filé de Peixe

	Concorrência	Conluio	Diferença (%)
Preço Médio	2.97	3.544	-16.20
Variância	0.283	0.078	262.82
Coef. Variação	0.095	0.022	331.82
Custo Médio	0.771	0.722	6.79
Variância	0.173	0.114	51.75
Coef. Variação	0.224	0.158	41.77

Fonte: Abrantes-Metz, Froeb, Geweke e Taylor (2006), p. 473.

Figura 13.1: Preço e Custo do Filé de Peixe



Fonte: Abrantes-Metz, Froeb, Geweke e Taylor (2006), p. 472.

Na Tabela 13.2 é apresentado o resumo da Figura contendo o preço médio, o custo médio pré e pós conluio bem como a variância e o coeficiente de variação (variância sobre o preço médio) para estes dois indicadores.

Como descrito por Abrantes-Metz, Froeb, Geweke e Taylor, o preço caiu 16% com a mudança de regime e a variância aumentou 262%. A média e o desvio padrão do custo também são maiores, mas não grandes o suficiente para explicar a amplitude de variação no preço.





## Capítulo 14

# Bibliografia

### 14.1 Notas sobre Referências

- Introdução – para uma boa introdução ao problema da regulação econômica veja o capítulo 1 de Cabral (2017) e o capítulo 10 de Viscusi, Vernon e Harrington (1998). No capítulo 10, seção 10.4 e 10.5, de Cabral (1994) está contida uma breve exposição sobre o tema.
- Econometria: Variáveis instrumentais – para a introdução à variáveis instrumentais veja capítulo 12 de Stock e Watson (2015) e os capítulos 15 e 16 de Wooldridge (2016). Também veja os capítulos 2 e 3 de Angrist e Pischke (2015).
- Internet: Melhores sites
  1. <http://luiscabral.org/iio/links/> – home page do livro de Luis Cabral (2017) que contém links importantes de organização industrial.
- Guia para Análise de Atos de Concentração Horizontal, Julho 2016: [www.cade.gov.br/acesso-a-informacao/publicacoes-institucionais/guias\\_do\\_Cade/guia-para-analise-de-atos-de-concentracao-horizontal.pdf](http://www.cade.gov.br/acesso-a-informacao/publicacoes-institucionais/guias_do_Cade/guia-para-analise-de-atos-de-concentracao-horizontal.pdf)



# Referências Bibliográficas

- Abrantes-Metz, Rosa, Luke Froeb, John Geweke e Christopher T. Taylor. “A Variance Screen for Collusion.” *International Journal of Industrial Organization*, 24 (3), 2006.
- Akerberg, Daniel A., Kevin Caves, e Garth Frazier. “Identification Properties of Recent Production Function Estimators.” *Econometrica*, 83 (6), pp. 2411-2451, 2015.
- Akgun, Ugur, Cristina Caffarra, Federico Etro, e Robert Stillman. “On the Welfare Impact of Mergers of Complements: Raising Rivals’ Cost Versus Elimination of Double Margin.” *Economics Letters*, 195, October 2020.
- Anderson, Theodore W. e Herman Rubin. “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations.” *Annals of Mathematical Statistics*, 21, pp. 570-582, 1950.
- Andrews Isaiah, Matthew Gentzkow, Jesse M. Shapiro. “Transparency in Structural Research.” *Journal of Business and Economic Statistics*, 38 (4), 2020.
- Andrews, Isaiah, James H. Stock, e L Sun. “Weak Instruments in IV Regression: Theory and Practice.” *Annual Review of Economics*, 11, 2019.
- Angrist, Joshua e Jörn-Steffen Pischke. *Mastering 'metrics: The Path from Cause to Effect*. Princeton, Princeton University Press, 2015.
- Backus, Matthew. “Why is Productivity Correlated to Competition?” *Econometrica*, 88 (6), 2020.
- Bain, Joe S. *Barriers to New Competetion*. Cambridge, Harvard University Press, 1956.
- Baldwin, Laura, Robert Marshall, e Jean-Francois Richard. “Bidder Collusion at Forest Service Timber Auctions.” *Journal of Political Economy*, 105, 1997.
- Bartelsman, Eric, John Haltiwanger, e Stefano Scarpetta. “Measuring and Analyzing Cross-country Differences in Firm Dynamics.” In *Producer Dynamics: New Evidence from Micro Data*. Cambridge, NBER, 2009.
- Berry, Steven T. “Estimating Discrete-Choice Models of Product Differentiation.” *RAND Journal of Economics*, 25 (2), 1994.

- Berry, Steven T., James A. Levinsohn, e Ariel Pakes. “Automobile Prices in Market Equilibrium.” *Econometrica*, 63 (4), 1995, pp. 841-890.
- Berry, Steven T., James A. Levinsohn, e Ariel Pakes. “Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy.” *American Economic Review*, 89 (3), 1999.
- Berry, Steven T., e Phil Haile. “Foundations of Demand Estimation.” In: Kate Ho, Ali Hortacsu e Alessandro Lizzeri (eds.). *Handbook of Industrial Organization*. Volume 4. Amsterdam, North Holland, 2021.
- Bloom, Nicholas, Renata Lemos, Daniela Scur, Raffaella Sadum, e John Van Reenen. “The New Empirical Economics of Management.” *Journal of the European Economic Association*, 12 (4), pp. 834-876, 2014.
- Bloom, Nicholas, e John Van Reenen. “Measuring and Explaining Management Practices Across Firms and Countries.” *Quarterly Journal of Economics*, 122, pp. 1351-1408, 2007.
- Borenstein, Severin e Andrea Shepard. “Dynamic Pricing in Retail Gasoline Markets.” *Rand Journal of Economics*, 27 (3), 1996.
- Bresnahan, Timothy. “Departures from Marginal-Cost Pricing in the American Automobile Industry.” *Journal of Econometrics*, 17 (2), 1981.
- Caballero, Ricardo. *Specificity and the Macroeconomics of Restructuring*. Yrjo Jahnsson Lectures. Cambridge, MIT Press, 2006.
- Caballero, Ricardo e M. Hammour. “On the Timing and Efficiency of Creative Destruction.” *Quarterly Journal of Economics*, 111, 1996.
- Cabral, Luis. *Economia Industrial*. Lisboa, McGraw-Hill, 1994.
- Cabral, Luis. *Introduction to Industrial Organization*. 2nd ed. Cambridge, MIT Press, 2017.
- Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, e Chad Syverson. “Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector.” *American Economic Review*, 106 (8), pp. 2110-2144, 2016.
- Che, Yeon-Koo e József Sákovics. “A Dynamic Theory of Holdup.” *Econometrica*, 72 (4), 2004.
- Chamberlin, Edward H. *The Theory of Monopolistic Competition*. Cambridge, Harvard University Press, 1933.
- Chernozhukov, Victor e Iván Fernández-Val. Bootstrapping. Lecture notes. MIT, Cambridge, 2017.
- Coase, Ronald H. “The Nature of the Firm.” *Economica*. 4, 1937.

- Coll, Steve. *The Deal of the Century: The Breakup of AT&T*. New York, Atheneum, 1986.
- Collard-Wexler, Allan e Jan De Loecker. “Reallocation and Technology: Evidence from the US Steel Industry.” *American Economic Review*, 105 (1), 2015.
- Conlon, Chris, e Jeff Gortmaker. “Best practices for differentiated products demand estimation with PyBLP.” *RAND Journal of Economics*, 51, 2020.
- Court, Andrew. “Hedonic Price Indexes with Automotive Examples.” in: *The Dynamics of Automobile Demand*, General Motors, New York, 1939.
- Crawford, Gregory S. e Ali Yurukoglu, “The Welfare Effects of Bundling in Multichannel Television Markets,” *American Economic Review*, 102 (2), 2012.
- De Loecker, Jan. “Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity.” *Econometrica*, 79 (5), pp. 1407-1451, 2011.
- De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. “The Rise of Market Power and the Macroeconomic Implications.” *Quarterly Journal of Economics*, 135 (2), pp. 561-644, 2020.
- De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, e Nina Pavcnik. “Prices, Markups, and Trade Reform.” *Econometrica*, 84 (2), 2016.
- De Loecker, Jan e Frederic Warzynski. “Markups and Firm-Level Export Status.” *American Economic Review*, 102 (6), pp. 2437-2471, 2012.
- Dearing, Adam. “Estimating Structural Demand and Supply Models using Tax Rates as Instruments.” *Journal of Public Economics*, 205, 2022.
- Deaton, Angus e John Muellbauer. “An Almost Ideal Demand System.” *American Economic Review*, 70 (3), 1980.
- Denning, Liam. “Time for Shale’s Loners to Settle Down.” *Bloomberg*, 20 de junho de 2017.
- Dixit, Avinash K. e Joseph E. Stiglitz. “Monopolistic Competition and Optimum Product Diversity.” *American Economic Review*, 67 (3), 1977.
- Doraszelski, Uli, G. Lewis, e Ariel Pakes. “Just Starting Out: Learning and Equilibrium in a New Market.” *American Economic Review*, 108(3), pp. 565-615, 2018.
- Eizenberg, Alon. “Upstream Innovation and Product Variety in the United States Home PC Market.” *Review of Economic Studies*, 81(3), 2014.
- Ellison, Glenn e Edward Glaeser. “Geographic Concentration in U.S. Manufacturing Industries: A Dashboard Approach.” *Journal of Political Economy*, 105 (5), pp. 889-927, 1997.

- Epple, Dennis e Bennet T. McCallum. "Simultaneous Equation Econometrics: The Missing Example." *Economic Inquiry*, 44 (2), pp. 374-384, 2006.
- Ericson, Richard e Ariel Pakes. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economics Studies*, 62 (1), 1995.
- Erikson, Tim e Ariel Pakes. "An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods." *American Economic Review*, 101 (5), 2011.
- Farrell, Joseph e Carl Shapiro. "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition." *B.E. Journal of Theoretical Economics: Policies and Perspectives*. 2010.
- Friedman, James. "A Noncooperative Equilibrium for Supergames." *Review of Economic Studies*, 28, 1971.
- Fuss, Melvyn, e Daniel McFadden (orgs.). *Production Economics: A Dual Approach to Theory and Applications*. Volume I: The Theory of Production. Amsterdam, North-Holland, 1978.
- Fuss, Melvyn, Daniel McFadden, e Yair Mundlak. "A Survey of Functional Forms in the Economic Analysis of Production." In: Fuss e McFadden (orgs.), 1978.
- Gorman, W. M. "On a Class of Preference Fields." *Metroeconomica*, 13 (2), 1961,
- Green, Edward J. e Robert H. Porter. "Noncooperative Collusion Under Imperfect Price Information." *Econometrica*, 52, 1984.
- Griliches, Zvi. "Hedonic Prices for Automobiles: An Econometric Analysis of Quality Change." In: *The Price Statistics of the Federal Government, General Series No. 73*. Columbia University Press for the National Bureau of Economic Research, New York, 1961.
- Gruley, Bryan, Kevin Crowley, Rachel Adams-Heard, e David Wethe. "Frackers Are in Crisis, Endangering America's Energy Renaissance." *Bloomberg Businessweek*, 21 de junho de 2020.
- Hall, Robert E. "The Relationship Between Price and Marginal Cost in US Industry." *Journal of Political Economy*, 96, pp. 921-947, 1988.
- Haltiwanger, John e James E. Harrington, Jr. "The Impact of Cyclical Demand Movements on Collusive Behavior." *Rand Journal of Economics*, 22 (1), 1991.
- Hansen, Bruce E. *Econometrics*. Princeton, Princeton University Press, 2022.
- Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50 (4), 1982.

- Harberger, Arnold C. "Monopoly and Resource Allocation." *American Economic Review*, 44, pp. 77-87, 1954.
- Hausman, Jerry, G. Leonard, e J. D. Zona. "Competitive Analysis with Differentiated Products." *Annales d'Economie et de Statistique*, 34, 1994.
- Hausman, Jerry, Serge Moresi e Mark Rainey. "Unilateral Effects of Mergers with General Linear Demand." *Economics Letters*, 111 (2), 2011, pp. 119-121.
- Hastings, Justine S., "Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California." *American Economic Review*, 94 (1), 2004.
- Hastings, Justine S. e Richard Gilbert., "Vertical Integration in Gasoline Supply: An Empirical Test of Raising Rivals' Costs." *Journal of Industrial Economics*, 53 (4), 2005.
- Hendricks, Ken e Robert H. Porter. "An Empirical Study of an Auction with Asymmetric Information." *American Economic Review*, 78 (5), 1988.
- Hendricks, Ken e Robert H. Porter. "An Empirical Perspective on Auctions." In: Mark Armstrong e Robert H. Porter (eds.). *Handbook of Industrial Organization*. Volume 3. Amsterdam, North Holland, 2007.
- Hicks, John. "Annual Survey of Economic Theory: The Theory of Monopoly." *Econometrica*, 3, pp. 1-20, 1935.
- Hicks, John. *Value and Capital*. London, Oxford University Press, 1946.
- Ho, Kate. "Insurer-Provider Networks in the Medical Care Market." *American Economic Review*, 99 (1), 2009.
- Holmes, Thomas e Jim A. Schmitz, Jr. "Competition and Productivity: A Review of Evidence." *Annual Review of Economics*, 2, 2010.
- Hoppenhayn, Hugo e Richard Rogerson. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy*, 101 (5), 1993.
- Hortacsu, Ali e Chad Syverson, "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices," *Journal of Political Economy*, 115, 2007.
- Hughes, Jonathan E., Christopher R. Knittel, e Daniel Sperling. "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand." *The Energy Journal*, 29 (1), 2008, pp. 93-114.
- Hulten, Charles. "Total Factor Productivity: A Short Biography." In: Charles R. Hulten, Edwin R. Dean, e Michael J. Harper, (orgs.) *New Developments in Productivity Analysis*. Chicago, University of Chicago Press, 2001.

- Joskow, Paul L. "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants." *Journal of Law, Economics, and Organization*, 1 (1), 1985.
- Kalouptsidi, Myrto. "Detection and Impact of Industrial Subsidies: The Case of Chinese Shipbuilding." *Review of Economic Studies*, 85, 2018.
- Lee, David S., Justin McCrary, Marcelo J. Moreira, e Jack Porter. "Valid *t*-ratio Inference for IV." *American Economic Review*, 112 (10), 2022.
- Lee, Robin S., Lecture Notes on Vertical Markets. Harvard University, 2018.
- Leibenstein, H. "Allocative efficiency vs. "X-efficiency"." *American Economic Review*, 56 (3), 1966.
- Levinsohn, James e Amil Petrin. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70 (3), 2003.
- Mankiw, N. Gregory, e Michael D. Whinston. "Free Entry and Social Inefficiency." *Rand Journal of Economics*, 17, 1986.
- Marschak, Jacob e William H. Andrews, Jr. "Random Simultaneous Equations and the Theory of Production." *Econometrica*, 12, pp. 143-205, 1944.
- Mason, Edward S. "Price and Production Policies of Large-Scale Enterprise." *American Economic Review*, 29, pp. 61-74, 1939.
- Mason, Edward S. "The Current State of the Monopoly Problem in the United States." *Harvard Law Review*, 62, pp. 1265-1285, 1949.
- Mikusheva, Anna, e Jesse M. Shapiro. "Isaiah Andrews, 2021 John Bates Clark Medalist." *Journal of Economic Perspectives*, 36 (1), 2022.
- Miller, Nathan H., Steven Berry, Fiona Scott Morton, Jonathan Baker, Timothy Bresnahan, Martin Gaynor, Richard Gilbert, George Hay, Ginger Jin, Bruce Kobayashi, Francine Lafontaine, James Levinsohn, Leslie Marx, John Mayo, Aviv Nevo, Ariel Pakes, Nancy Rose, Daniel Rubinfeld, Steven Salop, Marius Schwartz, Katja Seim, Carl Shapiro, Howard Shelanski, David Sibley, e Andrew Sweeting. "On The Misuse of Regressions of Price on the HHI in Merger Review." *Journal of Antitrust Enforcement*, 10 (2), 2022.
- Miller, Nathan H., e Matthew C. Weinberg. "Understanding the Price Effects of the MillerCoors Joint Venture." *Econometrica*, 85 (6), 2017.
- Moreira, Marcelo J. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica*, 71, pp. 1027-1048, 2003.
- Moszkowski, Erica. Notes on Demand Estimation. Não-publicado, 2019.



- Muendler, Marc-Andreas. "Trade, Technology, and Productivity: a Study of Brazilian Manufacturers, 1986-1998." Unpublished manuscript, 2004.
- Noel, Michael D. "Edworth Price Cycles: Evidence from the Toronto Retail Gasoline Market." *Journal of Industrial Economics*, 55 (1), 2007.
- Nosko, Chris. "Competition and Quality Choice in the CPU Market." Chicago Booth working paper, 2014.
- Olley, Steven e Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64, 1263–1297, 1996.
- Pakes, Ariel. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica*, 54 (4), 1986.
- Pakes, Ariel. "A Reconsideration of Hedonic Price Indexes with an Application to PC's." *American Economic Review*, 93 (5), 2003.
- Pakes, Ariel. Markups. 2018 (<https://scholar.harvard.edu/files/pakes/files/ftcnovember2018.pdf>).
- Pakes, Ariel. IO Class Notes: Collusion. Harvard University, 2019a.
- Pakes, Ariel. "A Helicopter Tour of Some Underlying Issues in Empirical Industrial Organization." *Annual Review of Economics*, 13, 2021.
- Pakes Ariel, Jack Porter, Kate Ho, and Joy Ishii. "Moment Inequalities and Their Application." *Econometrica*, 83 (1), 2015.
- Pflueger, Carolin E. e Su Wang. "A Robust Test for Weak Instruments in Stata." *Stata Journal*, 15 (1), 2015.
- Pinker, Frank. Demand Estimation Notes. Não-publicado, 2021.
- Porter, Robert H. "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886." *Bell Journal of Economics*, 14, 1983.
- Porter, Robert H. "Detecting Collusion." *Review of Industrial Organization*, 26 (2), pp. 147-167, 2005.
- Porter, Robert H. e Douglas Zona. "Detection of Bid Rigging in Procurement Auctions." *Journal of Political Economy*, 101, pp. 518-538, 1993.
- Reiss, Peter C. "Just How Sensitive are Instrumental Variable Estimates?" *Foundations and Trends in Accounting*, 10, No. 2-4, pp. 204-237, 2016.
- Reitzig, Markus. "Strategic Management of Intellectual Property." *MIT Sloan Management Review*, 45 (3), 2004.

- Rezende, Giuliana B. *Uma Análise sobre a Demanda por Imóveis no Mercado Imobiliário de Brasília*. Monografia de Graduação, Universidade de Brasília, 2009.
- Riordan, Michael H. “Competitive Effects of Vertical Integration.” In: *Handbook of Antitrust Economics*, edited by Paolo Buccirossi. Cambridge, MIT Press, 2008.
- Robinson, Joan. *The Economics of Imperfect Competition*. Londres, Macmillan, 1933.
- Robinson, Marc. “Collusion and the Choice of Auction.” *Rand Journal of Economics*, 16, 1985.
- Rotemberg, Julio J. e Garth Saloner. “A Supergame-Theoretic Model of Price Wars During Booms.” *American Economic Review*, 76 (3), 1986.
- Schmalensee, Richard, “Gaussian Demand and Commodity Bundling,” *Journal of Business*, 62, 1984.
- Schmalensee, Richard. “Inter-Industry Studies of Structure and Performance.” In: Richard Schmalensee e Robert Willig (orgs.). *Handbook of Industrial Organization*. Amsterdam: North-Holland, 1989.
- Schumpeter, Joseph A. *Capitalism, Socialism, and Democracy*. 1942.
- Shapiro, Carl. Unilateral Effects Calculations. Não-publicado, 2010.
- Shapiro, Carl e Hal S. Varian. *A Economia da Informação*. Rio de Janeiro, Campus, 1999.
- Staiger, D. e James H. Stock. “Instrumental variables regression with weak instruments.” *Econometrica*, 65, 1997.
- Stigler, George. *The Organization of Industry*. Homewood, Richard D. Irwin, 1968.
- Stone, Alan. *Regulation and Its Alternatives*. Washington, Congressional Quarterly Press, 1982.
- Stock, James H. e Mark W. Watson. *Introduction to Econometrics*. 3rd edition. New York, Pearson, 2015.
- Stock, James H. e Motohiro Yogo. “Testing for Weak Instruments in Linear IV Regression.” In: Donald W. Andrews (org.), *Identification and Inference for Econometric Models*. New York, Cambridge University Press, 2005.
- Stone, Richard. “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand.” *Economic Journal*, 64 (1), 1954.
- Sutton, John. “Gibrat’s Legacy.” *Journal of Economic Literature*, 35 (1), 1997.
- Tirole, Jean. *The Theory of Industrial Organization*. Cambridge, MIT Press, 1988.

- Verboven, Frank. “International Price Discrimination in the European Car Market.” *Rand Journal of Economics*, 27, 1996.
- Villas-Boas, Sofia B. “Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data.” *Review of Economic Studies*, 74 (2), 2007.
- Viscusi, W. Kip, John M. Vernon e Joseph E. Harrington, Jr. *Economics of Regulation and Antitrust*. Cambridge, MIT Press, 1998.
- Williamson, Oliver. “Transactions-cost Economics: the Governance of Contractual Relations.” *Journal of Law and Economics*, 22, 1979.
- Whinston, Michael D. *Lectures on Antitrust*. Cambridge, MIT Press, 2006.
- Wollmann, Thomas. “Trucks without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles.” *American Economic Review*, 108(6), 2018.
- Wollmann, Thomas. “Stealth Consolidation: Evidence from an Amendment to the Hart-Scott-Rodino Act.” *American Economic Review: Insights*, 1 (1), 2019.
- Wooldridge, Jeffrey M. “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables.” *Economics Letters*, 104 (3), pp. 102-114, 2009.
- Wooldridge, Jeffrey M. *Introdução à Econometria: Uma Abordagem Moderna*. 6ª edição. São Paulo, Cengage Learning, 2016.
- Wright, Philip G. “Moore’s Economic Cycles.” *Quarterly Journal of Economics*, 29, 1915.
- Wright, Philip G. *The Tariff on Animal and Vegetable Oils*. New York, Macmillan, 1928.