

ESTIMAÇÃO E IDENTIFICAÇÃO DE DEMANDA E DE OFERTA*

Cristian Huse
Da London School of Economics

Alberto Salvo
Da London School of Economics

1.1 INTRODUÇÃO

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ou, denotando a fatia (participação) da receita total da indústria correspondente

ao produto i (*share* de valor) como $s_i(p) = \frac{p_i D_i(p)}{\sum_{j=1}^n p_j D_j(p)}$, e lembrando a definição

das elasticidades-preço própria e cruzada, temos:³

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

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

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

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

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

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

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

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

ou:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Assim, estima-se que o *markup* do produto 1 aumente de 50% para 52% mediante a fusão, ao passo que o *markup* do produto 2 eleve-se de 33% para 40%. Com relação ao efeito da fusão sobre preços, assumindo inicialmente que a fusão não alterará custos marginais, ou seja, que *eficiências não são previstas* ($\tilde{c}_i = c_i$), temos a partir dos *markups* para o produto 1 que⁶ $c_1 = (1 - 0.5) p_1^{pre}$ e $\tilde{c}_1 = c_1 = (1 - 0.520) p_1^{pos}$, e, portanto, $p_1^{pos} / p_1^{pre} = (1 - 0.5) / (1 - 0.520) = 1.042$,

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

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

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

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

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

1.2 DEMANDA

1.2.1 Especificação geral e considerações iniciais

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

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

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

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

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

1.2.2 Produtos homogêneos versus diferenciados

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

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

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

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

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

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

Um exemplo de especificação de demanda por produto homogêneo. Estaremos analisando posteriormente vários aspectos da estimação estrutural da indústria de açúcar dos Estados Unidos na virada do século XX, por Genesove e Mullin (1998). Convém, portanto, ilustrar aqui a especificação de demanda por produtos homogêneos fazendo referência a este trabalho (relegamos a discussão sobre identificação de demanda à Subseção 2.5). Os autores estimam (13) utilizando quatro formas funcionais alternativas⁹ — como veremos na Seção 5, o intuito é verificar a robustez dos resultados à escolha de forma funcional — a saber:

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

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

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

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

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

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

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

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

Pela equação (14) vê-se claramente que um desafio presente na estimação de demanda por produtos diferenciados é o elevado número de parâmetros a serem estimados.¹¹ Com J bens ($J \geq 100$, em estudos clássicos tanto no caso de automóveis, como no caso de cereais), o número de elasticidades-preço (próprias e cruzadas) a ser estimado é da ordem de J^2 , de modo que mesmo um sistema de demanda extremamente restritivo consistiria em um grande número de parâmetros a serem estimados¹² (por contraste, há apenas uma elasticidade-preço a ser estimada em um mercado de produtos homogêneos). Como veremos nos modelos a seguir, maneiras de lidar com esse número excessivo de parâmetros incluem (além da opção de agregar produtos): *a*) decompor a decisão de compra do consumidor em estágios, de forma que alguns bens são melhores substitutos de determinada marca do que outros; *b*) utilizar restrições advindas da teoria econômica para restringir padrões de substituição, como restrições de homogeneidade; e *c*) modelar a decisão de consumo de forma que o consumidor escolha produtos indiretamente através das características que esses produtos oferecem, em vez de escolher diretamente os produtos. Como o número de características pode ser substancialmente inferior ao número de produtos (por exemplo na demanda por veículos), a dimensão do problema é reduzida.

1.2.3 Produtos diferenciados 1: modelos contínuos

1.2.3.1 Sistema de escolha em estágios múltiplos

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

onde α_n é um efeito fixo específico à região que novamente captura diferenças demográficas e de preferências invariantes no tempo, as variáveis Z_{nt} são definidas como na equação inferior (15), e ε_{nt} corresponde ao erro econométrico.²⁰

Elasticidades-preço de demanda. As elasticidades-preço próprias (para $i = j$) ou cruzadas (para $i \neq j$) podem ser calculadas a partir das estimativas dos coeficientes γ , β_i e γ_{ij} das equações superior e inferior, conforme segue:²¹

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

(A função indicadora $1[i = j]$ retorna 1 se $i = j$, retornando 0 no caso contrário.) Relegamos a dedução e breve discussão desta expressão ao Apêndice. Dois pontos merecem ser comentados. Primeiro, os erros-padrão das elasticidades podem ser calculados a partir dos erros-padrão associados aos coeficientes estimados (usando, por exemplo, o método delta: ver o Apêndice²²). Segundo, as estatísticas t normalmente apresentadas ao reportar a significância dos resultados devem levar em conta que a hipótese nula tem de ser adaptada, conforme a equação (18) deixa claro. Em vez de testar a hipótese nula de a estimativa ser igual a zero, ou seja, que $\eta_{ij} = 0$, hipóteses nulas mais apropriadas são $\eta_{ij} = w_j - 1$ para a significância das elasticidades próprias e $\eta_{ij} = w_j$ para a significância das elasticidades cruzadas.²³ Isso (infelizmente) poucas vezes é feito, não surpreendendo, portanto, estatísticas t apresentadas de magnitude bastante alta para as elasticidades próprias (onde o último termo de (18) é, por definição, igual a -1).

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

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

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

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

1.2.3.2 A métrica de distância

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

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

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

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

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

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

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

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

Pode-se deduzir a partir de (20) que a elasticidade-preço cruzada η_{ij} depende das características de ambas as marcas, i e j . Apesar desse aspecto do modelo parecer tão-somente sensato, tal propriedade flexível não é satisfeita no modelo *logit*, como veremos na Subseção 2.4.²⁵

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

Uma alternativa aos modelos contínuos consiste na classe de modelos de escolha discreta [McFadden (1974, 1978a e 1984); Cardell (1989); Berry (1994); Berry, Levinsohn e Pakes (1995); e Nevo (2001)]. Modelos de escolha discreta resolvem o problema da dimensionalidade tratando produtos como cestas de características²⁶ [Gorman (1955) e Lancaster (1966)]. No caso de automóveis, por exemplo, cada modelo será definido como a combinação de características tais como seu tamanho, peso, potência, consumo de combustível, e itens opcionais (ar condicionado, direção hidráulica). Assim, as preferências dos indivíduos são definidas no espaço de características, fazendo da dimensão desse espaço a dimensão relevante para a análise empírica. Cada consumidor vai então escolher a cesta de características que maximiza sua utilidade, sendo que as preferências dos consumidores diferem com relação às características. A demanda agregada, ou de mercado, é obtida como

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

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

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

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

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

onde:

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

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

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

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

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

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

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

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

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

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

a) atributos não-observados, pois o vetor de características afetando a escolha do indivíduo é conhecido apenas em parte pelo econometrista;²⁸

b) variações não-observadas nas preferências individuais, uma consequência da heterogeneidade individual, que introduz variância nas preferências;

c) erros de medida, pelo fato de as quantidades de cada característica observada não serem perfeitamente conhecidas; e

d) má especificação funcional, pelo fato de a função utilidade individual não ser conhecida, mas aproximada, o que envolve a adoção de hipóteses sobre a sua forma (linearidade, aditividade).

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

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

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

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

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

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

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

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

29. Tais conceitos serão abordados em maior detalhe adiante.

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

1.2.4.1 Modelo *logit*

O modelo *logit* pode ser escrito como:

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

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

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

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

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

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

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

resultando em

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Uma alternativa à especificação CES é dada por:

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

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

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

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

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

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

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

Modelos LDV podem ser escritos de forma geral como:

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

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

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

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

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

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

como consequência, a sua variância é dada por:

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

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

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

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

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

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

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

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

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

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

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

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

enquanto no caso de um modelo *logit*,

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

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

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

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

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

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

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

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

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

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

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

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

Normalmente, os efeitos marginais são computados nos valores médios amostrais dos regressores³⁸ (como no pacote Stata, por exemplo), e o método delta pode ser usado para obter fórmulas analíticas para os erros-padrão das estimativas. Alternativamente, podem-se comparar as elasticidades geradas pelos modelos, opção geralmente preferida em função de sua interpretabilidade econômica.

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

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

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

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

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

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

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

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

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

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

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

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

$$f(\varepsilon) = \frac{1}{\phi_1} \exp\left[-\exp\left(-\frac{(\varepsilon - \phi_2)}{\phi_1}\right) - \frac{(\varepsilon - \phi_2)}{\phi_1}\right].$$

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

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

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

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

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

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

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

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

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

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

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

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

1.2.4.3 MNL com características não-observáveis

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

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

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

42. Tal propriedade será estudada em maior detalhe adiante.

onde, agora, $\delta_{jt} = x'_{jt} \beta - \alpha p_{jt} + \xi_{jt}$. Ou seja, a utilidade que o produto j proporciona ao indivíduo i no período (e/ou mercado geográfico) t é descrita como a soma da qualidade média do bem j com um termo idiossincrático ε_{ijt} — tal índice de qualidade é composto por características observadas (pelo econometrista e pelos consumidores) x_{jt} e não-observadas (pelo econometrista, mas observadas pelos consumidores⁴³) ξ_{jt} . É importante salientar que tal modelo difere do *logit* clássico pela presença do termo adicional ξ_{jt} representando as características não-observadas dos produtos; no entanto, a hipótese de erro ε_{ijt} aditivo e i.i.d. (independente e identicamente distribuído) com distribuição de valor extremo do tipo I é mantida.

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

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

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

O primeiro fato a se notar aqui é que se tem um sistema de J equações (s_1, \dots, s_J) e J incógnitas (ξ_1, \dots, ξ_J), de forma que, para cada θ , existe apenas um ξ para o qual vale a igualdade. Intuitivamente, para obter ξ basta apenas que se *inverta* o sistema $s_j^{obs} = s_j(\xi, \dots; \theta_0)$, escrevendo ξ como função dos demais elementos.⁴⁵

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

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

45. Tal resultado técnico encontra-se fora do escopo do trabalho. Para fixar idéias, a função $s_j(\cdot)$ deve ser suficientemente “suave”. Ou seja, deve ser diferenciável em quase toda parte com relação a δ , e as derivadas devem satisfazer $\frac{\partial s_j}{\partial \delta_j} > 0$, $\frac{\partial s_j}{\partial \delta_k} < 0$, $k \neq j$. Para

sua derivação, bem como discussão dos detalhes técnicos, ver Berry (1994) e Berry, Levinsohn e Pakes (1995).

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

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

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

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

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

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

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

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

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

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

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

1.2.4.4 Uma análise crítica do modelo *logit*

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

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

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

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

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

$$P(\text{automóvel}) = P(A) = 70\%$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

daí resulta,

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

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

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

daí resulta,

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

As elasticidades são, então, dadas por:

a) Caso linear

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

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

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

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

b) Caso logarítmico

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

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

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

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

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

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

1.2.4.5 Modelo *logit* aninhado (NL)

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

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

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

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

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

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

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

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

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

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

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

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

resultando em um *market share*:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

com

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

onde:

x_{jk} e ξ_j são, respectivamente, características observadas e não-observadas dos produtos;

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

λ_k reflete o impacto da característica k sobre a utilidade média;

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

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

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

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

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

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

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

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

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

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

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

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

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

onde $\delta_j = \sum_{k=1}^K x_{jk} \lambda_k + \xi_j$. A título de comparação com a literatura anterior, vale

notar que, fixando-se um único v_p , tem-se um modelo *logit* clássico. Comparando-se tal especificação com a do modelo *logit*, a introdução de v_i 's diversos resulta na quebra da hipótese de consumidor representativo — ou na introdução de heterogeneidade entre consumidores, que reagirão a mudanças de preços relativos de acordo com o seu tipo. Intuitivamente, espera-se que tal heterogeneidade nas respostas gere padrões de substituição mais realistas entre os produtos.

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

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

A especificação logarítmica da renda é adotada para que a renda de fato influa nas escolhas do consumidor⁵³ (ver a nota 31 para maiores detalhes). O bem

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

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

exterior tem utilidade condicional indireta seguindo $u_{ij} = \alpha \ln(y_i) + \sigma_0 v_{i0} + \varepsilon_{i0}$, valendo notar que o seu preço p_0 é suposto ser 0, e σ_0 representa o desvio-padrão do componente não-observado das preferências dos consumidores, v_{i0} . Nevo (2000 e 2001) introduz o uso de atributos observados dos consumidores — mais precisamente, ele extrai realizações de características demográficas de dados de censo, o que pode ser interpretado como o uso da função distribuição empírica dessas características, tendo uma interpretação não-paramétrica.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

onde:

J é o número de observações;

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

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

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

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

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

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

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

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

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

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

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

1.2.4.7 Modelos da família GEV

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

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

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

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

onde:

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

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

ξ_j é o termo de características não-observadas do produto j .

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

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

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

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

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

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

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

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

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

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

onde G_j é a derivada parcial de j com relação a e^{δ_j} .

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

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

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

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

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

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

onde $\alpha_F = \frac{1-\rho_F}{2-\rho_F-\rho_B}$ e $\alpha_B = \frac{1-\rho_B}{2-\rho_F-\rho_B}$, fazendo com que $\alpha_F + \alpha_B = 1$ e $\rho_F, \rho_B \in (0, 1]$. Tal como no exemplo anterior, os parâmetros ρ parametrizam o grau de correlação intra-ramo entre os produtos e, caso seu valor seja igual a 1, o PD correspondente não é importante para a segmentação de mercado. Em particular, $\rho_F = 1$ ($\rho_B = 1$) implica que o modelo é um NL para *brands* (*frontier*), apenas. As *market shares* associadas são dadas por

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

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

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

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

1.2.5 Identificação de demanda

1.2.5.1 Motivação

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

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

1.2.5.2 “Classes” de instrumentos de demanda

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

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

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

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

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

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

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

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

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

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

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

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

Características de produtos. Em modelos de produtos diferenciados em que as características⁵⁹ são tratadas como exógenas (predeterminadas), é razoável supor que a competição varie com a localização, resultando em *markups* diferentes; ou seja, em última análise, os preços variam com a localização das características — idéia que aparece na literatura ao menos desde Bresnahan (1987). Seguindo tal linha, Berry, Levinsohn e Pakes (1995) supõem que $E(\xi | X) = 0$, ou seja, que as características não-observadas, condicionadas às características observadas, têm

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.3 INCORPORANDO OFERTA: CUSTOS E CONDUITA

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

1.3.1 Estimação de um modelo estrutural estático

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

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

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

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

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

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

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

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

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

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

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

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

63. Observe que em nossa exposição optamos por especificar (34), e portanto θ_{it} , de maneira diferente de Bresnahan (1989), que escreve a receita marginal como $P_t + \theta_{it} q_{it} P_1(Q_t, \cdot)$. Enquanto em nossa especificação θ_{it} pode variar entre 0 e 1 (com $\theta_{it} = 1$ correspondendo a conluio perfeito; ver adiante), na especificação de Bresnahan conluio perfeito corresponde a $\theta_{it} = Q_t/q_{it} = 1/s_{it}$. Pela especificação de Bresnahan, podemos usar o problema da firma, $\max_{q_{it}} \geq 0 P(Q_t, \cdot) q_{it} - C(q_{it}, \cdot)$, e a CPO dele resultante, $P_t + q_{it} P_1(Q_t, \cdot) (dq_{it}/dq_{it}) = C_1(q_{it}, \cdot)$ para motivar o parâmetro de conduta como uma “variação conjectural”, onde $\theta_{it} = dq_{it}/dq_{it}$. Por essa interpretação, ao incrementar sua quantidade em dq_{it} , a firma i detém uma “conjectura” com respeito ao incremento agregado resultante da indústria dQ_t . Tal interpretação tem sido criticada por não estar sedimentada em um modelo teórico de oligopólio [ver Corts (1999) e Reiss e Wolak (2002)].

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

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

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

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

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

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

1.3.1.1 Estimação 1: estimação de conduta como parâmetro livre

Vários artigos na literatura de estimação de conduta⁶⁴ estimam um sistema formado pela equação de demanda (31) e por equações de precificação (34) para cada

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

firma, de forma a obter estimativas livres dos parâmetros de demanda α (e portanto das elasticidades), custos β e conduta θ_{it} . A estimativa livre de θ_{it} pode ser interpretada como uma medida de quão próximo está o equilíbrio observado das previsões teóricas associadas a hipóteses de conduta alternativas. Por exemplo, uma estimativa de θ_{it} próxima de 0 é interpretada como evidência de que a indústria tem comportamento que tende a competitivo. Na prática, apesar da especificação (34) permitir a variação de conduta ao longo do tempo, o parâmetro de conduta é normalmente restringido a um só valor médio ao longo do tempo (ou seja, $\theta_{it} = \theta_{1i}, \forall t$), ou a um número limitado de “quebras” (por exemplo, $\theta_{it} = \theta_{1i}$ se $0 \leq t \leq T$ e $\theta_{it} = \theta_{2i}$ se $t > T$), coincidentes com eventos relevantes à indústria.⁶⁵ Adicionalmente, é bastante comum ver o sistema especificado de forma agregada, no nível de indústria, por não se dispor de dados no nível da firma — omite-se portanto o subscrito i para firma na equação de precificação (34), $C_1(\cdot)$ é então o custo marginal médio na indústria,⁶⁶ e o parâmetro de conduta θ , recebe a interpretação do “grau médio de conluio na conduta” (*average collusiveness of conduct*) [Bresnahan (1989)].

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

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

66. É fácil ver a partir de (34) que quando conduta corresponde a concorrência perfeita ($\theta_i = 0$) ou a conluio perfeito ($\theta_i = 1$), $C_1(\cdot)$ é tanto o custo médio aritmético ($\sum_i C_1(q_{it})/n$) (para ver isso, somar as equações de precificação de todas as n firmas, dividindo então por n) como o custo médio ponderado pela participação $\sum_i s_i C_1(q_{it})$ (multiplicar a equação de precificação de cada firma i por s_i e somar todas as n equações resultantes), dada a simetria de custos marginais entre firmas implícita no *equilíbrio*. Já quando conduta na indústria corresponde a Cournot ($\theta_i = s_i$), a primeira dessas manipulações algébricas (ou seja, $C_1(\cdot)$ sendo o custo médio aritmético) gera o parâmetro médio de conduta $(\sum_i \theta_i)/n = (\sum_i s_i)/n = 1/n$, enquanto a segunda manipulação algébrica ($C_1(\cdot)$ sendo o custo médio ponderado pela participação) gera o parâmetro médio de conduta $\sum_i s_i \theta_i = \sum_i s_i^2 = HHI$ (o índice de Herfindahl-Hirschman).

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

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

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

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

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

1.3.1.2 Identificação

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

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

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

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

A identificação dos parâmetros de conduta θ em uma indústria de bens diferenciados é estudada por Nevo (1998). Ele confirma a generalidade do resultado de Bresnahan (1982): em princípio os parâmetros de conduta podem ser identificados junto com os parâmetros de demanda e de custo. Na prática, porém, Nevo alerta para a dificuldade de se dispor de quantidade suficiente de variáveis exógenas de demanda. Devido ao maior número de parâmetros de conduta em uma indústria de bens diferenciados, a identificação de conduta na ausência de dados sobre custo (ou lucros) fica comprometida devido às restrições de exclusão que dificilmente serão satisfeitas. Enquanto o número mínimo de variáveis endógenas em cada equação de precificação é 1 em uma indústria de bens homogêneos — o número mínimo correspondendo ao caso mais simples de custos marginais constantes em quantidade⁷¹ —, este número mínimo em uma indústria de bens diferenciados sobe para o número de bens — novamente o número mínimo correspondendo ao caso de custos marginais constantes em quantidade — (veja-se a demonstração no próximo parágrafo). Como solução para o problema de inferir conduta juntamente com os parâmetros de demanda e de custo, Nevo propõe que seja utilizada a metodologia de seleção de modelos, apresentada na Subseção 1.3.1.4.

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

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

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

71. A variável $\frac{P}{E}$ na equação (35), por exemplo, é endógena. A depender da especificação do custo marginal $c_n = C_1(q_n, \dots)$, pode haver outras variáveis endógenas — se o custo marginal variar em função da quantidade.

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

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

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

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

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

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

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

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

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

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

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

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

1.3.1.3 Ilustração 1: estimação de conduta como parâmetro livre

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

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

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

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

1.3.1.4 Estimação 2: seleção estatística de modelos

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

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

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

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

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

1.3.1.5 Ilustração 2: seleção estatística de modelos

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

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

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

1.4 A EXPERIÊNCIA INTERNACIONAL NO ANTITRUSTE

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

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

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

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

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

1. Quais os efeitos unilaterais da fusão?

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

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

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

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

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

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

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

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

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

Muito em função da dificuldade de tratamento, os SDC's, especialmente o DOJ americano, tendem a se concentrar no estudo dos efeitos unilaterais das fusões; apesar do interesse recente do DOJ⁸³ e, principalmente, da Comissão Européia sobre o assunto,⁸⁴ o fato de implementações empíricas do estudo de efeitos coordenados não estarem devidamente desenvolvidas é um fator limitante ao seu estudo.⁸⁵

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

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

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

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

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

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

1.4.1 Modelos calibrados de simulação de efeitos unilaterais

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

1.4.1.1 O modelo PCAIDS

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

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

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

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

e pode ser escrita na forma matricial por:

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

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

86. Considerado em detalhe na Seção 1.2.

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

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

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

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

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

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

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

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

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

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

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

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

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

A matriz de elasticidades correspondente é então obtida como:

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

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

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

1.4.1.2 O modelo ALM

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Resolvendo-se a equação acima para α_j chega-se a:

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

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

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

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

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

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

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

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

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

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

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

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

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

onde $s_{m|l}$ é o *market share* combinado das firmas 1 e 2, pré-fusão.⁸⁷ Ou seja, a margem da firma é decrescente na substituíbilidade entre os bens interiores (β) e na substituíbilidade entre bens interiores e o bem exterior (η_l), e crescente no *market share* pré-fusão combinado das firmas 1 e 2 ($s_{m|l}$), em função de se ter $\beta \cdot \bar{p} - \eta_l > 0$.

1.4.1.3 Análise crítica dos modelos calibrados de simulação

Os modelos abordados nesta seção são especialmente úteis quando se trabalha sob restrições de tempo e obtenção de dados, e aí residem sua maior virtude e sua maior fragilidade. Por serem simplificações de modelos estudados anteriormente, além das propriedades indesejáveis dos modelos originais, eles herdam características indesejáveis adicionais. Seu uso é recomendado apenas em estudos de caráter exploratório, ou quando restrições severas de tempo e dados forem observadas.

1.5 CONSIDERAÇÕES PRÁTICAS NA ESTIMAÇÃO ESTRUTURAL

Nesta seção oferecemos comentários de natureza prática relevantes ao analista que tem ao seu alcance dados sobre alguma indústria e deseja estimar a estrutura dessa indústria. Procuramos tecer a discussão da forma mais intuitiva possível.

1.5.1 Requisitos de dados

Tipicamente — e deixando de lado por agora a existência de efeitos dinâmicos — identifica-se cada observação como refletindo o ponto de equilíbrio em um mercado estático. Observações podem corresponder: *a*) a um mesmo mercado geográfico ao longo do tempo, sendo que cada período no tempo é então tratado como uma realização distinta desse mercado (por exemplo, dados referentes ao Distrito Federal em janeiro de 2005 correspondem a uma realização de mercado, ao passo que dados referentes ao Distrito Federal em fevereiro de 2005 correspondem a outra realização); *b*) a diferentes mercados geográficos em um mesmo período no tempo (dados referentes a diferentes mercados geográficos são tratados como diferentes realizações de um mesmo processo de geração de dados ou de distintos processos com parâmetros em comum a serem estimados); ou *c*) a uma combinação dos dois anteriores. O valor de dados em painel — onde, por exemplo, dados referentes a diferentes mercados geográficos são coletados ao longo do tempo — provém da disponibilidade de observações tanto na dimensão temporal como na *cross-section*.

87. Para uma discussão mais detalhada, incluindo aplicações e discussões referentes a excedente do consumidor e bem-estar, ver Werden e Froeb (1994) e Werden, Froeb e Tardiff (1996).

De maneira geral, quanto maior o número de observações à disposição do analista, menor será o erro-padrão das estimativas. No entanto, não é necessariamente verdade que mais dados (ou dados mais freqüentes) sejam sempre melhores. Por exemplo, Genesove e Mullin (1998) optam por agregar dados para uma periodicidade trimestral, apesar de terem dados semanais disponíveis, em função da alta correlação serial observada nos últimos, que precisaria ser tratada de forma satisfatória (ver a seguir). Além disso, uma premissa implícita em muitos trabalhos empíricos envolvendo estimação estrutural (apesar de poucas vezes explicitada pelo pesquisador) é que as observações ao longo do tempo são realizações de um mesmo processo de geração de dados, cujos parâmetros são invariantes no tempo. Tal premissa pode fazer sentido quando o período de tempo ao qual os dados correspondem é razoavelmente curto, digamos três a quatro anos. No entanto, quando o período de tempo for mais longo, ou a indústria sob estudo estiver em franca evolução (por exemplo, telefonia móvel), o analista deverá considerar a possibilidade de que esteja ocorrendo uma variação ou quebra estrutural, com mudança nos parâmetros estruturais, como, por exemplo, uma alteração nas preferências dos consumidores. Se tal possibilidade for concreta, o analista deverá incorporar a variação ou quebra estrutural em sua especificação, adicionando — para citar um exemplo simples — efeitos de tempo ou uma variável de tempo.

Como exposto acima, para efetuar a estimação estrutural o analista irá requerer séries de preços e quantidades, seja na dimensão temporal e/ou na dimensão da *cross-section*, junto com deslocadores exógenos de demanda (incluídos na especificação de demanda, mas tipicamente excluídos da especificação de oferta, utilizados, portanto, na identificação do lado de oferta) e deslocadores exógenos de oferta (incluídos na especificação de oferta, mas tipicamente excluídos da especificação de demanda, podendo, portanto, ser utilizados na identificação de demanda).

Ilustramos os requisitos de dados com referência à estimação de modelos discretos de demanda, pelo fato de estes modelos necessitarem de uma disponibilidade maior de dados (lembrar que, nesses modelos, como o consumidor demanda características de produtos em vez de demandar produtos em si, são necessários dados referentes a características da indústria em estudo).

Dados agregados de mercado (market level data). O caso clássico do uso de dados de mercado para estimação de modelos discretos é Berry, Levinsohn e Pakes (1995). Os requisitos mínimos de dados são:

- Preços, quantidades vendidas e características (podendo incluir dados de publicidade) dos produtos no mercado.

Em termos da notação acima, tais informações nos fornecem p_j , x_j e q_j para todo produto $j = 1, \dots, J$ dentro do mercado.

Para que se obtenham os *market shares* dos produtos, necessita-se do tamanho do mercado. No caso de automóveis, tal como em Berry, Levinsohn e Pakes (1995 e 2004), para o mercado americano, uma primeira aproximação é dada pelo número de unidades familiares (que não seria tão satisfatória para o caso brasileiro, pelo fato de o país ser bem mais pobre); para Bresnahan, Stern e Trajtenberg (1997), o tamanho do mercado é tomado como sendo o número de trabalhadores de escritório no país. Em resumo, tem-se que:

- Tamanho do mercado = dados + hipóteses.

De posse do tamanho do mercado, tem-se agora s_j , $j = 0, 1, \dots, J$. Não é incomum que pesquisadores façam experimentos, alterando o tamanho do mercado a fim de checar a robustez dos resultados obtidos.

Opcionalmente, tal como de início feito em Nevo (2000 e 2001), podem-se incluir características demográficas (tamanho da unidade familiar, existência de crianças etc.) no modelo [variáveis z , em termos de notação adotada, por exemplo, em (29)]. Uma forma de fazê-lo é extrair realizações de dados do censo, o que tem uma interpretação não-paramétrica. Alternativamente, poder-se-ia estimar uma distribuição paramétrica sobre os dados disponíveis e, de posse das estimativas dos parâmetros, gerar variáveis aleatórias com essa mesma distribuição.

As principais vantagens dos dados em nível de mercado são duas:

- São mais fáceis de serem obtidos.
- O risco de viés de seleção é relativamente pequeno (em contraste com o que acontece com microdados — ver adiante).

As principais desvantagens giram em torno do tratamento menos satisfatório de heterogeneidade do consumidor:

- O fato de a estimação ser potencialmente mais difícil (há pouca informação sobre as características do consumidor que comprou determinado produto).
- O fato de se usar informações sobre a distribuição de, por exemplo, características demográficas, e não as características associadas a cada unidade familiar.

Microdados — micro data. No caso de microdados, os requisitos mínimos são:

- Preços e características (podendo incluir dados de publicidade) de todos os produtos no mercado.

- As escolhas feitas pelos consumidores.

Em termos da notação acima, tais informações nos fornecem p_{ij} , x_{ij} e q_{ij} para todo produto $j = 1, \dots, J$ dentro do mercado e consumidor $i = 1, \dots, I$.

Tal como no caso anterior, opcionalmente, pode-se incluir características demográficas de cada família no modelo.

As principais vantagens dos microdados referem-se ao tratamento de heterogeneidade:

- Oportunidade de avaliar em detalhe o impacto das características demográficas nas escolhas.
- Oportunidade de observar o consumidor/unidade familiar em vários períodos, tendo acesso a seu comportamento dinâmico.
- Em tese, a estimação é mais simples, pela maior informação fornecida (confronte-se com o caso anterior).

As principais desvantagens são:

- Dificuldade de conseguir os dados.
- Viés de seleção, ou seja, a possibilidade de a amostra considerada não ser representativa da população para a qual se deseja extrapolar o estudo (considere os efeitos de se modelar a demanda por automóveis para uma amostra de classe média alta e as conseqüências de seu uso para toda a população brasileira).
- Finalmente, a potencial inconsistência das estimativas, pela dificuldade de se controlar a endogeneidade.

Um comentário importante diz respeito à distribuição (ou cobertura) de produtos em mercados de produtos diferenciados. A situação ideal é aquela em que todas as marcas de interesse para o estudo encontrem-se sempre disponíveis no ponto de varejo no momento em que o consumidor toma sua decisão de compra. Afinal, é o que se assume ao procurar inferir a substituíbilidade entre marcas. Na prática, poucas marcas estão disponíveis o tempo todo no ponto de varejo, seja porque os estoques acabaram no estabelecimento ou porque o estabelecimento opta por não distribuir aquela marca. É possível que os erros decorrentes de tal ocorrência se anulem no agregado, ao trabalharmos com dados agregados de mercado (*market level data*) ou ao agregarmos as decisões individuais de cada consumidor quando os dados em questão sejam microdados. Mas é pouco provável que tal distorção seja trivial quando da introdução de uma nova marca, em que sua

distribuição esteja em evolução ao longo de um período mais longo de tempo. Sempre que possível valerá a pena para o analista procurar obter informações a respeito de distribuição, incorporando-as à estimação. Tais dados são coletados por agências de coleta de dados como a ACNielsen e a IRI⁸⁸ [para algumas aplicações ver Mariuzzo, Walsh e Whelan (2004) e Slade (2004)].

Vale mencionar ainda o cuidado que deverá ser tomado pelo analista quando a prática de descontos ao consumidor através de cupons ou outras vantagens for difundida em determinada indústria sob estudo (ainda que tal prática no Brasil seja tímida em comparação com, por exemplo, os Estados Unidos). O desafio que se apresenta aqui é que dados sobre cupons poderão não ser observados pelo analista, ou observados somente imperfeitamente: como então obter os preços líquidos pagos pelos consumidores? Vale também mencionar que os preços coletados por agências como ACNielsen podem corresponder não aos preços médios no período (ou seja, receita no período dividida por quantidade vendida no período) mas sim a preços praticados *no momento em que a coleta é feita* (por exemplo, uma vez ao mês, ou uma vez por semana), podendo esse preço variar substancialmente entre um momento de coleta e o próximo. Essa distorção não estará presente quando o analista estiver trabalhando com microdados.

1.5.2 Nível de agregação de dados

Uma consideração prática importante diz respeito ao nível de agregação dos dados. A agregação pode se dar em várias dimensões: *a*) agregação no espaço geográfico; *b*) agregação por embalagem, versão ou sabor; *c*) agregação por canal de vendas (e, para cada canal de vendas, por estabelecimento); e *d*) agregação no tempo. A título de ilustração, um analista hipotético que estivesse examinando a indústria de bebidas carbonatadas (refrigerantes), com toda sorte de dados a sua disposição, poderia optar por um nível de (des)agregação tal que uma observação equivaleria a um vetor de preços, quantidades vendidas e características para cada sabor (laranja), para cada marca (Fanta), para cada tamanho de embalagem (garrafa familiar de 2 litros em PET), correspondente a vendas através de supermercados com cinco ou mais caixas (*checkouts*) no município de Campinas na semana de 5 de fevereiro de 2005. Em outro extremo, uma observação poderia ser definida como um vetor de preços, quantidades vendidas e características (médias) por marca (Fanta), para

88. Um ponto semelhante deve ser feito com respeito à tentativa pelo analista de obter dados sobre promoções realizadas por produtores e/ou varejistas junto a consumidores. A literatura em *marketing* demonstra claramente a eficácia de promoções em aumentar o volume de vendas. Apesar de a teoria econômica não informar de maneira explícita como essa atividade promocional e publicitária deve ser modelada, faz sentido conceber que a posse pelo analista de dados sobre promoções poderá melhorar sensivelmente a acurácia das estimativas.

todos os sabores e tamanhos de embalagem, através de todos os canais de venda (varejistas *self-service*, varejistas tradicionais com balcão de vendas e bares), para todo o Brasil, no ano de 2005.

Naturalmente, quanto mais desagregados estiverem os dados, maior é o conjunto de possibilidades que o analista terá em mãos na hora de partir para a modelagem e a estimação (desde que a coleta de dados de maneira mais desagregada não tenha introduzido maiores erros de medida). No entanto, a utilização dos dados na forma mais desagregada possível não necessariamente será a mais indicada. Em alguns casos, os dados em nível demasiadamente desagregado, em virtude do processo de coleta, poderão conter muito ruído. Será aconselhável que o analista então proceda à agregação dos dados. Mas qual será o nível ótimo de agregação? A resposta aqui dependerá de questões institucionais e muito provavelmente não haverá uma resposta definitiva, devendo o analista testar a robustez de seus resultados estimando, quando possível, dados com diferentes níveis de agregação. Uma boa precaução consiste em considerar se os níveis de agregação não ignoram importante variação que possa estar ocorrendo em níveis menos agregados. Voltando ao exemplo de refrigerantes, conhecimento institucional poderá sugerir que a estimação de elasticidades-preço deverá utilizar dados suficientemente desagregados por embalagem e por canal de vendas. A sensibilidade a preços do consumidor de Guaraná Antarctica na embalagem familiar de 2 litros, vendida através de grandes supermercados, certamente será distinta da sensibilidade a preços do consumidor de Guaraná Antarctica na embalagem de consumo individual de 350 ml, vendida através de bares (assim como a distribuição dessa mesma marca poderá ser tradicionalmente mais forte no canal bar do que no canal varejo). Ao estimar uma equação de demanda, o analista poderá então utilizar: *a*) todos os dados, porém de forma desagregada por embalagem e canal de vendas, incluindo, portanto, efeitos fixos em sua especificação (assim como interações desses efeitos fixos com, por exemplo, preços); ou *b*) dados referentes a cada tamanho de embalagem e cada canal de vendas separadamente, estimando uma equação distinta para cada combinação de embalagem e canal. Similarmente, a agregação na dimensão geográfica deverá considerar a possibilidade de que as preferências de consumidores, assim como o comportamento das firmas servindo esses consumidores, poderão ser distintos em diferentes mercados geográficos. Por exemplo, por motivos institucionais, a penetração de bebidas dietéticas (por exemplo, *Coca Light*) é sabidamente maior na cidade do Rio de Janeiro do que em outras grandes cidades como São Paulo. De posse de dados em painel, um analista, estimando uma equação como a (15), deverá então permitir tal variação através da inclusão em sua especificação de

efeitos fixos de cidade-marca (e possivelmente interagindo esses efeitos com preços, sendo a identificação do coeficiente de preços dada, então, pela variação temporal).⁸⁹

Quanto à agregação no tempo, pode haver um *trade-off* entre eficiência (quanto maior o número de observações, maior o número de graus de liberdade) e correlação serial: é concebível que um analista, dispondo de observações com variação temporal semanal, opte por agregar os dados na dimensão temporal para períodos mensais. Apesar de estar reduzindo o número de observações em $3/4$, a correlação serial entre observações mensais poderá ser menor do que entre observações semanais, podendo ser tratada de modo mais satisfatório, utilizando-se, por exemplo, o estimador de covariância de Newey e West (1987).

1.5.3 Escolha de variável dependente (regressão reversa)

Em algumas situações, particularmente na estimação de demanda por produtos homogêneos, o analista deve escolher não somente a forma funcional (ver subseção adiante) como também a variável dependente, a saber, preços ou quantidades — por exemplo, ao passo que Genesove e Mullin (1998) tomam quantidades como a variável dependente, Parker e Roller (1997) escolhem preços. Aqui, até onde sabemos, a literatura é escassa. A análise dos resíduos (*fitted residuals*) poderá dar subsídios ao analista com respeito a qual regressão é preferível. Vale ressaltar que a robustez das estimativas à escolha de determinada variável dependente pode ser verificada através da estimação da regressão reversa. Carlton (2003) comenta que “parece ser uma regularidade empírica que um (pesquisador) freqüentemente pode obter uma elasticidade alta ao simplesmente estimar preços em quantidades do que o inverso!” (p. 8; tradução livre).

89. Em diversos momentos é concebível que o analista trabalhe, mediante algumas premissas, em nível *menos* agregado do que o nível de agregação em que os dados estão disponíveis. Para citar um exemplo [Salvo (2005)], assumo que dados de demanda por cimento (preços, quantidades, atividade na construção civil etc.) estão disponíveis somente por *estado* (unidade da federação), mas o analista deseja trabalhar no nível de *município*. (Note que há em média 200 municípios por estado, totalizando aproximadamente 5.600 municípios no Brasil.) Uma possibilidade consiste em, inicialmente, estimar uma curva de demanda para cada estado a partir dos dados estaduais. Observe que a curva de demanda (inversa) de determinado estado é o resultado da soma horizontal das curvas de demanda (inversas) de todos os municípios localizados naquele estado. Agora assumo que o preço de reserva do consumidor que mais valoriza o produto (*o highest-valuation consumer*) em cada município não varia entre municípios naquele estado (ou seja, o intercepto no eixo de preço é o mesmo para a curva de demanda de cada município naquele estado). Assumo também que na medida em que o tamanho de mercado dos municípios (dado pelo tamanho de suas populações) varia, as curvas de demanda dos municípios alteram-se de maneira “multiplicativa”, ou seja, elas giram ao redor do intercepto de preço. A partir da curva de demanda (inversa) do estado S dada por, digamos, $p = \alpha_S - \beta_S q$, o analista pode obter uma estimativa da curva de demanda do município M , ou seja $p = \alpha_M - \beta_M q$, onde $\alpha_M = \alpha_S$ e $\beta_M = \frac{Y_S}{Y_M} \beta_S$, sendo que Y_M denota a população do município M e $Y_S := \sum_{M \in S} Y_M$ é a população total do estado S . É fácil verificar que (a um dado preço) a elasticidade-preço de demanda no nível do estado equivale à elasticidade-preço de demanda no nível do município.

1.5.4 Implementação prática

Tratamos da implementação prática de modelos discretos, visto que as versões mais sofisticadas exigem a programação de rotinas complexas por parte do analista.

1.5.4.1 Implementação de modelos discretos⁹⁰

Ao se implementar modelos discretos como o *logit*, a variável dependente usada é a diferença entre o logaritmo do *market share* de cada produto subtraído do logaritmo do *market share* do bem exterior, $\ln(s_j) - \ln(s_0)$ (obtidos após a definição do mercado potencial, M) e regressores mais usados — além do preço — são despesas com publicidade, características dos produtos estudados, variáveis *dummies* para marcas e períodos distintos. Para as versões clássicas do modelo *logit*, propriedades teóricas dos estimadores podem ser obtidas em Amemiya (1985), ou ainda em livros-texto como Wooldridge (2002).

Nosso tratamento se baseia principalmente no GMM. Para chegar à formulação do modelo, considere as fórmulas dos *market shares logit*,

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

$$s_0 = \frac{1}{1 + \sum_{k=1}^J \exp(\delta_k)}$$

divida a primeira equação pela segunda, e tome o logaritmo para obter $\ln(s_j) - \ln(s_0) = \delta_j$. Recorde que no modelo *logit* com características não-observadas [Berry (1994)], tem-se $\delta_j = x_j' \theta + \xi_j$ ($\beta \subseteq \theta$), onde x_j engloba características observadas mais o preço do produto j . Nesse caso, a solução do sistema relacionando *shares* geradas pelo modelo a *shares* observadas é trivial, fornecendo o que interpretamos como sendo o termo de erro

$$\xi_j(\theta, x, s) = \ln(s_j) - \ln(s_0) - x_j' \theta$$

90. A maior parte dos estimadores discutidos nesta subseção encontra-se implementada no pacote econométrico *Stata* — os modelos e os respectivos comandos são: *Logit - logit*; *Probit - probit*; *Multinomial logit - mlogit*; *Nested logit - nlogit*. Além disso, o comando para o método de variáveis instrumentais é o *ivreg*.

Uma vez ilustrado esse passo fundamental, recordamos e aprofundamos a discussão sobre aspectos teóricos do modelo, que valem independentemente da especificação do mesmo — usaremos o *logit* como exemplo pela sua tratabilidade.

Primeiro, vale notar que as propriedades assintóticas do modelo estão relacionadas ao número de produtos J , bem como ao tamanho da amostra de consumidores a partir da qual se calculam os *market shares*. Segundo, algumas das características observadas — tipicamente preço — podem estar correlacionadas com ξ , requerendo o uso de instrumentos. Terceiro, ressaltamos que os *shares* $s(\theta, x, \xi, P)$ são gerados pela agregação das demandas individuais dos agentes (indivíduos, unidades familiares etc.) com distribuição de preferências P , por exemplo, a função distribuição empírica de uma amostra aleatória gerada pela distribuição verdadeira de preferências, P^0 . Quarto, recordamos o sistema relacionando *market shares* observados e *market shares* gerados pelo modelo, $s(\theta_0, x, \xi, P^0) = s^{obs}$, onde θ_0 é o valor verdadeiro do vetor de parâmetros. A estratégia a ser seguida para estimação é exatamente aquela descrita na seção referente a Berry, Levinsohn e Pakes (1995), onde a minimização da distância entre *shares* observados e *shares* gerados pelo modelo, $\|s(\theta_0, x, \xi, P^0) - s^{obs}\|$ é mostrada como não sendo factível (tanto em função das não-linearidades envolvidas, como pelo fato de os argumentos serem funções uns dos outros), levando à proposição da adoção do GMM [Berry (1994)].

Voltando-se ao caso *logit*, interprete-se o termo ξ_j como um erro de um modelo de regressão de *shares* observados em características observadas. Dada a possibilidade de correlação entre componentes de x e o erro ξ , considere a decomposição $x := (x_1, x_2)$, onde x_1 são características de produtos exógenas, e x_2 são características endógenas (o preço, para fixar idéias). A hipótese de identificação clássica é supor a existência de um conjunto de instrumentos z , tipicamente constituído por x_1 (instrumentos incluídos) e instrumentos (excluídos) para x_2 (preço), e a condição de momento amostral associada é dada por:

$$\begin{aligned} G_J(\theta, s, P^0) &= \frac{1}{J} \sum_{j=1}^J z_j \cdot \xi_j(\theta, x, s, P^0) \\ &= \frac{1}{J} \sum_{j=1}^J z_j \cdot [\ln(s_j) - \ln(s_0) - x_j' \theta] \end{aligned}$$

É interessante notar que o estimador acima pode ser melhor entendido como um estimador IV (variáveis instrumentais) linear, de acordo com os resultados de Berry (1994), podendo ser estimado usando-se 2SLS.

Em termos teóricos, a condição de momento resulta da hipótese de covariância nula entre o conjunto exógeno de instrumentos, z , e características não-observadas, ξ , via

$$E[G_J(\theta)] := E\left[\frac{1}{J} \sum_{j=1}^J z_j \cdot \xi_j(\theta, x, s, P^0)\right] = 0$$

em $\theta = \theta_0$ (valor verdadeiro do vetor de parâmetros), e a estimativa de θ é obtida pela minimização de uma norma da condição de momento amostral,

$$\frac{1}{J} \sum_{j=1}^J z_j \cdot \xi_j(\theta, x, s, P).$$

As hipóteses sobre $\{\xi_j\}_{j=1}^J$ são muito semelhantes às feitas sobre o erro num modelo de regressão linear. Primeiro, supõe-se que sejam variáveis aleatórias independentes. Além disso, supõe-se que $E[\xi_j | x_j] = 0$ e $\sup_j E[\xi_j^2 | x_j] < \infty$ com probabilidade igual a 1. Vale ressaltar que tais hipóteses permitem que tal processo exiba heterocedasticidade condicional.

Como mencionado anteriormente, nem sempre é possível avaliar os *market shares* gerados pelo modelo, $s(x, \xi, \theta, P^0)$, analiticamente, de forma que é preciso simular a distribuição P^0 , calculando-se $s(x, \xi, \theta, P^m)$. O acréscimo na variância decorrente do erro de simulação pode, como mencionado, ser estimado por meio de um procedimento de *bootstrap*, tal como destacado na seção sobre BLP.⁹¹

1.5.5 Comparação e escolha de formas funcionais (demanda e oferta)

A imposição de uma forma funcional é um mal necessário em virtude do problema de dimensionalidade. Idealmente poderíamos utilizar métodos não-paramétricos caso dispuséssemos de quantidade suficiente de dados, mas na prática há de se adotar estrutura. No entanto, conforme discutido anteriormente, a premissa de forma funcional não deverá estar informando o padrão de substituição,⁹² papel

91. Para uma discussão sobre *bootstrap*, ver Wooldridge (2002, cap. 12).

92. A título de ilustração, vale voltar à Subseção 2.4.4, onde é feita uma análise crítica do modelo *logit*, em que o segundo exemplo compara as elasticidades das especificações lineares e logarítmicas em preços. A lição, no caso do *logit*, é que as elasticidades são consequência direta da forma funcional empregada. Assim, sabe-se que um modelo linear em preços vai dar uma elasticidade mais alta que um modelo logarítmico *a priori*, para preços maiores que a unidade.

este que deve ser desempenhado pelos dados: nesse sentido, é desejável que o modelo de demanda seja suficientemente “flexível”.⁹³

Vale lembrar que a robustez das estimativas à escolha da forma funcional pode (e deve) ser verificada através da estimação de diferentes formas funcionais para verificar a robustez. Por exemplo, na estimação de demanda por um produto homogêneo, Genesove e Mullin (1998) estimam quatro diferentes formas funcionais, a saber: linear, log-linear, exponencial e quadrática. Afortunadamente, as elasticidades-preço de demanda estimadas pelos autores segundo as quatro especificações são próximas, sugerindo que a escolha de forma funcional não está gerando a estimativa de elasticidade. Menos óbvia é a estratégia que o analista deverá seguir caso as estimativas dependam sensivelmente da escolha de forma funcional, e em que a análise dos resíduos não recomende uma forma funcional (ou um subconjunto destas) em detrimento de outras. Nesse caso, o analista deverá reportar o achado e possivelmente ampliar o intervalo de confiança de suas estimativas, levando em consideração um intervalo que englobe os intervalos estimados pelas diferentes formas funcionais.

Já em uma indústria de produtos diferenciados, Crooke *et alii* (1999) mostram como a escolha de forma funcional pode afetar sensivelmente a magnitude dos efeitos de um ato de concentração sobre preços. Através de experimentos de Monte Carlo, eles obtêm aumentos de preços consideravelmente maiores com funções de demanda de elasticidade constante do que com funções de demanda lineares ou *logit*.⁹⁴ Werden, Froeb e Tschantz (*forthcoming*) argumentam, à luz do Antitrust Logit Model (ALM), que o modelo *logit* tende a gerar previsões de aumentos de preços modestos em face de atos de concentração.⁹⁵

Reproduzimos (além de elaborarmos) aqui alguns comentários feitos por Hosken *et alii* (2002) e Werden *et alii* (1994) a respeito das vantagens e desvantagens de diferentes formas funcionais de demanda (no que tange, particularmente, à estimação de demanda de produtos diferenciados):

- Sistema de demanda linear: A principal vantagem é a simplicidade no cálculo dos efeitos de uma fusão [ver Werden, Froeb e Tardiff (1996) para maiores

93. Uma excelente discussão sobre a escolha de formas funcionais na modelagem de demanda e a flexibilidade oferecida por diferentes modelos é dada por Deaton e Muellbauer (1980b).

94. Intuitivamente, na medida em que o preço aumenta, a demanda se torna mais elástica *mais rapidamente* nas funções de demanda linear e *logit*, em comparação com a função de demanda log-linear (onde a elasticidade permanece constante). Esse aumento rápido da elasticidade atua como um freio no aumento de preços decorrente da fusão.

95. Werden (1996) propõe uma maneira de analisar um ato de concentração que passa ao largo da escolha da forma funcional e, portanto, do efeito desta sobre a magnitude de variação de preços: esta análise consiste em computar a redução de custos marginais necessária para exatamente compensar o efeito unilateral da fusão sobre preços. Como esse cálculo envolve, por definição, preços pós-fusão iguais a preços pré-fusão, a forma como as elasticidades variam em função de preços — e portanto a escolha da forma funcional — não mais é importante.

detalhes]. Entre as suas desvantagens está a falta de garantia de que as elasticidades-preço cruzadas tenham o sinal “esperado”: por exemplo, a elasticidade-preço cruzada entre marcas substitutas (assumindo que o efeito renda seja baixo⁹⁶) deveria ser positiva, mas nem sempre o é.

- Sistema de demanda log-linear (elasticidade constante): A vantagem desse sistema é que as elasticidades podem ser lidas imediatamente a partir dos coeficientes de preço estimados pela especificação. Entre as desvantagens, está *a*) a restrição de que as elasticidades permaneçam constantes — em vez de aumentar, o que seria mais natural no caso da elasticidade própria — na medida em que o preço aumenta; *b*) a inversão comum de sinais em relação ao esperado; e *c*) restrições provenientes de teoria econômica, como a soma das fatias de gastos do consumidor sendo igual a 1, não são necessariamente satisfeitas.

- Sistema de demanda AIDS: Entre suas vantagens está a importante propriedade de flexibilidade. Hosken *et alii* (2002) afirmam que “muitas das estimativas de sistemas de demanda apresentados às Agências (de antitruste, como FTC e DOJ) usam o sistema de escolha em estágios múltiplos com a especificação de demanda AIDS” (p. 26; tradução livre). Entre as desvantagens, está a necessidade de estimar um elevado número de parâmetros e, portanto, a necessidade de o analista dispor de uma base de dados satisfatória para manter a variância de suas estimativas dentro de um patamar aceitável (além da necessidade de o analista adotar premissas identificadoras que podem suscitar alguma polêmica em certas situações — ver a Subseção 1.2.5).⁹⁷

- Sistema de demanda *logit* (MNL): Entre as vantagens, estão a facilidade de estimação e a baixa exigência de dados (devido ao reduzido número de parâmetros). Como vimos, essa facilidade vem a um custo — a falta de flexibilidade: a propriedade IIA (independência das alternativas irrelevantes) pode ser irrealista. Lembrar que no modelo *logit*, por exemplo, se o preço de um produto sobe, os consumidores substituem-no por outros produtos em proporção aos *shares* desses outros produtos.⁹⁸ Isso decorre de a elasticidade-preço cruzada de demanda de cada produto $j \neq i$ com relação ao preço do produto i ser a mesma — não levando em conta informações

96. Lembrar a equação de Slutsky.

97. Outra limitação de sistemas desse tipo (contínuos) é o tratamento da heterogeneidade individual. O analista poderá ter acesso a dados ao nível do consumidor (microdados) mas não fica clara a forma como esses dados devem ser incorporados ao sistema.

98. Em defesa do modelo *logit* (ALM), Werden *et alii* (1994) citam o argumento de Willig (1991) de que a propriedade IIA é justificada em uma situação em que as marcas, sendo fusionadas, não estão situadas a uma distância particularmente próxima ou distante no espaço de características. Werden e seus co-autores afirmam que, em virtude do papel central de *shares* no modelo *logit*, “o ALM serve, portanto, como um filtro comparável ao filtro oferecido pelos *market shares* na análise antitruste tradicional, sendo o ALM uma melhoria substancial por oferecer predições quantitativas sobre aumentos de preços” (p. 9; tradução livre).

sobre a proximidade desses produtos no espaço de características. Note que como nesse modelo a substituição é proporcional aos *shares* de mercado, uma fusão entre marcas com *shares* elevados levará a um aumento de preços maior do que uma fusão entre marcas com *shares* pequenos. Além disso, vale lembrar que as elasticidades são consequência direta da forma funcional adotada para preços, como demonstrado na Subseção 1.2.4.4.

- Generalização do *logit* 1 — Sistema de demanda *logit* aninhado (NL): Ao passo que no modelo *logit* o componente aleatório de utilidade está distribuído de forma idêntica e independente entre consumidores, o que não possibilita a correlação de preferências por subgrupos de marcas, no *logit* aninhado marcas “mais próximas” são colocadas em um mesmo ninho pelo analista, permitindo que haja correlação de preferências por subgrupos de marcas. A vantagem do *logit* aninhado, portanto, é que quando o preço de determinada marca sobe, é possível que boa parte da substituição ocorra para outras marcas do mesmo ninho (a depender de um parâmetro que é estimado, indexando tanto a proximidade entre marcas de um mesmo ninho, quanto a distância entre as marcas desse e de outros ninhos). Assim, a fusão entre marcas de um mesmo ninho poderá ter um impacto muito maior em preços do que a fusão entre marcas de ninhos distintos. As desvantagens desse modelo, como já discutimos, são a arbitrariedade na definição de ninhos (contornável, em princípio, via teste de Vuong), a persistência da IIA intra-ramos, e o fato de elasticidades próprias exibirem os mesmos problemas das do modelo *logit* clássico.

- Generalização do *logit* 2 — Sistema de demanda PD-GEV (BST): Outra forma de adicionar flexibilidade ao modelo *logit* é especificar dimensões de diferenciação. Marcas com características parecidas estão mais próximas nesse espaço de diferenciação multidimensional. Assim como no *logit* aninhado, onde a proximidade entre marcas de um ninho é estimada, no BST a importância de cada dimensão de diferenciação em explicar as escolhas de consumidores também é estimada (através do parâmetro ρ). Assim, o efeito sobre preços será maior em uma fusão envolvendo marcas com classificações próximas. Um aspecto positivo desse sistema em comparação com o *logit* aninhado, é que ele permite tratar potenciais segmentações de mercado diferentes de forma simétrica, ao passo que o *logit* aninhado impõe uma estrutura hierárquica. Isso torna o modelo menos passível de críticas em aplicações práticas. No entanto, não surpreende que o esforço de implementação ante o *logit* aninhado seja maior.

- Métrica de distância: Sua principal vantagem é a flexibilidade. No entanto, tal modelo tem um componente de arbitrariedade na definição de vizinhança,

além de haver críticas quanto a sua aplicação para a análise de bem-estar [ver Nevo (2004)].

- Generalização do *logit* 3 — Sistema de demanda *logit* com coeficientes aleatórios (*mixed logit*, BLP): Trata-se da versão mais geral e flexível do *logit*, que se acredita fornecer elasticidades mais condizentes com a intuição econômica. Pode ser aplicado tanto a dados agregados, como a microdados, mas seu custo de implementação e o tempo gasto no processo de estimação, tendem a ser bastante altos.

1.5.6 Testes de especificação

Uma prática crescente e sã na literatura empírica em organização industrial é aquela em que o analista reporta o resultado de testes de especificação. Em vez de oferecermos um tratamento compreensivo desse teste de especificação, remetemos o leitor ao livro-texto de Wooldridge (2002).⁹⁹ Visto que o contrário foge ao escopo do presente trabalho, nos limitamos aqui a um breve comentário do objetivo de alguns desses testes:

- Testes de sobreidentificação [ver Wooldridge (2002, subseções 6.2.2 e 8.5.2)]: A exogeneidade das variáveis instrumentais quando o número destas é igual ao número de variáveis endógenas (ou seja, a equação está *exatamente* identificada) é por definição uma hipótese mantida, ou seja, não-testável. No entanto, quando a equação estiver sobreidentificada, dado um número de instrumentos superior ao número de variáveis endógenas, é possível que a redução da variância das estimativas, obtida através do aumento no número de condições de momento (cada instrumento representa uma condição adicional que tende a reduzir a variância das estimativas), esteja sendo obtida à custa de consistência. Este é mais um exemplo do *trade-off* recorrente em econometria entre eficiência e consistência dos estimadores. Daí decorre a importância de que se teste a sobreidentificação.¹⁰⁰

- Testes de endogeneidade [ver Wooldridge (2002, subseção 6.2.1)]: Esses testes visam avaliar a correlação entre a variável supostamente endógena e o resíduo da estimação. Ignorar a endogeneidade dos preços pode gerar resultados tais como elasticidades-preço próprias de demanda inesperadamente baixas (demandas inelásticas) e, no limite, coeficientes associados ao preço positivos [Trajtenberg (1989)]. A endogeneidade dos preços pode ser modelada com o uso de técnicas econométricas clássicas, como 2SLS — ver a seção sobre identificação de demanda.

99. Um tratamento compreensivo, incluindo comandos do pacote econométrico Stata, é oferecido por Baum, Schaffer e Stillman (2003).

100. Davidson e MacKinnon (1993) comentam que “testes de sobreidentificação deveriam ser realizados rotineiramente quando se computam estimativas por variáveis instrumentais” (p. 236; tradução livre).

A instrumentação dos preços tende a gerar elasticidades com magnitudes mais próximas do que se esperaria em um dado mercado. Adicionalmente, espera-se que o uso de variáveis demográficas — e o controle de heterogeneidade em geral, com o uso de efeitos fixos — traga melhores resultados.

- Heterocedasticidade e autocorrelação dos resíduos [ver Wooldridge (2002, seção 7.8.5)]: A análise dos resíduos estimados poderá informar sobre a validade das premissas adotadas pelo analista quanto à distribuição e correlação temporal nos erros.

1.5.7 Análise de sensibilidade e comparação entre modelos

Dada a complexidade de muitos dos modelos de demanda considerados, nem sempre a análise das implicações das hipóteses feitas é imediata, sejam elas feitas sobre o modelo estrutural, ou de cunho estocástico. Assim, uma forma muito adotada na literatura para checar a razoabilidade dessas hipóteses e/ou comparar modelos alternativos é por meio de análises de sensibilidade e robustez. Casos típicos incluiriam, como já descrito, verificar se os valores das elasticidades cruzadas para especificações lineares e logarítmicas em preço se equivalem. Mais geralmente, é importante checar quão robustos são os resultados a definições de variáveis como o *market share* do bem exterior, forma funcional — especialmente a forma como variáveis demográficas interagem com as características, hipóteses distribucionais, e assim por diante. Além disso, convém confrontar os resultados com a intuição econômica do pesquisador.

Modelos simétricos (tais como o *logit*) são os menos indicados em termos de estimação de demanda e análise de fusões, principalmente por não levarem em conta uma dimensão fundamental no que se refere a produtos diferenciados — a proximidade entre alternativas no espaço de características, com conseqüências indesejáveis nas elasticidades cruzadas e, conseqüentemente, na definição de mercado relevante. Apesar disso, modelos *logit* e derivados têm sido usados com frequência [Werden e Froeb (1994)]. Entretanto, exceto quando tempo for a *única* variável relevante, não recomendamos o seu uso e, ainda assim, que sejam feitos testes de especificação.

Comparar modelos contínuos multiestágios e modelos discretos com coeficientes aleatórios é algo bem mais complexo. Modelos multiestágio (e mesmo discretos, como o *logit* aninhado) requerem a definição *a priori* dos segmentos em cada estágio, que não podem ser nem muito pequenos, nem muito grandes, e são freqüentemente difíceis de definir — uma saída seria definir especificações alternativas e verificar a robustez dos resultados para especificações diferentes.

Muitas vezes é de interesse comparar, por exemplo, modelos alternativos apresentados pelas partes envolvidas em um ato de concentração. Fazemos aqui alguns comentários sobre como comparar modelos alternativos, que devem ser usados em conjunto com as considerações feitas nas seções anteriores:

- Comparando MNL e NL: No que se refere a testes de especificação para o modelo *logit*, Hausman e McFadden (1984) propõem dois tipos de testes, sendo o primeiro uma aplicação do “Teste de Hausman” [Hausman (1978a)], em que o MNL é o modelo eficiente sob a hipótese nula e inconsistente sob a alternativa, e o NL o modelo consistente sob a hipótese alternativa; e o segundo obtido através da estimação de um modelo NL, do qual o MNL é um caso particular, e da aplicação da “santíssima trindade” de testes (Razão de Verossimilhança, Wald, Multiplicadores de Lagrange). O segundo teste decorre de um resultado demonstrado por McFadden (1978), de acordo com o qual a estrutura de estágios é consistente com RUM se, e somente se, determinados coeficientes (correspondentes a σ na equação (28)) estão contidos no intervalo (0,1) — caso eles tendam a 1, o modelo se reduz a um MNL e, caso eles sejam maiores do que 1, há substituição entre ramos diferentes. Isso implica que a especificação dos estágios estimada não é consistente com a hipótese RUM, e uma configuração alternativa da árvore de decisão deve necessariamente ser considerada. Trajtenberg (1989) compara MNL e NL em seu estudo de tomógrafos.

- Especificações alternativas em NL e escolha em estágios múltiplos (*multi-stage budgeting*): A divisão de produtos em grupos tanto em modelos NL, como no caso de modelos de escolha em estágios, é um tanto arbitrária.¹⁰¹ Adicionalmente, no caso do NL, a hipótese de choques i.i.d. intra-ramo, é passível de críticas. Além do problema da arbitrariedade da escolha dos elementos constituintes de cada estágio, deve-se atentar para o tamanho dos ramos de um dado estágio. De forma intuitiva, espera-se que algumas das características de produtos em um mesmo mercado estejam correlacionadas — por exemplo, potência, número de cilindros e dimensões do carro, tal como no caso dos automóveis, em Goldberg (1995). Adotando-se um ramo com poucos produtos, eles tenderão a ser mais homogêneos, gerando um alto grau de colinearidade, mas eliminando a necessidade de incluir todas as características dos produtos. No caso do NL, incluir ramos com muitos produtos tende a mitigar a colinearidade, mas volta-se ao problema da IIA presente no MNL. Além disso, cabe ressaltar que, apesar de capturar

101. Rubinfeld (2000) lembra que a substituição entre marcas pertencentes a segmentos (ninhos) distintos em um sistema de escolha em estágios múltiplos opera através da alocação de gastos entre os segmentos, em um estágio anterior, em decorrência da variação no índice de preços de cada segmento. Isso gera um viés para baixo na elasticidade-preço cruzada entre essas marcas pertencentes a distintos segmentos.

correlações entre elementos de um mesmo ramo, o NL não consegue capturar correlações entre ramos distintos, de forma que, quando alternativas não podem ser separadas em diferentes ramos de modo satisfatório, o NL não é aplicável. Para ambos os modelos, sugerimos que, primeiro, especificações alternativas sejam testadas, a fim de comprovar a robustez dos resultados obtidos. A seguir, e apesar de não termos conhecimento de sua implementação empírica, recomendamos o teste de Vuong (descrito no Apêndice) para comparar especificações alternativas de seqüências de “ninhos” para modelos NL e de escolha em estágios múltiplos.¹⁰² Tal teste seria incorporado quando, por exemplo, duas estruturas de ninhos passassem pelos testes de especificação do NL, sendo ambas consistentes com a hipótese RUM, e não tendo uma resposta sobre qual delas é preferível.

- Comparando MNL e BLP: Supondo-se que o econometrista tenha acesso ao método de estimação BLP, uma estratégia de teste óbvia é impor a hipótese nula de coeficientes não-aleatórios (ou seja, $\beta^o = \beta^u = 0$) e comparar modelos BLP e MNL por meio do teste de razão de verossimilhança [ver Wooldridge (2002), ou Newey e McFadden (1994)]. Nesse caso, se a diferença entre as funções de verossimilhança não for substancial, rejeitaria o modelo BLP em favor do MNL. No caso de não se ter acesso ao método de estimação BLP, é teoricamente possível (mesmo que de implementação não-trivial) construir um teste de multiplicadores de Lagrange e testar-se a hipótese nula de coeficientes não-aleatórios ($\beta^o = \beta^u = 0$) ou, equivalentemente, de um MNL. Tal como para a aplicação do teste de Vuong para a comparação de especificações NL ou multiestágios, não é de nosso conhecimento que tal teste tenha sido implementado.

1.5.8 Estimação simultânea de demanda e de oferta, ou estimação em estágios

O *trade-off* entre eficiência e consistência manifesta-se novamente na opção pelo analista de estimar simultaneamente os lados de demanda e de oferta, ou de proceder à estimação em estágios. A vantagem de estimar simultaneamente equações de demanda e de oferta (por exemplo, equações de precificação, além de possivelmente outras equações tais como as de demanda por fatores¹⁰³) é que o maior número de restrições (condições) na estimação deve resultar em menor variância das estimativas dos parâmetros estruturais. No entanto, caso haja erro de especificação no lado de oferta — o que tende a ser mais comum relativamente à

102. Capps, Church e Love (2003) sugerem o critério de dominância de verossimilhança de Pollak e Wales (1991), também um teste não-aninhado.

103. Para a inclusão de equações de demanda por fatores, ver, por exemplo, Gollop e Roberts (1979) e Appelbaum (1979), ou a discussão em Bresnahan (1989).

especificação de demanda — a estimação de oferta simultaneamente à estimação de demanda poderá “contaminar” esta última, levando a estimativas inconsistentes de demanda.

Por esse motivo, muitas vezes é preferível proceder à estimação em estágios: primeiro estima-se o lado de demanda, possivelmente utilizando variáveis de oferta na condição de instrumentos para lidar com o problema de endogeneidade. De posse dos parâmetros estimados de demanda, estes são embutidos na especificação de oferta que é então estimada, em um segundo estágio. Naturalmente, como os parâmetros de demanda a serem incorporados à especificação de oferta neste segundo estágio são estimativas procedentes da inferência no primeiro estágio, e não os verdadeiros valores da população, no segundo estágio deve-se calcular erros-padrão pelo método de *bootstrap*, ou utilizar a teoria de estimadores de dois estágios [ver *2-step estimators*, em Wooldridge (2002, cap. 12)].¹⁰⁴ Isso levará em consideração o erro de amostragem (*sampling error*) no primeiro estágio.

Slade (2004), por exemplo, cita que “uma vantagem do procedimento em duas etapas é que a especificação incorreta da CPO não irá contaminar as estimativas de demanda, nas quais tipicamente depositamos maior confiança” (p. 15; tradução livre).

1.5.9 Elementos dinâmicos

As especificações de demanda e de oferta apresentadas são estáticas: assume-se que não operem efeitos dinâmicos importantes na indústria tanto pelo lado da demanda quanto pelo lado da oferta, ou que estes não impactem a consistência das estimativas. É fácil conceber situações em que tal premissa pode falhar, tais como:

- Empilhamento de estoques pelo consumidor (*consumer stockpiling*): Quando o consumidor pode acumular estoques de algum produto não-durável (*non-durable*), mas armazenável (*storable*). Por exemplo, ao se deparar com um preço promocional de detergente em pó, o consumidor adquire estoques para várias semanas de consumo. Naturalmente, a elasticidade-preço de demanda de curto prazo, que leva em consideração a reação imediata das vendas à flutuação de preços de acordo com o calendário promocional, poderá ser maior do que a elasticidade-preço de demanda de longo prazo, que poderá ser relevante segundo a finalidade do analista (por exemplo, querendo estimar os efeitos sobre preços de um ato de concentração).¹⁰⁵ Remetemos o leitor a Hendel e Nevo (2001 e 2002).

104. Para uma discussão não-técnica do *bootstrap*, ver Brownstone e Valletta (2001).

105. Lembrar que, *caeteris paribus*, elasticidades-preço de demanda próprias mais elevadas limitam o aumento de preços após a fusão. Portanto, um viés para cima (em termos absolutos) na estimativa de elasticidade significa um viés para baixo no efeito estimado de preços de um ato de concentração.

- Formação de hábito (*habit formation*) ou efeitos de rede (*network effects*): Quando o consumo de amanhã depende sensivelmente do consumo de hoje, tais como no consumo de bens geradores de dependência (*addictive goods*, como narcóticos, tabaco e álcool) dizemos que ocorre formação de hábito. De forma semelhante, operam efeitos de rede em uma indústria quando externalidades estão presentes, tais como tipicamente é o caso em bens de informação, comunicação e entretenimento — a minha utilidade de consumo de *software* irá depender do número de outros consumidores, visto que quanto maior for o mercado, maior serão a variedade e a qualidade ofertadas de *software*; por outro lado, quanto maiores a variedade e a qualidade ofertadas maior será o número de usuários. Uma leitura sugerida é Pollak e Wales (1992), mas ver também Sutton (1998) para uma discussão interessante, acompanhada de exemplos, de efeitos de rede em ação.

- Aprendendo ao fazer (*learning by doing*): Em algumas indústrias, como a indústria química, custos são decrescentes na produção histórica acumulada. Quanto maior for a experiência acumulada de produção, maior é a produtividade. Este é outro exemplo de externalidade, matematicamente isomorfo ao de efeitos de rede, só que operando pelo lado de oferta e não de consumo [ver, novamente, Sutton (1998)].

- Bens duráveis: A demanda presente por alguns bens, como o de equipamentos industriais e agrícolas, irá depender de condições futuras, como a expectativa de renda ou o preço futuro de aço ou de soja. De maneira semelhante, a utilidade de um automóvel é usufruída não somente no período presente como também em períodos futuros [ver Hendel e Nevo (2004)].

Como capturar esses efeitos dinâmicos na estimação estrutural, identificando os parâmetros de forma satisfatória, não é uma questão trivial e é matéria de pesquisa atualmente, apesar de sua complexidade. Algumas vezes variáveis defasadas (em níveis e/ou em diferenças) são adicionadas à especificação, na tentativa de capturar esses efeitos. Entretanto deve-se atentar para o problema de identificação, assim como a possível multicolinearidade entre o valor presente da variável em questão e os valores defasados.

1.5.10 Relação entre elasticidades no varejo e elasticidades derivadas do produtor

É comum o analista estar interessado em avaliar, por exemplo, um ato de concentração entre *produtores* (por exemplo, os fabricantes de chocolate Nestlé e Garoto) mas dispor de dados coletados por agências de pesquisa de mercado (como a ACNielsen) apenas junto a *varejistas* (através, por exemplo, de *scanners*). Apesar

de ter sido tipicamente ignorado na literatura empírica e na prática antitruste, o relacionamento entre as elasticidades-preço de demanda estimadas com dados de varejo e os efeitos sobre preços de uma fusão entre produtores não é trivial.¹⁰⁶ As elasticidades-preço da demanda “final” que os varejistas como um todo observam (e que o analista estima) não são necessariamente as mesmas que as elasticidades enfrentadas pelo produtor, a não ser que certas premissas restritivas sejam satisfeitas (por exemplo, que os varejistas sejam perfeitamente competitivos, ou que os varejistas apliquem um *markup* percentual fixo sobre o preço de compra¹⁰⁷). Ver, por exemplo, Hausman e Leonard (2002), que abordam esse assunto, mas sobretudo Hosken *et alii* (2002), para uma discussão intuitiva.¹⁰⁸

Hosken *et alii* (2002) também discutem o importante aspecto de que várias vezes o contrato de fornecimento entre produtores de bens diferenciados e os varejistas de tais produtos envolve formatos de preço (termos comerciais) mais complicados do que meramente preços lineares em quantidades (ou seja, onde o preço para uma unidade é R\$ 5, e o preço para 100 unidades é ainda R\$ 5 por unidade). Por exemplo, é comum observar a prática de descontos por volume, ou bonificações em dinheiro caso um volume de vendas preestabelecido entre produtor e varejista seja atingido. Novamente, esse formato não-linear de preços pode ter implicações não-triviais para a análise dos efeitos de uma fusão entre produtores, utilizando-se dados de varejo (vendas e condições, normalmente lineares, do varejista para o consumidor).

Sugerimos ao leitor interessado nesses contratos verticais de fornecimento e formas de lidar com os problemas citados, os trabalhos empíricos de Villas-Boas (2001); Manuszak (2002); Asker (2004); Bonnet, Dubois e Simioni (2004); Brenkers e Verboven (2004); e Villas-Boas e Zhao (2005). Esses trabalhos fazem parte de uma literatura ainda incipiente que trata de analisar como dados de *scanner*, coletados ao nível de varejo, podem nos informar sobre a estrutura da indústria fornecedora que fica a montante na cadeia vertical de suprimentos.

106. Observe que os dados de varejo normalmente disponíveis ao analista estão agregados para todos os estabelecimentos (dentro de determinado canal, como supermercados com mais de dez *checkouts*) em uma mesma região geográfica. Caso o analista estivesse trabalhando com dados desagregados por estabelecimento, ele teria de lidar com a substituição entre estabelecimentos, questão que não será necessariamente relevante na análise de ato de concentração entre produtores.

107. Hosken *et alii* (2002) lembram que tal comportamento por parte de um varejista será otimizador somente sob condições muito restritivas (por exemplo, varejistas vendendo somente um produto, enfrentando demanda de elasticidade constante e incorrendo em custos marginais 0 além do preço de custo).

108. Os autores especulam que “para alguns produtos onde a lealdade (do consumidor) à marca não seja particularmente forte, a ameaça por parte de varejistas de tirar esses produtos das prateleiras irá disciplinar os aumentos de preço de produtores. Nesses casos, as elasticidades ao nível de varejo poderão subdimensionar as elasticidades ao nível do produtor. Por outro lado, algumas marcas fortes não seriam descontinuadas por varejistas a não ser que o produtor praticasse um aumento substancial de preços. Ai então, as elasticidades ao nível de varejo poderão superdimensionar as elasticidades ao nível do produtor” (p. 26; tradução livre).

1.5.11 Calculando custos marginais diretamente

Em algumas situações é possível ao analista calcular o custo marginal diretamente, em vez de inferi-lo a partir da estimação estrutural. Este tipicamente poderá ser o caso quando a tecnologia de produção for simples (por exemplo, a tecnologia de coeficientes fixos) e houver disponibilidade de dados (por exemplo, quando o analista observa os coeficientes fixos bem como o preço de fatores). Exemplos são oferecidos por Genesove e Mullin (1998), que estudam a indústria de refinamento de açúcar (em que uma dada quantidade de açúcar bruto é convertida em açúcar refinado e os autores observam o preço de açúcar bruto importado), e Salvo (2004), que estuda a indústria de cimento (onde os coeficientes fixos e preços de fatores também são observados, mas crucialmente o autor observa os embarques de cimento de cada fábrica para cada mercado local, possibilitando o cálculo do custo de transporte). Apesar do considerável esforço exigido do analista, que terá de conhecer detalhes institucionais da indústria, o cálculo do custo marginal — quando possível — oferece a grande vantagem de substituir a estimação deste, ou mesmo possibilitar a verificação do custo marginal estimado.

Quando a tecnologia for mais complexa, pode ser possível que o analista calcule, a partir de algumas premissas, um limite superior do custo marginal. Essa informação também poderá ser utilizada pelo analista para conferir as estimativas econométricas de custo marginal, ou para aumentar a eficiência da estimação (ao incorporar as restrições decorrentes do limite superior na estimação).

1.6 RECOMENDAÇÕES DE POLÍTICA

Apesar de o presente trabalho ter trazido à tona uma variedade de aspectos técnicos da fronteira da literatura, os aspectos institucionais envolvidos no antitruste merecem tanta ou mais atenção do que a técnica. Assim, nossas principais recomendações são no sentido do fortalecimento institucional do SBDC.

Em primeiro lugar, recomendamos o fortalecimento de mecanismos de obtenção de dados de qualidade. Além disso, os dados requeridos e utilizados em estudos feitos por todas as partes em um ato de concentração, ou na instauração de um processo administrativo, devem ser depositados junto ao SBDC, permitindo, assim, a replicação dos resultados obtidos.

Em segundo lugar, recomendamos que análises feitas pelo SBDC sejam feitas antes da concretização do ato. Mesmo uma análise tecnicamente correta e bem fundamentada pode sofrer contestações por implicar a reversão de uma operação, contribuindo para o desgaste da imagem do SBDC diante do grande público, assim como contestações na justiça da parte perdedora.

Em terceiro lugar, recomendamos a valorização da carreira de analista da concorrência. Entendemos que um SBDC não pode prescindir de quadros bem formados e em processo de constante aperfeiçoamento: o material coberto neste trabalho mostra a importância do conhecimento de Organização Industrial e Econometria uma vez que se tenha optado pela utilização de modelos estruturais. Em particular, vemos cursos de extensão e pós-graduação, bem como intercâmbios com SDCs estrangeiros, como fontes sadias de aperfeiçoamento de pessoal.

Do ponto de vista da adoção de modelos descritos no trabalho, entendemos que, num primeiro estágio, o SBDC veja como mais conveniente adotar métodos parcimoniosos como o PCAIDS (utilizado pelo SDC da Nova Zelândia) e o ALM (utilizado pelo SDC americano). No entanto, sob os critérios de *precisão de previsão, tempestividade, parcimônia, operacionalidade e simplicidade de interpretação*, entendemos que, num horizonte de longo prazo, métodos mais adequados seriam, por exemplo, o NL e o sistema de escolha em estágios, desde que adotados com as precauções detalhadas no texto. Tal recomendação encontra-se alinhada com o seguinte comentário de Hosken *et alii* (2002, p. 26), em texto para discussão da Federal Trade Commission: “Many of the demand system estimations presented to the Agencies use the multi-stage budgeting approach with the AIDS demand specification.”

Como mencionamos no texto, os métodos sugeridos encontram-se implementados em pacotes econométricos disponíveis no mercado, não exigindo a programação de rotinas complexas por parte do analista, como no caso de métodos como o modelo logit com coeficientes aleatórios: estimar um modelo de tal complexidade exige uma série de escolhas feitas caso a caso, teste de especificações alternativas, além da adequação aos dados disponíveis para o estudo, requerendo conhecimento e maturidade que seriam obtidos, por exemplo, ao longo de um curso de Doutorado em Economia.¹⁰⁹

109. Isso não significa que modelos dessa natureza não possam ser — ou não tenham sido — implementados no âmbito da análise antitruste brasileira. É de nosso conhecimento, por exemplo, que modelos dessa complexidade têm sido estimados pela Secretaria de Direito Econômico em análise recente de ato de concentração no mercado brasileiro de tratores.

BIBLIOGRAFIA

- AKAIKE, H. Information theory and an extension of the likelihood ratio principle. In: PETROV, B. N., CSAKI, F. (eds.). *Second International Symposium of Information Theory*. Minnesota Studies in the Philosophy of Science, Budapest: Akademiai Kiado, 1973.
- ALSTON, J. M., FOSTER, K. A., GREEN, R. D. Estimating elasticities with the linear approximate almost ideal demand system: some Monte Carlo results. *Review of Economics and Statistics*, v. 76, p. 351-356, 1994.
- APPELBAUM, E. Testing price taking behaviour. *Journal of Econometrics*, v. 9, p. 283-294, 1979.
- ANDERSON, S. P., DE PALMA, A., THISSE, J.-F. *Discrete choice theory of product differentiation*. The MIT Press, 1992.
- ASKER, J. W. *Diagnosing foreclosure due to exclusive dealing*. NYU Stern, 2004, mimeo.
- BAJARI, P., BENKARD, C. L. *Demand estimation with heterogeneous consumers and unobserved product characteristics: a hedonic approach*. Stanford University, 2004, mimeo.
- BAKER, J., BRESNAHAN, T. F. The gains from Merger or collusion in product-differentiated industries. *Journal of Industrial Economics*, v. 33, n. 4, p. 427-444, 1985.
- BARTEN, A. P. *Theorie en empirie van een Volledig Stelsel van Vraagvergelijkingen*. Rotterdam: University of Rotterdam, 1966 (Doctoral Dissertation).
- BAUM, C. F., SCHAFFER, M. E., STILLMAN, S. *Instrumental variables and GMM: estimation and testing*. Department of Economics, Boston College, 2003 (Working Paper, 545).
- BEN-AKIVA, M. *Structure of passenger travel demand models*. MIT, 1973 (Ph.D. Dissertation).
- BERNDT, E. R., FUSS, M. A., WAVERMAN, L. *Dynamic models of the industrial demand for energy*. Electric Power Institute, EA-580, Palo Alto, CA, 1977.
- BERNHEIM, B. D., WHINSTON, M. D. Multimarket contact and collusive behavior. *Rand Journal of Economics*, v. 21, p. 1-26, 1990.
- BERRY, S. Estimating discrete-choice models of product differentiation. *Rand Journal of Economics*, v. 25, p. 242-262, 1994.
- BERRY, S., LEVINSOHN, J., PAKES, A. Automobile prices in market equilibrium. *Econometrica*, v. 63, p. 841-890, 1995.
- . Differentiated products demands systems from a combination of micro and macro data: the new car market. *Journal of Political Economy*, v. 112, n. 1, p. 68-105, 2004.
- BERRY, S., LINTON, O., PAKES, A. Limit theorems for estimating the parameters of differentiated product demand systems. *Review of Economic Studies*, 2004 (forthcoming).
- BERRY, S., PAKES, A. Some applications and limitations of recent advances in empirical industrial organization: Merger analysis. *American Economic Review*, v. 83, n. 2, p. 247-252, 1993.
- . *The pure characteristics demand model*. Harvard University, 2002, mimeo.
- BONNET, C., DUBOIS, P., SIMIONI, M. *Two-part tariffs versus linear pricing between manufacturers and retailers: empirical tests on differentiated products markets*. Toulouse, 2004, mimeo.
- BRENKERS, R., VERBOVEN, F. *Liberalizing a distribution system: the European car market*. K. U. Leuven, 2004, mimeo.

- BRESNAHAN, T. F. The oligopoly solution concept is identified. *Economic Letters*, v. 10, p. 87-92, 1982.
- . Competition and collusion in the American automobile oligopoly: the 1955 price war. *Journal of Industrial Economics*, v. 35, p. 457-482, 1987.
- . Empirical studies of industries with market power. In: SCHMALENSEE, R., WILLIG, R. D. (eds.). *Handbook of Industrial Organization*. New York: North-Holland, v. II, p. 1.011-1.057, 1989.
- . *The apple cinamon cheerios war: valuing new goods, identifying market power, and economic measurement*. Stanford University, 1997, mimeo.
- BRESNAHAN, T. F., STERN, S., TRAJTENBERG, M. Market segmentation and the sources of rents from innovation: personal computers in the late 1980s. *Rand Journal of Economics*, Special Issue, 1997.
- BROWNSTONE, D., VALLETTA, R. The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. *Journal of Economic Perspectives*, v. 15, p. 129-141, 2001.
- BUSE, A., CHAN, W. H. Invariance, price indices and estimation in almost ideal demand systems. *Empirical Economics*, v. 25, p. 519-539, 2000.
- CAPPS Jr., O., CHURCH, J., LOVE, H. A. Specification issues and confidence intervals in unilateral price effects analysis. *Journal of Econometrics*, v. 113, p. 3-31, 2003.
- CARDELL, N. S. *Extensions of the multinomial logit: the bedonic demand model, the non-independent logit model, and the ranked logit model*. Harvard University, 1989 (Ph.D. Dissertation).
- . Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, v. 13, p. 185-213, 1997.
- CARLTON, D. W. Using economics to improve antitrust policy. Milton Handler Lecture, *Columbia Business Law Review*, 2003 (forthcoming).
- CHAMBERLAIN, G. Multivariate regression models for panel data. *Journal of Econometrics*, v. 18, n. 1, p. 5-46, 1982.
- CHRISTENSEN, L. R., JORGENSON, D. W., LAU, L. J. Transcendental logarithmic utility functions. *American Economic Review*, v. 65, p. 367-383, 1975.
- CORTS, K. S. Conduct parameters and the measurement of market power. *Journal of Econometrics*, v. 88, p. 227-250, 1999.
- CYSNE, R. P. *et alii*. Demanda por cerveja no Brasil: um estudo econométrico. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 31, n. 2, p. 249-268, 2001.
- DAVIS, D. D., WILSON, B. J. *Differentiated product competition and the antitrust logit model: an experimental analysis*. Virginia Commonwealth University, 2003, mimeo.
- DAVIS, P., HUSE, C., VAN REENEN, J. *Evaluating incentives for collusion in differentiated products markets: measuring the coordinated effects of Mergers*. 2005 (in preparation).
- DAVIS, P. J. *Demand models for market level data*. MIT Sloan School of Management, 2001, mimeo.
- . *Lecture notes in economics of industry*. Department of Economics, London School of Economics, 2004.

- DEATON, A., MUELLBAUER, J. An almost ideal demand system. *American Economic Review*, v. 70, p. 312-326, 1980a.
- . *Economics and consumer behavior*. Cambridge University Press, 1980b.
- DIXIT, A., STIGLITZ, J. E. Monopolistic competition and optimum product diversity. *American Economic Review*, v. 67, p. 297-308, 1977.
- DUBÉ, J.-P. *Product differentiation and Mergers in the carbonated soft drink industry*. University of Chicago, 2004, mimeo.
- ELLISON, S. F. *et alii*. Characteristics of demand for pharmaceutical products: an examination of four cephalosporins. *Rand Journal of Economics*, v. 28, n. 3, 1997.
- ENCAOUA, D., JACQUEMIN, A. Degree of monopoly, indices of concentration and threat of entry. *International Economic Review*, v. 21, p. 87-105, 1980.
- EPSTEIN, R. J., RUBINFELD, D. L. Merger simulation: a simplified approach with new applications. *Antitrust Law Journal*, v. 69, p. 883-919, 2002.
- FARRELL, J., SHAPIRO, C. Horizontal Mergers: an equilibrium analysis. *American Economic Review*, v. 80, n. 1, p. 107-126, 1990.
- FIUZA, E. P. S. *Automobile demand and supply in Brazil: effects of tax rebates and trade liberalization on markups in the 1990s*. 2002, mimeo.
- FREEDMAN, D. A., PETERS, S. Bootstrapping a regression equation: some empirical results. *Journal of the American Statistical Association*, v. 79, p. 97-106, 1984.
- GASMI, F., LAFFONT, J. J., VUONG, Q. H. A structural approach to empirical analysis of collusive behavior. *European Economic Review*, v. 34, p. 513-523, 1990.
- . Econometric analysis of collusive behaviour in a softdrink market. *Journal of Economics and Management Strategy*, v. 1, p. 277-311, 1992.
- GASTRI, F., VUONG, Q. An econometric analysis of some duopolistic games in frices an advertising. In: RHODES, R. FOMBY, T. *Advances in econometrics: econometric methods and models for industrial organization*. JAI pres Inc., p. 225-254, 1991.
- GENESOVE, D., MULLIN, W. P. Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *Rand Journal of Economics*, v. 29, p. 355-377, 1998.
- GOLDBERG, P. K. Product differentiation and oligopoly in international markets: the case of the US automobile industry. *Econometrica*, v. 63, n. 4, p. 891-951, 1995.
- GOLLOP, F., ROBERTS, M. Firm interdependence in oligopolistic markets. *Journal of Econometrics*, v. 10, p. 313-331, 1979.
- GORMAN, W. M. A possible procedure for analysing quality differentials in the egg market. *Review of Economic Studies*, v. 47, p. 843-856, 1955, 1980.
- . Separable utility and aggregation. *Econometrica*, v. 27, p. 469-481, 1959.
- . The structure of utility functions. *Review of Economic Studies*, v. 5, p. 369-390, 1968.
- . Separability and aggregation. In: BLACKORBY, C., SHORROCKS, A. F. (eds.). *Collected works of W. M. Gorman*, 1995.

- GREEN, R., ALSTON, J. Elasticities in AIDS models. *American Journal of Agricultural Economics*, v. 72, p. 442-445, 1990.
- HAUSMAN, J. A. Specifications tests in econometrics. *Econometrica*, v. 46, p. 1.251-1.272, 1978.
- . Specification and estimation of simultaneous equations models. In: GRILICHES, Z., INTILIGATOR, M. (eds.). *Handbook of econometrics*. Amsterdam: North-Holland, 1983.
- . Valuation of new goods under perfect and imperfect competition. In: BRESNAHAN, T. F., GORDON, R. (eds.). *The economics of new goods. Studies in Income and Wealth*, v. 58. Chicago: National Bureau of Economic Research, 1997a.
- . *Reply to prof. Bresnahan*. MIT, 1997b, mimeo.
- HAUSMAN, J. A., LEONARD, G. K. Economic analysis of differentiated products mergers using real world data. In: HAUSMAN, J. A., LEONARD, G. K., MASON, G. *Law Review*, v. 5, n. 3, 1997.
- . The competitive effects of a new product introduction: a case study. *Journal of Industrial Economics*, v. L, n. 3, p. 235-261, 2002.
- HAUSMAN, J. A., LEONARD, G. K., ZONA, J. D. Competitive analysis with differentiated products. *Annales D'Economie et de Statistique*, v. 34, p. 159-180, 1994.
- HAUSMAN, J. A., MCFADDEN, D. Specification tests for the multinomial logit model. *Econometrica*, v. 52, n. 5, p. 1.219-1.240, 1984.
- HENDEL, I., NEVO, A. *Sales and consumer inventory*. NBER, 2001 (Working Paper, 9.048).
- . *Measuring the implications of sales and consumer stockpiling behavior*. Berkeley: University of California, 2002, mimeo.
- . Intertemporal substitution and storable products. *Journal of the European Economic Association*, v. 2, n. 2, p. 536-547, 2004.
- HICKS, J. R. *Value and capital*. Oxford University Press, 1936.
- HORIZONTAL MERGER GUIDELINES*. US Department of Justice and the Federal Trade Commission, 1992.
- HOSKEN, D. *et alii*. *Demand system estimation and its application to horizontal Merger analysis*. Federal Trade Commission, 2002, mimeo.
- IVALDI, M. *et alii*. *The economics of unilateral effects*. European Commission, Interim Report for DG Competition, 2003a.
- . *The economics of tacit collusion*. European Commission, Interim Report for DG Competition, 2003b.
- IVALDI, M., VERBOVEN, F. Quantifying the effects from horizontal Mergers in European competition policy. *International Journal of Industrial Organization*, 2004 (forthcoming).
- IWATA, G. Measurement of conjectural variations in oligopoly. *Econometrica*, v. 42, p. 947-966, 1974.
- KULLBACK, S., LEIBLER, R. On information and sufficiency. *Annals of Mathematical Statistics*, v. 22, n. 1, p. 79-86, 1951.

- KÜHN, K.-U. *The coordinated effects of Mergers in differentiated products markets*. University of Michigan, 2004, mimeo.
- LANCASTER, K. J. A new approach to consumer theory. *Journal of Political Economy*, v. 74, n. 2, p. 132-157, 1966.
- LEE, M.-J. *Methods of moments and semiparametric econometrics for limited dependent variable models*. Berlin: Springer-Verlag, 1966.
- LEONTIEF, W. Composite commodities and the problem of index numbers. *Econometrica*, v. 4, p. 39-59, 1936.
- MANSKI, C. F. The structure of random utility models. *Theory and Decision*, v. 8, p. 229-254, 1977.
- MANUSZAK, M. *The impact of upstream Mergers on retail gasoline markets*. Carnegie Mellon University, 2002, mimeo.
- MARIUZZO, F., WALSH, P. P., WHELAN, C. *Embedding consumer taste for location convenience into a structural model of equilibrium*. Dublin: Trinity College, 2004, mimeo.
- MCFADDEN, D. Conditional logit analysis of qualitative choice behavior. In: ZAREMBKA, P. (ed.). *Frontiers of econometrics*. New York: Academic Press, 1974.
- . Modeling the choice of residential location. In: KARLGVIST, A. *et alii* (eds.). *Spatial interaction theory and planning models*. Amsterdam: North-Holland, 1978a.
- . The general linear profit function. In: FUSS, M. A., MCFADDEN, D. (eds.). *Production economics: a dual approach to theory and applications*, Amsterdam: North-Holland, v. 1, 1978b.
- . Econometric models of probabilistic choice. In: MANSKI, C., MCFADDEN, D. (eds.). *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press, 1981.
- . Econometric analysis of qualitative response models. In: GRILICHES, Z., INTILLIGATOR, M. (eds.). *Handbook of econometrics*, Amsterdam: North-Holland, v. III, 1984.
- . A method of simulated moments for estimation of discrete response models without numerical Integration. *Econometrica*, v. 57, n. 5, p. 995-1.026, 1989.
- MCFADDEN, D., TRAIN, K. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, v. 15, p. 447-470, 2000.
- MOTTA, M. *Competition policy: theory and practice*. Cambridge University Press, 2004.
- NEVO, A. Identification of the oligopoly solution concept in a differentiated-products industry. *Economic Letters*, v. 59, p. 391-395, 1998.
- . Mergers with differentiated products: the case of the ready-to-eat cereal industry. *Rand Journal of Economics*, v. 31, p. 395-421, 2000.
- . Measuring market power in the ready-to-eat cereal industry. *Econometrica*, v. 69, p. 307-342, 2001.
- . *Lecture notes in industrial organization*. Cemmap Masterclass, 2004.

- NEWKEY, W. K., MCFADDEN, D. Large sample estimation and hypothesis testing. In: ENGLE, R. F., MCFADDEN, D. (eds.). *Handbook of econometrics*, Amsterdam: North-Holland, v. IV, 1994.
- NEWKEY, W. K., WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, v. 55, p. 703-708, 1987.
- PAKES, A. *Lecture notes in graduate industrial organization*. Department of Economics, Harvard University, 2003.
- PAKES, A., MCGUIRE, P. Computation of Markov perfect equilibria: numerical implications of a dynamic differentiated product model. *Rand Journal of Economics*, v. 25, n. 4, p. 555-589, 1994.
- PAKES, A., POLLARD, D. Simulation and the asymptotics of optimization estimators. *Econometrica*, v. 57, n. 5, p. 1.027-1.057, 1989.
- PARKER, P. M., RÖLLER, L. H. Collusive conduct in duopolies: multimarket contact and cross-ownership in the mobile telephone industry. *Rand Journal of Economics*, v. 28, p. 304-322, 1997.
- PASHARDES, P. Bias in estimating the almost ideal demand system with the stone index approximation. *Economic Journal*, v. 103, p. 908-915, 1993.
- PETERS, C. *Evaluating the performance of Merger simulation: evidence from the US airline industry*. Northwestern University, 2001, mimeo.
- PETRIN, A. Quantifying the benefits of new products: the case of the Minivan. *Journal of Political Economy*, v. 110, p. 705-729, 2002.
- PINKSE, J., SLADE, M. E. Mergers, brand competition, and the price of a pint. *European Economic Review*, v. 48, n. 3, p. 617-643, 2004.
- PINKSE, J., SLADE, M. E., BRETT, C. Spatial price competition: a semi-parametric approach. *Econometrica*, v. 70, p. 1.111-1.155, 2002.
- POLLAK, R. A., WALES, T. J. The likelihood dominance criterion: a new approach to model selection. *Journal of Econometrics*, v. 47, p. 227-242, 1991.
- . *Demand system specification and estimation*. Oxford: Oxford University Press, 1992.
- PORTER, R. H. A study of cartel stability: the joint executive committee, 1880-1886. *The Bell Journal of Economics*, v. 14, n. 2, p. 301-314, Autumn 1983.
- REISS, P. C., WOLAK, F. Structural econometric modeling: rationales and examples from industrial organization. *Handbook of Econometrics*, v. 6, 2002 (manuscript of 2002).
- RIVERS, D., VUONG, Q. Model selection tests in nonlinear dynamic models. *The Econometrics Journal*, v. 5, n. 1, p. 1-39, 2002.
- ROBERTS, M. Testing oligopolistic behavior: an application of the variable profit function. *International Journal of Industrial Organization*, v. 2, p. 367-383, 1984.
- RUBINFELD, D. Market definition with differentiated products: the post/Nabisco cereal Merger. *Antitrust Law Journal*, v. 68, p. 163-182, 2000.
- SALVO, A. E. *Inferring conduct under the threat of entry: the case of the Brazilian cement industry*. London School of Economics, 2004 (Sticerd Working Paper, EI 38).

- . *Price competition, Mergers and structural estimation in oligopoly*. London School of Economics, 2005 (Ph.D. Dissertation).
- SCHMALENSEE, R. Entry deterrence in the ready-to-eat breakfast cereal industry. *Bell Journal of Economics*, v. 9, p. 305-327, 1978.
- SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v. 6, p. 461-464, 1978.
- SHAPIRO, C. Mergers with differentiated products. *Antitrust*, v. 10, n. 2, p. 23-30, 1996.
- SLADE, M. E. Market power and joint dominance in UK brewing. *Journal of Industrial Economics*, v. LII, p. 133-163, 2004.
- SPENCE, M. Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies*, v. 43, p. 217-235, 1976.
- STONE, J. Linear expenditure systems and demand analysis: an application to the pattern of British demand. *Economic Journal*, v. 64, p. 511-527, 1954.
- SUSLOW, V. Estimating monopoly behavior with competitive recycling: an application to Alcoa. *Rand Journal of Economics*, v. 17, p. 389-403, 1986.
- SUTTON, J. *Technology and market structure: theory and history*. Cambridge, MA: The MIT Press, 1998.
- THEIL, H. The information approach to demand analysis. *Econometrica*, v. 6, p. 375-380, 1965.
- TRAJTENBERG, M. The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy*, v. 97, n. 2, p. 444-479, 1989.
- VERBOVEN, F. The nested logit model and representative consumer theory. *Economics Letters*, v. 50, n. 1, p. 57-63, 1996a.
- . International price discrimination in the European car market. *Rand Journal of Economics*, v. 27, n. 2, p. 240-268, 1996b.
- VILLAS-BOAS, J. M., ZHAO, Y. Retailer, manufacturers, and individual consumers: modeling the supply side in the ketchup marketplace. *Journal of Marketing Research*, v. 42, p. 83-95, 2005.
- VILLAS-BOAS, S. B. *Vertical contracts between manufacturers and retailers: an empirical analysis*. University of California at Berkeley, 2001, mimeo.
- VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica*, v. 57, p. 307-333, 1989.
- WERDEN, G. J. The history of antitrust market delineation. *Marquette Law Review*, v. 76, p. 123-215, 1992a.
- . Four suggestions on market delineation. *Antitrust Bulletin*, p. 107-121, 1992b.
- . A robust test for consumer welfare enhancing Mergers among sellers of differentiated products. *Journal of Industrial Economics*, v. 44, n. 4, p. 409-413, 1996.
- . Simulating the effects of differentiated products Mergers: a practitioner's guide. *Proceedings of the NE-165 Conference*, 1997.
- WERDEN, G. J., FROEB, L. M. The effects of Mergers in differentiated products industries: logit demand and Merger policy. *Journal of Law, Economics, & Organization*, v. 194, p. 407-426, 1994.

- WERDEN, G. J., FROEB, L. M., TARDIFF, T. J. The use of the logit model in applied industrial organization. *International Journal of the Economics of Business*, v. 3, n. 1, p. 83-105, 1996.
- WERDEN, G. J., FROEB, L. M., TSCHANTZ, S. The logit model for simulating unilateral competitive effects. In: HARKRIDER, J. (ed.). *Econometrics in antitrust*. ABA Section on Antitrust Law, 1994 (forthcoming).
- WILLIG, R. D. Merger analysis, industrial organization theory, and the Merger guidelines. *Brookings Papers on Economic Activity, Microeconomics*, p. 281-332, 1991.
- WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press, 2002.

APÊNDICE I

DEMANDA

I.1 Dedução da função de demanda AIDS — expressão (15)

Seguimos Deaton e Muellbauer (1980a) ao deduzirmos a função de demanda AIDS no formato de *share* de valor. Consideremos uma classe de preferências representada por uma função de custos do seguinte tipo:

$$\log c(u, p) = (1 - u) \log \{a(p)\} + u \log \{b(p)\}$$

onde u denota utilidade, p denota preços, e $a(p)$ e $b(p)$ são funções homogêneas de primeiro grau em preços, definidas conforme segue.¹¹⁰ $\log \{a(p)\}$ é função quadrática dos logaritmos dos preços

$$\log \{a(p)\} = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j$$

e $\log \{b(p)\} - \log \{a(p)\}$ é dado por

$$\log \{b(p)\} - \log \{a(p)\} = \beta_0 \prod_k p_k^{\beta_k}$$

onde α_i , β_i e γ_{ij}^* são parâmetros, tal que $\sum_i \alpha_i = 1$ e $\sum_j \gamma_{kj}^* = \sum_k \gamma_{kj}^* = \sum_j \beta_j = 0$. (Essas restrições são requeridas para que a função de custos seja linearmente

110. Observe que $c(u, p) = [a(p)]^{1-u} [b(p)]^u$. Com algumas exceções (ver o Apêndice de Deaton e Muellbauer (1980), o valor de u cairá entre 0 e 1 tal que $a(p)$ e $b(p)$ podem ser interpretados como os custos de subsistência ($u = 0$) e saciedade (ou bliss, $u = 1$), respectivamente.

homogênea em preços.) A função de custos (ou melhor, o respectivo log) pode ser escrita como

$$\log c(u, p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j + u \beta_0 \prod_k p_k^{\beta_k} \quad (39)$$

Pelo Lema de Shephard, $\frac{\partial c(u, p)}{\partial p_i} = q_i(u, p)$, a demanda compensada pelo bem i , e portanto:

$$\frac{\partial \log c(u, p)}{\partial \log p_i} = \frac{p_i}{c(u, p)} \frac{\partial c(u, p)}{\partial p_i} = \frac{p_i q_i(u, p)}{c(u, p)} = s_i(u, p) \quad (40)$$

onde s_i denota a participação orçamentária (*share* de valor) do bem i . Essa participação orçamentária é então obtida derivando (39):

$$s_i(u, p) = \frac{\partial \log c(u, p)}{\partial \log p_i} = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i u \beta_0 \prod_k p_k^{\beta_k} \quad (41)$$

onde γ_{ij} denota a média aritmética de γ_{ij}^* e γ_{ji}^*

$$\gamma_{ij} = \frac{1}{2} (\gamma_{ij}^* + \gamma_{ji}^*) \quad (42)$$

e, portanto, $\gamma_{ij} = \gamma_{ji}$ (simetria). A maximização de utilidade implica que $c(u, p)$ seja igual ao dispêndio total Y . A função de custos definida por (39) pode então ser invertida, gerando:

$$\begin{aligned} u \beta_0 \prod_k p_k^{\beta_k} &= \log Y - \left(\alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j \right) \\ &= \log Y - \left(\alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j \right) \end{aligned} \quad (43)$$

onde se utiliza (42) no último passo.

Defina agora o índice de preços P dado pela expressão translog

$$\log P := \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj} \log p_k \log p_j \quad (44)$$

A participação orçamentária do bem i dada por (41) pode então ser expressa, usando (43) e (44), como função de preços e do dispêndio:

$$s_i(u, p) = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log \left(\frac{Y}{P} \right) \quad (45)$$

As restrições aplicadas aos parâmetros de (39) e a restrição de simetria implicam

$$\sum_i \alpha_i = 1 \quad \sum_i \gamma_{ij} = \sum_i \beta_i = 0 \quad (\text{"somatórios"})$$

$$\sum_j \gamma_{ij} = 0 \quad (\text{homogeneidade})$$

$$\gamma_{ij} = \gamma_{ji} \quad (\text{simetria})$$

Contanto que essas restrições sejam observadas, (45) caracteriza um sistema de funções de demanda cujo somatório corresponde ao dispêndio total ($\sum_i s_i = 1$), são homogêneos de grau 0 com relação a preços e dispêndio e satisfazem a propriedade de simetria de Slutsky. Variações nos preços relativos impactam demanda através dos parâmetros γ_{ij} : uma variação percentual unitária no preço do bem j tem um efeito γ_{ij} sobre a participação orçamentária do bem i , mantido constante o dispêndio real $\frac{Y}{P}$. Variações no dispêndio real impactam demanda através dos parâmetros β_i . Esses parâmetros somam 0 e são positivos para “bens de luxo” e negativos para “bens de necessidade”.

Na prática [ver a aplicação empírica de Deaton e Muellbauer (1980a), bem como a nota 18], o índice de preços P dado em (44) pode ser aproximado por um índice de preços de P^S [Stone (1954)]:

$$\log P^S := \sum_i s_i \log p_i \quad (46)$$

Convém lembrar que essa dedução refere-se à demanda do consumidor individual. A agregação das demandas individuais, gerando demanda agregada, baseia-se em extenso material teórico-econômico, ressaltando-se o trabalho de Gorman (1995).

1.2 Dedução da elasticidade-preço referente ao sistema de escolha em dois estágios — expressão (18)

Lembrando que $s_{int} = \frac{p_{int} q_{int}}{Y_{nt}}$, temos que

$$\log q_{int} = \log Y_{nt} - \log p_{int} + \log s_{int}$$

onde s_{int} é dado pela equação de nível inferior (15). Assim, as elasticidades-preço de demanda podem ser calculadas como (suprimindo os subscritos n e t , para simplificar a exposição, e lembrando que a função indicadora $1[i = j]$ retorna 1 se $i = j$, retornando 0 no caso contrário)

$$\eta_{ij} = \frac{\partial \log q_i}{\partial \log p_j} = \frac{\partial (\log Y - \log p_i + \log s_i)}{\partial \log p_j} = \frac{\partial \log Y}{\partial \log p_j} - 1[i = j] + \frac{\partial \log s_i}{\partial \log p_j} \quad (47)$$

onde, diante de (15),

$$\frac{\partial \log s_i}{\partial \log p_j} = \frac{1}{s_i} \frac{\partial s_i}{\partial \log p_j} = \frac{1}{s_i} \left(\beta_i \frac{\partial \log Y}{\partial \log p_j} - \beta_i \frac{\partial \log P}{\partial \log p_j} + \gamma_{ij} \right) \quad (48)$$

Usando (48) em (47), temos

$$\eta_{ij} = \frac{\partial \log Y}{\partial \log p_j} \left(1 + \frac{\beta_i}{s_i} \right) - 1 [i = j] + \frac{1}{s_i} \left(\gamma_{ij} - \beta_i \frac{\partial \log P}{\partial \log p_j} \right) \quad (49)$$

Recordando o índice de preços de Stone (16), temos que

$$\frac{\partial \log P}{\partial \log p_j} = \frac{\partial (\sum_i w_i \log p_i)}{\partial \log p_j} = w_j + \sum_i \frac{\partial w_i}{\partial \log p_j} \log p_i \simeq w_j \quad (50)$$

assumindo que $\frac{\partial w_i}{\partial \log p_j} \simeq 0$.¹¹¹ Quanto a $\frac{\partial \log Y}{\partial \log p_j}$, ainda que $Y = \sum_i p_i q_i$, podemos aproximar¹¹² Y pelo produto do consumo total $Q = \sum_i q_i$ e o índice de preços Stone P ; ou seja:

$$Y = \sum_i p_i q_i \simeq P \sum_i q_i = PQ$$

o que equivale a

$$\log Y = \log P + \log Q$$

onde $\log Q$ é dado pela equação de nível superior (17). Dessa forma,

$$\frac{\partial \log Y}{\partial \log p_j} = \frac{\partial \log P}{\partial \log p_j} + \frac{\partial \log Q}{\partial \log p_j} = \frac{\partial \log P}{\partial \log p_j} \left(1 + \frac{\partial \log Q}{\partial \log P} \right)$$

111. Essa aproximação segue, por exemplo, Hausman e Leonard (2002). Caso se utilize o índice de preços de Stone (46), $\sum_i s_{it} \log p_{it}$, em vez da versão *média* do índice de preços de Stone (16), $\sum_i w_i \log p_{it}$, temos que (suprimindo por conveniência o subscrito t)

$$\frac{\partial \log P}{\partial \log p_j} = s_j + \sum_i \frac{\partial s_i}{\partial \log p_j} \log p_i \quad \text{onde} \quad \frac{\partial s_i}{\partial \log p_j} = \gamma_{ij} + \beta_i \frac{\partial \log Y}{\partial \log p_j} - \beta_i (\alpha_j + \sum_k \gamma_{kj} \log p_k). \quad \text{Ver nota 18.}$$

112. Podemos evitar esta aproximação ao definirmos a variável Q na regressão (17) como sendo dada pela divisão da receita total da indústria Y pelo índice de preços de Stone P (ou seja, $Q = Y/P$), em vez de definirmos Q como o consumo total (ou seja, $Q = \sum_i q_i$) como fizemos na apresentação do estágio superior. Na prática, as estimativas devem ser próximas.

que pode ser reescrita, diante de (50) e (17), como

$$\frac{\partial \log Y}{\partial \log p_j} = w_j (1 + \gamma)$$

Usando esta equação assim como novamente a (50) em (49), temos finalmente que:

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

que é a expressão para as elasticidades dada no texto.¹¹³

APÊNDICE II

ECONOMETRIA

II.1 O método delta¹¹⁴

O método delta é uma forma de obter a distribuição de funções de um estimador. Seja $\{\hat{\theta}_N\}$ uma seqüência de estimadores de um vetor p -dimensional $\theta \in \Theta$, e suponha que $\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow^d N(0, V)$, onde V é uma matriz positiva definida de dimensão $p \times p$.

Considere agora uma função $f: \Theta \rightarrow \mathbb{R}^q$ continuamente diferenciável no espaço $\Theta \subset \mathbb{R}^p$, com $q \leq p$, e suponha que θ seja um ponto interior de Θ . Defina ainda $f_\theta := \partial f / \partial \theta$ como o gradiente de f , com dimensão $q \times p$. Então:

$$\sqrt{N} \left(f(\hat{\theta}_N) - f(\theta_0) \right) \rightarrow^d N \left(0, f_\theta(\theta_0) \cdot V \cdot f_\theta(\theta_0)' \right)$$

Demonstração informal: como θ_0 é ponto interior de Θ e $\hat{\theta}_N$ é um estimador consistente de θ_0 , tem-se que $\hat{\theta}_N$ é ponto interior de um subconjunto convexo

113. Essa equação é utilizada, por exemplo, por Hausman e Leonard (2002). Para uma discussão sobre diferentes expressões para as elasticidades utilizadas na literatura, ver Alston, Foster e Green (1994).

114. A discussão desta seção segue de perto Wooldridge (2002, cap. 12), cuja leitura recomendamos.

de Θ contendo θ_0 com probabilidade tendendo a 1. Assim, com probabilidade tendendo a 1, pode-se usar a expansão $f(\hat{\theta}_N) = f(\theta_0) + f_\theta(\theta^*) \cdot (\hat{\theta}_N - \theta_0)$ onde θ^* é um ponto interior de um segmento ligando $\hat{\theta}_N$ e θ_0 . Como θ^* está no segmento (estando, inclusive, mais próximo de θ_0 que $\hat{\theta}_N$, ele também vai convergir em probabilidade para θ_0 . Resulta daí que $f_\theta(\theta^*) \xrightarrow{p} f_\theta(\theta_0)$ pelo Teorema de Slutsky. Pode-se então escrever

$$\begin{aligned} \sqrt{N}(f(\hat{\theta}_N) - f(\theta_0)) &= f_\theta(\theta^*) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) \\ &= f_\theta(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) + [f_\theta(\theta^*) - f_\theta(\theta_0)] \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) \\ &= f_\theta(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) + o_p(1) \cdot O_p(1) \\ &= f_\theta(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0) + o_p(1) \end{aligned}$$

onde a segunda linha é obtida somando-se e subtraindo-se $f_\theta(\theta_0) \cdot \sqrt{N} \cdot (\hat{\theta}_N - \theta_0)$ ao lado direito da igualdade na primeira linha; a terceira linha é obtida da consistência de $f_\theta(\theta^*)$ (pois $f_\theta(\theta^*) \xrightarrow{p} f_\theta(\theta_0)$ equivale a $f_\theta(\theta^*) - f_\theta(\theta_0) = o_p(1)$), e do fato de $\sqrt{N} \cdot (\hat{\theta}_N - \theta_0)$ ser limitada em probabilidade ($O_p(1)$); a última linha vem do fato de o produto entre uma variável aleatória que converge para 0 e uma limitada em probabilidade convergir também para 0.

Conclui-se então que, multiplicando-se $f_\theta(\theta_0)$ a uma variável aleatória $N(0, V)$ obtém-se $N(0, f_\theta(\theta_0) \cdot V \cdot f_\theta(\theta_0)')$, e o resultado, $\sqrt{N}(f(\hat{\theta}_N) - f(\theta_0)) \rightarrow^d N(0, f_\theta(\theta_0) \cdot V \cdot f_\theta(\theta_0)')$. A estimação consistente da matriz de covariância é feita plugando-se $\hat{\theta}_N$ no lugar de θ_0 no gradiente $f_\theta(\cdot)$.

II.2 O teste de Vuong

Focaremos agora critérios de seleção de modelos não-aninhados, ou seja, modelos em que um não pode ser obtido como caso particular do outro. Apesar de possuir algumas inconveniências, tais como desempenho insatisfatório em pequenas amostras, utilizaremos o teste de Vuong (1989), não só por ser bem estabelecido, como por não ter competidores bem estabelecidos [ver ainda Rivers e Vuong, (2002), para generalizações do Teste de Vuong].

O teste de Vuong é baseado no critério de informação de Kullback-Leibler [KLIC - Kullback e Leibler (1951)], uma medida de “distância”,¹¹⁵ ou discrepância, entre dois modelos. O KLIC é definido por:

$$KLIC := E_0 [\ln h_0(Y_i | X_i)] - E_0 [\ln f(Y_i | X_i, \beta^*)]$$

onde:

$h_0(\cdot|\cdot)$ é a densidade condicional verdadeira de Y_i dado X_i , que é, na verdade, desconhecida;

E_0 é a esperança tomada de acordo com o modelo verdadeiro; e

β^* são os valores pseudo-verdadeiros de β , ou seja, as estimativas de β quando $f(Y_i|X_i)$ não é o modelo verdadeiro.

A partir do KLIC, define-se o “melhor” modelo como sendo aquele que minimiza a discrepância, pelo fato de estar tão próximo quanto possível do modelo verdadeiro. Tal modelo é obtido pela maximização de $E_0 [\ln f(Y_i | X_i, \beta^*)]$, tal como no critério de quasi-máxima verossimilhança (QML). Tal observação permite concluir que escolher entre modelos consiste em comparar os valores esperados de suas log-verossimilhanças, e é exatamente nisso que se baseia o teste de Vuong.

Para fixar idéias, considere dois modelos, $M_f = \{F_\beta, \beta \in B\}$ e $M_g = \{G_\gamma, \gamma \in \Gamma\}$, onde F e G são funções distribuição, caracterizando os processos geradores de dados dos dois modelos, com densidades f e g , respectivamente. Os modelos acima definidos são chamados (estritamente) não-aninhados quando $M_f \cap M_g = \emptyset$, que pode ser melhor entendido como refletindo o fato de não haver nenhuma densidade condicional $(Y_i|X_i)$ que seja elemento de M_f e M_g , para quase todo x .

Exemplo. Considere modelos de regressão linear em que se tem interesse de testar a distribuição dos erros. Se M_f contempla o caso de erros com distribuição Normal e M_g o caso de erros com distribuição logística, os modelos são não-aninhados. ■

Uma dificuldade óbvia no que concerne a testes de hipóteses decidindo entre M_f e M_g é que se for imposta a nula de acordo com a qual o modelo M_f é verdadeiro, sua rejeição não implicará nenhuma informação quanto à correção do modelo

115. O termo distância é referido entre aspas pelo fato de, a rigor, o KLIC não ser uma distância, por não valer $KLIC(f, g) = KLIC(g, f)$, propriedade fundamental a que uma distância deve satisfazer.

M_g . Naturalmente, então, vem a idéia de se considerar a nula como supondo a equivalência entre os modelos comparados, ou seja,

$$H_0 : E_0 \left[\ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right] = 0$$

De acordo com a hipótese nula, ambos os modelos estão igualmente próximos da especificação correta.¹¹⁶ É importante notar que agora existem duas hipóteses alternativas, dadas por:

$$H_f : E_0 \left[\ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right] > 0$$

em que o modelo M_f é preferido ao M_g , e

$$H_g : E_0 \left[\ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right] < 0$$

segundo a qual o modelo M_g é preferido ao M_f .

Apesar de a esperança acima ser desconhecida, Vuong (1989) demonstra que, sob condições de regularidade,

$$\frac{1}{N} LR_N(\hat{\beta}, \hat{\gamma}) \rightarrow^{a.s.} E_0 \left[\ln \frac{f(Y_i | X_i, \beta^*)}{g(Y_i | Z_i, \gamma^*)} \right]$$

onde:

$$LR_N(\hat{\beta}, \hat{\gamma}) = \ln f(Y | X, \hat{\beta}) - \ln g(Y | Z, \hat{\gamma})$$

116. Note que não se fala em *fou g* serem "os modelos verdadeiros" (ou estarem corretamente especificados) — os modelos são aqui tratados simetricamente.

é a estatística da razão de verossimilhança. Sob a hipótese nula, tem-se então a seguinte convergência assintótica:

$$\frac{LR_N(\hat{\beta}, \hat{\gamma})}{\sqrt{N} \cdot \hat{w}_N} \rightarrow^d N(0, 1)$$

onde:

$$\hat{w}_N^2 = \frac{1}{N} \sum_{i=1}^N \left[\ln \frac{f(Y_i | X_i, \hat{\beta})}{g(Y_i | X_i, \hat{\gamma})} \right]^2 - \left[\frac{1}{N} \sum_{i=1}^N \ln \frac{f(Y_i | X_i, \hat{\beta})}{g(Y_i | Z_i, \hat{\gamma})} \right]^2$$

Ou seja, a estatística de teste pode ser entendida como a razão de log-verossimilhança média, convenientemente normalizada.

Definindo-se as hipóteses alternativas H_f e H_g segundo as quais os modelos f e g , respectivamente, são preferidos, o teste pode ser interpretado da seguinte forma: não se pode rejeitar a hipótese nula caso a estatística de teste esteja suficientemente próxima de 0; não se pode rejeitar H_f (H_g) caso a estatística de teste seja significativamente maior (menor) que 0. Em termos formais, Vuong (1989) mostra que, sob H_f , tem-se

$$\frac{LR_N(\hat{\beta}, \hat{\gamma})}{\sqrt{N} \cdot \hat{w}_N} \xrightarrow{as} +\infty$$

enquanto, sob H_g ,

$$\frac{LR_N(\hat{\beta}, \hat{\gamma})}{\sqrt{N} \cdot \hat{w}_N} \xrightarrow{as} -\infty$$

O resultado teórico sugere um procedimento simples ao se implementar o teste. Primeiro, escolhe-se um valor crítico c associado a um quantil de interesse da distribuição Normal-padrão; se a estatística de teste é maior que o valor crítico c , o modelo M_f é o preferido; caso a estatística de teste tome valor menos que $-c$, o modelo M_g é o preferido. Se o valor da estatística de teste estiver situado entre $-c$ e c , não se pode discriminar entre os modelos.

Caso o número de coeficientes difira entre os modelos, é preciso fazer uma correção de graus de liberdade, e Vuong sugere o uso de critérios como os de Akaike (1973) ou Schwarz (1978). Como o critério BIC de Schwarz é sabidamente mais indicado que o de Akaike, sugerimos o seu uso na correção, o que gera a estatística de teste corrigida:

$$\widetilde{LR}_N(\hat{\beta}, \hat{\gamma}) := LR_N(\hat{\beta}, \hat{\gamma}) - \left[\frac{p}{2} \ln(N) - \frac{q}{2} \ln(N) \right]$$

onde p e q são as dimensões dos vetores β e γ , respectivamente.

Ao usar o teste de Vuong, deve-se levar em conta que esse é um teste pouco potente — a convergência da estatística de teste para a distribuição Normal precisa de amostras de tamanho razoável para ocorrer.