Economia Industrial Empírica

Cláudio R. Lucinda[1]

30 de outubro de 2011

[1]Professor, Fundação Getúlio Vargas – Escola de Economia de São Paulo e Escola de Administração de Empresas de São Paulo. Endereço: Rua Itapeva, 474, 11º andar. CEP: 01420-001. Fone: +55 (11) 3281-7765. E-mail: claudio.lucinda@fgv.br .

Sumário

Prefácio

Tenho dois motivos pelos quais consolidei as minhas notas de aula em uma única apostila, que pretendo transformar em livro. O primeiro deles é consolidar em um único lugar um enorme conjunto de técnicas que estão dispersas em muitos lugares na literatura e, em especial, com notações e técnicas diferentes.

O segundo motivo é a falta de referências para os alunos de economia sobre como dar uma roupagem empírica a conceitos que, atualmente, são muito vistos apenas na teoria. Apesar de muito importantes para o entendimento da realidade, não são utilizados para a resolução de problemas práticos de política econômica e estratégia empresarial.

Capítulo 1

Introdução

Este capítulo tem por objetivo, em primeiro lugar, apresentar uma breve introdução a como evoluiu a análise empírica no campo da Organização Industrial, além de dar alguma características mais gerais da forma como a análise econométrica é praticada neste campo atualmente.

Neste sentido, o ponto básico deste livro é apresentar modelos empíricos que nos ajudem a compreender o comportamento de agentes econômicos – empresas e consumidores – em um contexto de mercado. Os modelos aqui apresentados estão fortemente vinculados ao que é chamado Nova Organização Industrial Empírica (*New Empirical Industrial Organizarion,* ou NEIO), e surgiu com uma contraposição ao paradigma de análise empírica prevalecente em boa parte dos anos 60 e 70 do século passado, conhecido como Estrutura-Conduta-Desempenho (ou *Structure-Conduct-Performance* – SCP).

Este paradigma SCP foi inicialmente desenvolvido por Edward S. Mason na década de 30 em Harvard, e se baseia na suposição que o <u>desempenho</u> das empresas ou mercados em particular é uma decorrência da <u>conduta</u> dos compradores e vendedores em tais mercados, envolvendo decisões de preços e outras práticas comerciais, agressividade da competição, estratégias tecnológicas, entre outros. A <u>conduta</u>, por sua vez, dependeria da <u>estrutura</u> do mercado, englobando fatores tais como tamanho e distribuição de tamanho entre as diferentes empresas, o grau de diferenciação dos produtos percebida pelos consumidores, grau de integração vertical entre as empresas. Esta estrutura de mercado, finalmente, seria o resultado de condições básicas de mercado, tais como condições de oferta de matérias primas, ambiente institucional, aspectos sazonais e climáticos. A figura a seguir ilustra bem estas relações.

Note-se que esta relação de causalidade, neste paradigma, é implícita pelo analista. Segundo Scherer (1970, p. 5)[42], "..we seek theories which permit us to predict ultimate market performance from the observation of structure, basic conditions and conduct.". Ainda que os pesquisadores desta linha reconhecessem que, em boa parte das situações práticas, a causalidade não era tão simples assim – a conduta afetando a estrutura de mercado ao mesmo tempo em que é afetada

Figura 1.1: Paradigma SCP

CONDIÇÕES BÁSICAS

OFERTA	DEMANDA
Matéria-Prima	Elasticidade da Demanda
Tecnologia	Crescimento da Demanda
Durabilidade do Produto	Substitutos
Valor/Peso	Tipo de Comercialização
Sindicalização	Sazonalidade

↓

ESTRUTURA DE MERCADO

Número de Compradores/Vendedores
Diferenciação de Produtos
Barreiras à Entrada
Estrutura de Custos
Integração Vertical
Conglomeração

↓

CONDUTA

Precificação
Estratégia de Produtos
Pesquisa e Desenvolvimento
Publicidade
Estratégias Jurídicas

↓

DESEMPENHO

Eficiência Alocativa e Produtiva
Progresso
Pleno Emprego
Equidade

por ela – esta questão e suas implicações empíricas não foram levadas adiante.

Outro ponto importante do paradigma SCP é a pouca ênfase dada à estimação econométrica de elementos fundamentais para a análise teórica, tais como Custos Marginais e Elasticidades. Por exemplo, Joe Bain (1941, p. 272), afirma:

> "...the direct statistical measurement of the elasticity of demand or of marginal costs is in most cases next to impossible...One may proceed instead with a qualitative analysis of the characteristics of industrial markets (including numbers, degree of dierentiation of the product, etc.) in order to nd where monopoly power may be expected to existe and where (assuming prot maximization) it is exploited"

As limitações do paradigma SCP levaram uma mudança de foco na análise empírica em Organização Industrial, dando origem à já mencionada Nova Organização Industrial Empírica, que se diferencia da SCP nas seguintes dimensões:

- Para a NOIE, as margens entre Preço e Custo Marginal não são diretamente observáveis. No modelo SCP, utilizavam-se dados contábeis para a obtenção de estimativas de CMg e, a partir daí, obtinham-se medidas de índices de Lerner e avaliava-se o poder de mercado das empresas. Agora, infere-se o Custo Marginal a partir do comportamento das empresas ou estima-se diretamente o poder de mercado sem calcular diretamente o Custo Marginal;

- Diferentemente do SCP, as comparações entre indústrias são pouco reveladoras em termos de poder de mercado; ou seja, considerar uma empresa como possuidora de poder der mercado somente com base no fato que a margem é maior do que outras indústrias (análise comum no paradigma SCP) é pouco revelador. Na abordagem da NEIO, cada indústria possui elementos distintos que tornam esta análise extremamente problemática.

- A conduta de uma empresa – bem como a de uma indústria como um todo – é algo que deve ser estimado, a partir de elementos comportamentais sobre a forma da demanda, do comportamento das empresas e interação estratégica.

- Desta forma, a inferência sobre o poder de mercado é fortemente atrelada com a teoria econômica, na medida que as hipóteses alternativas estão claramente explicitadas.

Estas diferenças implicam também em mudanças na forma de análise econométrica utilizada. Enquanto que no paradigma SCP a análise empírica era realizada em grande medida com base nas chamadas "formas reduzidas[1]", com o reconhecimento da inter-relação entre todos os aspectos do comportamento dos agentes no mercado temos que a abordagem econométrica mais relevante é a chamada *Econometria Estrutural*. Iremos detalhar mais aprofundadamente o que entendemos por econometria estrutural a seguir

[1] Formas Reduzidas – Abordagem econométrica em que se busca caracterizar o comportamento de uma variável y em função apenas de variáveis exógenas.

1.1 Modelagem Econométrica Estrutural

Podemos entender a análise empírica em economia como organizada em dois campos: descritivo e estrutural, sendo que os dois estarão preocupados - direta ou indiretamente – com a distribuição conjunta entre duas ou mais variáveis, que iremos denotar aqui por $f(x, y)$. A primeira abordagem, a descritiva, é focada em construir e resumir dados econômicos, com um foco principal em descrever tendências, padrões e associações que podem estimular pesquisas subsequentes. É importante notar que trabalhos empíricos muito importantes na literatura podem ser considerados como pertencentes a esta linha, como o de A. W. Phillips sobre a relação entre inflação e desemprego, a relação que deu origem à curva de Engel, entre outros.

Para isto, um elemento chave é a distribuição conjunta entre as variáveis cuja relação se busca entender, coisa que denominaremos $f(x, y)$. No entanto, na maior parte dos casos, a distribuição conjunta entre estas variáveis não é diretamente estimada, mas sim elementos que são derivados desta distribuição, como por exemplo:

- A distribuição condicional de y dado x, $f(y|x)$;
- A Esperança condicional de y dado x, $E(y|x)$;
- A Correlação (ou covariância) condicional de y dado x, $Corr(y|x)$ ou $Cov(y|x)$
- Um quantil específico – α– da distribuição de y dado x – $Q_\alpha(y|x)$;
- O Melhor Preditor Linear de y dado um valor para x – $BLP(y|x)$

Como na maior parte das vezes não temos dados suficientes para estimar diretamente por métodos não-paramétricos ou semiparamétricos a distribuição conjunta $f(x, y)$, ou de qualquer um dos elementos listados acima, os pesquisadores da área em geral lançam mão de técnicas paramétricas. Só que, para conseguirmos trabalhar com alguns destes pontos sem que tenhamos que passar pela distribuição conjunta, devemos trazer da teoria econômica elementos que restrinjam adequadamente a distribuição conjunta de y e x. Com estas restrições, podemos estimar a distribuição condicional – ou os momentos condicionais – e, daí, se necessário for, voltar à distribuição conjunta. Esta é a essência da abordagem econométrica estrutural.

Note-se que, tanto na abordagem descritiva quanto na abordagem estrutural busca-se, ao fim e ao cabo, caracterizar a distribuição conjunta. No entanto, na abordagem econométrica estrutural buscam-se parêmetros ou primitivas econômicas da distribuição conjunta, enquanto que nos modelos descritivos a idéia principal é caracterizar simplesmente a distribuição conjunta. Note-se que **a busca destas primitivas ou parâmetros da distribuição conjunta é sempre dependente destas premissas que limitam a distribuição conjunta.** Os elementos essenciais de um modelo estrutural são as hipóteses econômicas e estatísticas, as quais deveriam ser, pelo menos, razoáveis econômica e estatisticamente.

Mas preste atenção: mesmo quando o analista não deriva explicitamente o modelo estrutural, **qualquer conclusão de ordem causal ou comportamental está implicitamente se baseando em um modelo estrutural**. Em muitos livros de econometria este ponto não é claro o suficiente[2], dando margem a alguns a pensar que existe um animal denominado "análise de forma reduzida". Em geral, este animal depende de teoria econômica apenas para determinar o que vai na categoria y e o que vai na categoria x. Definidas estas variáveis, em geral se utiliza regressão linear (com os estimadores ajeitados adequadamente) para estimar os coeficientes das variáveis que estão do lado direito da igualdade, como na análise abaixo:

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \varepsilon_{it}$$

O que vamos discutir brevemente é que, ou este modelo deve abandonar qualquer pretensão de oferecer afirmações causais ou comportamentais, ou ele deve ser o resultado de um modelo estrutural propriamente dito. Para entender isso melhor, vamos supor que nenhuma das duas opções aconteça, e que saiamos estimando um modelo como o acima, na ilusão que ele nos dá uma relação entre x e y, mas que ele não seja explicitamente derivado de nenhuma teoria. O grande problema é que, além da nossa fé, nada garante que o modelo anterior seja o verdadeiro, ou este aqui:

$$x_{1it} = \gamma_1 + \gamma_2 y_{it} + \gamma_3 x_{2it} + \epsilon_{it}$$

Supondo que a fé não seja um guia adequado, o que podemos afirmar a partir do conjunto de estimativas $[\alpha, \beta_1, \beta_2]$ ou $[\gamma_1, \gamma_2, \gamma_3]$? Sem um modelo econômico, o máximo que podemos dizer da relação entre xe y vem da estatística. Geralmente, quando o analista se encontra nesta situação, acabamos por chegar em debates sobre a busca do "instrumento adequado", ou dos problemas de endogeneidade que ninguém sabe muito bem de onde vem. **O grande problema é que, na ausência de um modelo econômico que amarre especificamente uma especificação a uma média condicional da variável dependente, é impossível ligar a evidência de análise de regressão às predições de estática comparativa** (tudo o mais mantido constante, alterando a variável "tal", temos um dado efeito sobre a variável dependente...). Para entendermos melhor esta afirmação, vamos passar a um exemplo específico. Suponhamos que tenhamos informações sobre as seguintes variáveis:

- Q - Produção Total (unidades)
- TC - Custo Total
- p_K - Preço do Capital
- p_L - Preço do Trabalho

[2]Em especial, nos livros que são muito abundantes em aplicações de Economia do Trabalho...

Qual dos dois modelos a seguir seria o "certo"?

$$\ln TC = \theta_1 + \theta_2 \ln Q + \theta_3 \ln p_K + \theta_4 \ln p_L + \epsilon \quad (1.1)$$
$$\ln Q = \beta_1 + \beta_2 \ln TC + \beta_3 \ln p_K + \beta_4 \ln p_L + \epsilon \quad (1.2)$$

A primeira intuição de quase todos os economistas é de dizer que a primeira das especificações seria a "correta", pois ela pareceria uma função de custos como a que teríamos no livro do Varian [50] ou no MWG[35]. Se queremos usar os dados para derivar relações causais entre as variáveis, devemos nos perguntar duas questões:

1. Quais são as premissas econômicas e estatísticas que justificam cada um destes modelos?

2. Estas premissas fazem sentido para a questão em tela?

Vamos detalhar brevemente quais seriam as condições sob as quais cada uma das equações seria válida, começando com a primeira delas, a equação (1.1), que tem como variável dependente o Custo Total e os preços dos insumos como variáveis independentes. Neste caso, começamos imaginando que as empresas para as quais temos os dados operam de acordo com a boa e velha Cobb-Douglas, e que conheçam os valores de todos os parâmetros dela:

$$Q = AK^\alpha L^\beta$$

Além disso, vamos imaginar que estamos em uma indústria regulada e que cada uma das empresas tem uma tecnologia diferente (suposta como sendo i.i.d. e positiva). Cada uma das empresas deve ter a seguinte função de lucros:

$$\pi = pAK^\alpha L^\beta - p_K K - p_L L$$

Vamos supor que o regulador – interessado no bem-estar social, e, por isso, buscando minimizar o Custo Médio de produção – estabelece o preço ao qual a empresa pode vender o seu produto e, a partir daí, a empresa vai lá e escolhe a alocação de K e L. Fazendo a otimização da empresa, chegamos à seguinte função:

$$TC = C_0 p_K^\gamma p_L^{1-\gamma} Q^\delta A^{-\delta}$$

Em que $\delta = 1/(\alpha + \beta)$. Passando o log na equação acima, temos:

$$\ln TC = \ln C_0 + \gamma \ln p_K + (1-\gamma) \ln p_L + \delta \ln Q - \alpha \ln A$$

Só que não observamos o elemento A. Dessa forma, a equação acima vai ficar sendo o seguinte:

$$\ln TC = C_1 + \gamma \ln p_K + (1-\gamma) \ln p_L + \delta \ln Q + \ln u$$

Em que a nova constante, C_1 absorve a média condicional de A, dados os valores dos outros regressores:

$$C_1 = \ln C_0 + E[\ln A | \ln p_K, \ln p_L, \ln Q]$$

E o termo erro de média zero absorve a parte não esperada dos choques de produtividade:

$$\ln u = \frac{1}{\delta}(\ln A - E[\ln A | \ln p_K, \ln p_L, \ln Q])$$

Como assumimos que os termos A são i.i.d., temos que a esperança condicional que é somada no termo C_1 é igual à esperança não condicional. Além disso, como é o regulador – e não as empresas, que se supõe que conheçam o termo A– estabelece o preço para minimizar o custo médio de produção, podemos estimar os parâmetros da equação com ajuda de OLS. Isso ocorre porque a parte não observada dos choques de produtividade não está relacionada com a quantidade produzida – pois ela não foi produzida levando-se em conta o valor completo de A.

A seguir, vamos observar situações em que a equação (1.2) é a mais adequada para a estimação por OLS. Neste caso, os dados são os mesmos, mas as empresas não sabem o valor do parâmetro A na função Cobb-Douglas anteriormente colocada. Neste caso, ela terá de tomar as suas decisões sem saber o valor exato deste parâmetro, maximizando o lucro com base na esperança de A:

$$E[\pi(p, K, L)] = E[pAK^{\alpha}L^{\beta}] - p_K K - p_L L$$

Sobre o procedimento de preços, vamos supor que ele seja determinado por um regulador que, da mesma forma que a empresa, não sabe o valor exato do parâmetro A. Este regulador, como no caso anterior, irá colocar o preço de forma que a empresa terá lucros nulos em equilíbrio. Isso não impede que a empresa maximize os lucros de acordo com a forma a seguir:

$$E[\pi(p, K, L)] = p^{reg}E[AK^{\alpha}L^{\beta}] - p_K K - p_L L$$

As CPO deste problema de maximização são:

$$L = \frac{\alpha p_K}{\beta p_L} K$$

O que implica a seguinte função custo:

$$TC = \frac{\alpha + \beta}{\beta} p_K K$$

Que não depende do parâmetro A. Substituindo as duas equações na função de produção, chegamos a uma equação que relaciona a produção total aos custos totais e aos preços dos insumos:

$$Q = D_0 TC^{\alpha+\beta} p_K^{-\beta} p_L^{-\alpha} A$$

Passando logs dos dois lados, temos a seguinte função custo:

$$\ln Q = \ln D_0 + (\alpha + \beta) \ln TC - \beta \ln p_K - \alpha \ln p_L + \ln A$$

Nem as empresas nem o econometrista observa o valor de A, o que indica que ele deve tratá-los como erros aleatórios, o que indica que a parte prevista pelas outras variáveis independentes vai ser absorvida pela constante e desvios em relação a esta parte serão absorvidos pelo termo erro, ou seja:

$$\begin{aligned} \ln Q &= D_1 + (\alpha + \beta) \ln TC - \beta \ln p_K - \alpha \ln p_L + \eta \\ \eta &= \ln A - E[\ln A | \ln TC, \ln p_K, \ln p_L] \\ D_1 &= \ln D_0 + E[\ln A | \ln TC, \ln p_K, \ln p_L] \end{aligned}$$

Dadas as hipóteses anteriores, podemos estimar esta equação por OLS. Portanto, para um determinado conjunto de premissas do comportamento dos agentes, uma especificação pode ser a "correta", enquanto que sob um outro conjunto de premissas, outra especificação pode ser a adequada. A principal dificuldade, a princípio, é como pensar de forma organizada sobre estas premissas. Para ajudar, Reiss e Wolak (2003)[40] compilaram uma estrutura para a especificação de modelos econométricos estruturais em Organização Industrial. Esta estrutura pode ajudá-los tanto para montar os seus modelos – que serão cobrados nos papers que vocês tem de apresentar no final do curso – quanto para avaliar as premissas dos modelos econométricos encontrados na literatura.

As premissas de ordem econômica são:

- Descrição do Ambiente Econômico, incluindo:
  - Tamanho do mercado e instituições
  - Os agentes econômicos
  - A informação disponível para cada um dos agentes
- Uma lista de "primitivas":
  - Tecnologia
  - Preferências
  - Dotações de Fatores de Produção
- Variáveis exógenas aos agentes, como por exemplo:
  - Restrições ao comportamento dos agentes

  - Variáveis externas ao modelo que influenciam o comportamento dos agentes

- Variáveis de Decisão, Horizontes de Tempo e Funções Objetivo dos Agentes, tais como:
  - Maximização de Utilidade dos Consumidores
  - Maximização dos lucros da empresa
- Um conceito de solução de jogos de equilíbrio, tal como:
  - Leiloeiro Walrasiano com comportamento de competição perfeita pelos consumidores
  - Equilíbrio de Nash em cima das variáveis de decisão mencionadas antes;

Boa parte das técnicas utilizadas para o desenvolvimento adequado de premissas deste tipo serão desenvolvidas em livros-texto de Organização Industrial como o do Shy[44], o do Carlton e Perloff[18] ou o do Tirole[47]. O segundo conjunto de premissas tem recebido muito menos atenção do que merece, que é o passo de mapear o modelo econômico com as características acima em um modelo econométrico. **A principal diferença entre um modelo econômico e um modelo econométrico é que este último inclui elementos não observáveis que absorvem as implicações do fato que o modelo econômico não se ajusta perfeitamente aos dados observados.** Note-se que o processo de inclusão destes elementos não deve ser arbitrário, pois a escolha destes elementos pode possuir implicações importantes para a distribuição dos estimadores dos parâmetros relevantes do modelo. Existem quatro formas pelas quais estes elementos não observáveis podem ser incorporados em um modelo:

- **Incerteza do analista e das empresas sobre o ambiente econômico**: A incerteza do analista e a incerteza dos agentes sobre o ambiente econômico tem efeitos dramaticamente diferentes quando consideramos a construção de modelos econométricos. Como vimos anteriormente, a hipótese de que o termo A não é conhecido pelas empresas faz com que uma especificação seja de estimação adequada, enquanto uma premissa diferente leva a resultados distintos. Em geral, chamamos uma situação em que os agentes tomam uma decisão com base em uma variável que não é observada pelo econometrista como um caso de heterogeneidade não observável.
- **Erros de Otimização dos Agentes:**
- **Erros de Medida:** Em muitos casos, os erros de medida podem ser um elemento importante do componente aleatório.

A partir destas premissas econômicas e estatísticas, o passo seguinte envolve a montagem do modelo econométrico propriamente dito; com isto, estou querendo dizer impor hipóteses distribucionais e paramétricas específicas sobre as variáveis. Com isto, pode-se utilizar uma técnica de estimação (OLS, IV, 3SLS, GMM, ML – devidamente revisados no Apêndice) e testar as hipóteses impostas no modelo. Este processo pode ser considerado como possuindo quatro pontos:

- Seleção de Forma Funcional
- Seleção de Hipóteses Distribucionais
- Seleção de Técnica de Estimação
- Seleção de Testes de Especificação.

Em geral, o que iremos discutir com maior profundidade neste curso são os dois primeiros tópicos, com foco nos dois outros somente em alguns momentos. A primeira parte do curso irá discutir a questão da forma funcional e hipóteses distribucionais para um dos elementos importantes em um modelo econométrico estrutural em Organização Industrial: o lado da demanda. Antes de fazermos isso, precisaremos revisar alguns pontos referentes à teoria do consumidor, que será o tema da primeira parte deste livro.

Parte I

Demanda

Capítulo 2

Modelos Neoclássicos de Demanda

A primeira abordagem econométrica moderna utilizada para o entendimento do comportamento do consumidor será aqui caracterizada como "Modelagem Neoclássica de Demanda", e uma boa introdução ao tema pode ser encontrada em Stigler (1954)[45], bem como em Brown e Deaton (1972)[15]. Esta abordagem, ainda que diretamente baseada na teoria neoclássica de demanda, é o resultado de pesquisas sobre o comportamento do consumidor que datam desde o final do século XVIII, e que ganham força com os estudos de Ernst Engel em meados do século XIX sobre a relação entre consumo e renda, e os estudos de Henry Moore no começo do século XX sobre a relação entre consumo e preços dos bens. Estes dois temas retornarão na análise aqui apresentada como perguntas de pesquisa bastante relevantes, além de um outro tema que apenas poderia ser estudado com a formalização da teoria microeconômica: a análise de bem-estar decorrente de medidas de política econômica.

Do ponto de vista de estimação, devemos notar notar que a análise destes modelos começa no pós-guerra, quando começam a estar disponíveis de forma regular dados mais desagregados sobre os padrões de consumo dos indivíduos, com Pesquisas de Orçamentos Familiares. Uma introdução boa a este tema – sistema de demanda neoclássicos – pode ser encontrada em Pollak e Wales (1992)[39].

Neste capítulo, concentraremos nossa atenção nos principais modelos empíricos de demanda baseados na tradição microeconômica neoclássica, usualmente baseados na estimação de sistemas de equações. O primeiro passo desta análise é discutir, exatamente, quais são as características de um sistema de equações que podem caracterizá-lo como adequado. Entende-se um sistema de demanda como teoricamente plausível se ele for consistente com o processo de maximização da utilidade do consumidor. Em especial, este conceito pode ser operacionalizado verificando-se as seguintes condições[1]:

- *Adding-up* (ou exaustão da restrição orçamentária): supõe-se que o valor das demandas por

[1] Um excelente detalhamento sobre o tema encontra-se em Barten (1977) [8]

todos os bens exaure o valor da restrição orçamentária. Em palavras, isso quer dizer que o valor do dispêndio previsto pelo modelo tem de ser igual ao valor do dispêndio observado sempre. Podemos ver isso usando a relação $\sum_k p_k \frac{\partial g_k}{\partial w} = 1$ ou $\sum_k p_k \frac{\partial g_k}{\partial p_i} + q_i = 0$ ou , ou seja:

$$\sum_k p_k h_k = \sum_k p_k x_k = w$$

- Homogeneidade: as demandas hicksianas são homogêneas de grau zero nos preços, e as demandas Marshallianas no gasto total e nos preços, ou seja, para escalar $\theta > 0$. A intuição deste resultado é que mudanças inflacionárias perfeitamente antecipadas, que elevam os preços e o dispêndio total em uma mesma proporção, mantém o padrão de consumo inalterado. Ou ainda, o consumidor toma suas decisões independente da unidade de medida monetária (ou seja, o consumo é exatamente o mesmo se ele fizer a conta em Reais ou em centavos).
$$h_i(u, \theta p) = h_i(u, p) = x_i(\theta w, \theta p) = x_i(w, p)$$

A primeira condição possui um caráter mais formal, para guiar a modelagem, enquanto que a segunda, ainda que seja uma premissa sobre o comportamento individual racional, é de um caráter mais geral. As duas últimas propriedades, por outro lado, dizem respeito diretamente a como o comportamento do consumidor é entendido pela teoria microeconômica neoclássica, e um detalhamento de como estas questões são importantes encontra-se no Apêndice. Tais propriedades são diretamente relacionadas com a matriz de efeitos preço da função demanda hicksiana.

- Simetria: As derivadas cruzadas das demandas Hicksianas são simétricas.

- Negatividade: A matriz de derivadas $\bigtriangledown_p h(u, p)$ das demandas hicksianas com relação aos preços tem que ser negativa semidefinida. Esta propriedade pode ser testada por meio da nossa querida Equação de Slutsky.

Note-se que nem sempre temos estas condições atendidas por todas as especificações. Por exemplo, uma forma funcional que é muito comum é a log-linear e que não atende à primeira das condições. Para ilustrar isso melhor vamos supor que tenhamos um sistema de equações para N produtos, sendo que cada equação possui a seguinte forma funcional:

$$\ln q_i = \alpha_i + e_i \ln w + \sum_k e_{ik} \ln p_k + u_i$$

Esta função duplo log é muito comumente utilizada porque os coeficientes estimados são diretamente as elasticidades. Vamos definir o logaritmo da participação no gasto como sendo $\ln s_i = \ln q_i + \ln p_i - \ln w$. Substituindo isso na equação acima, temos que:

$$\ln s_i = \alpha_i + (e_i - 1) \ln w + (e_{ii} + 1) \ln p_i + \sum_{k \neq i} e_{ik} \ln p_k$$

Pela restrição de exaustão da restrição orçamentária, mencionada acima, temos que $\sum_k w_k e_k = 1$, o que indica que *ou tereemos todos as elasticidades renda iguais a um ou pelo menos uma delas tem que ser maior do que um.* Se o primeiro caso for verdade, o modelo não é muito interessante, uma vez que ele implica padrões de consumo iguais a todos os níveis de gasto, o que não é verdade. Se o segundo caso for verdade, existe necessariamente um produto de luxo, para o qual $e_i > 1$ e um produto inferior para o qual $e_i < 1$. O problema é que, neste caso, para alguns intervalos de renda o gasto é menor do que o orçamento e, em outras, a situação se reverte. O exemplo a seguir ilustra este ponto:

Exemplo 1: No arquivo `EVIEWS.txt` estão os dados para 270 famílias (Unidades de Consumo) brasileiras extraídos da pesquisa de orçamentos familiares de 1995/96, com as seguintes informações:

- P_CARNE: preço médio em R$ do quilo de carne enfrentado por família ao longo das 52 semanas da pesquisa.
- P_FRANGO: preço médio em R$ do quilo de frango enfrentado por família ao longo das 52 semanas da pesquisa.
- P_PEIXE: preço médio em R$ do quilo de peixe enfrentado por família ao longo das 52 semanas da pesquisa.
- P_PORCO: preço médio em R$ do quilo de carne de porco enfrentado por família ao longo das 52 semanas da pesquisa.
- V_CARNE: valor médio semanal em Reais dos gastos com carne por família.
- V_FRANGO: valor médio semanal – em Reais – dos gastos com carne de frango por família.
- V_PEIXE: valor médio semanal – em Reais – dos gastos com carne de peixe por família.
- V_PORCO: valor médio semanal – em Reais – dos gastos com carne de porco por família.
- Q_CARNE: consumo médio semanal – em quilogramas – de carne bovina por família.
- Q_FRANGO: consumo médio semanal – em quilogramas – de carne de frango por família
- Q_PEIXE: consumo médio semanal – em quilogramas – de carne de peixe por família
- Q_PORCO: consumo médio semanal – em quilogramas – de carne de porco por família.

Estimando o modelo duplo log para três dos bens[2], chegamos aos seguintes coeficientes:

	Sensibilidade com relação ao preço de				
	p_c	p_f	p_p	p_{po}	w
q_c	-1.0041	0.1399	0.0563	-0.0097	1.0795
	(-13.9963)	(2.3055)	(2.0931)	(-0.3743)	(68.1477)
q_f	0.0158	-0.9649	-0.1084	-0.0458	0.9911
	(0.1701)	(-12.2560)	(-3.1072)	(-1.3650)	(48.2296)
q_p	0.1745	-0.7029	-0.9115	0.0017	1.1714
	(0.4779)	(-2.2756)	(-6.6592)	(0.0132)	(14.5300)

OBS: Estatísticas t entre parênteses

Do ponto de vista das elasticidades-preço, os resultados são bastante razoáveis. Todos os produtos apresentam elasticidade-preço próxima de um, frango e peixe são considerados como substitutos à carne de boi, e peixe e frango como complementos. O grande problema aqui está na elasticidade-dispêndio da demanda. Com os coeficientes acima, pequenas alterações no dispêndio total ou fazem com que a participação no gasto da carne de porco seja negativa, ou o valor do gasto será diferente da soma das quantidades projetadas pelo modelo, multiplicada pelos preços.

Apesar desta rejeição da primeira condição, em alguns casos afirma-se que, para a análise de um único produto, o uso da equação duplo log acima não nos obriga a montar um sistema em que as outras equações sejam da mesma forma. No entanto, isso é verdade para apenas um intervalo pequeno dos dados.

2.1 Abordagem Diferencial e o Modelo de Rotterdam

Uma alternativa de modelagem empírica de demanda envolve aproximar diretamente a função demanda resultante do processo de maximização da utilidade do consumidor, que é o resultado do trabalho de Theil (1965)[46]. A idéia no artigo original era formular a função demanda em termos de teoria da informação e de números índices de preço. Atualmente, a melhor forma de derivar este tipo de abordagem é começar com a função demanda por um produto qualquuer:

$$x_l = x(\mathbf{p}, w)$$

Tirando a diferencial total desta função, temos:

$$dx_l = \frac{\partial x_l}{\partial w} dw + \sum_{j=1}^{L} \frac{\partial x_l}{\partial p_j} dp_j$$

Lembrando que o diferencial total de uma variável k, dk, pode ser entendido como $kd\log k$, ao multiplicarmos os dois lados por ${p_l}/{w}$ chegamos à seguinte equação:

$$s_l d\log x_l = \frac{\partial x_l}{\partial w} \frac{x_l}{x_l} \frac{p_l}{w} wd\log w + \sum_{j=1}^{L} \frac{\partial x_l}{\partial p_j} \frac{p_l}{w} p_j d\log p_j$$

O primeiro dos termos do lado direito da igualdade pode ser escrito como o produto da participação no orçamento do produto l e a elasticidade-renda da demanda do mesmo produto, o que seria o share marginal do produto l

$$\theta = s_l \times \epsilon_w^l$$

A equação fica sendo então:

$$s_l d\log x_l = \theta_l d\log w + \sum_{j=1}^{L} \frac{\partial x_l}{\partial p_j} \frac{p_l}{w} p_j d\log p_j$$

O grande problema é o termo da somatória. A paartir da microeconomia tradicional, pode se mostrar que:

$$\nabla_p \mathbf{x} = \lambda \mathbf{U}^{-1} - \left(\frac{\lambda}{{\partial\lambda}/{\partial w}} \right) \nabla_w \mathbf{x} (\nabla_w \mathbf{x})^T - (\bigtriangledown_w x(\mathbf{p}, w))(x(\mathbf{p}, w))^T$$

Em que λ é o multiplicador de lagrange associado com a restrição orçamentária no problema dual e a matriz $\mathbf{U}$ é a matriz Hessiana da função utilidade[3]. Ou seja, podemos escrever o efeito marginal do preço sobre a demanda marshalliana como sendo:

$$\frac{\partial x_i}{\partial p_j} = \lambda u_{ij} - \frac{\lambda}{\partial\lambda/\partial w}\frac{\partial x_i}{\partial w}\frac{\partial x_j}{\partial w} - \frac{\partial x_i}{\partial w}x_j$$

Substituindo isso na função original, nós temos:

$$\begin{aligned} s_l d\log x_l &= \theta_l d\log w + \sum_{j=1}^{L}\left(\frac{\lambda u_{ij}p_l p_j}{w} - \frac{\lambda/w}{\partial\lambda/\partial w}\theta_l\theta_j - \theta_j s_j\right) d\log p_j \\ s_l d\log x_l &= \theta_l\left(d\log w - \sum_j s_j d\log p_j\right) + \sum_{j=1}^{L}\left(\frac{\lambda u_{ij}p_l p_j}{w} - \frac{\lambda/w}{\partial\lambda/\partial w}\theta_l\theta_j\right) d\log p_j \end{aligned}$$

O termo $\sum_j s_j d\log p_j$ representa a variação em um índice de preços de Divisia – uma aproximação linear a um índice de preços que será usada mais adiante – e o representaremos como $d\log\mathbf{P}$. O primeiro termo em parênteses é um índice Divisia de quatidades[4], ou seja, o primeiro termo é $\theta_l d\log\mathbf{Q}$.

Agora, a equação fica sendo:

$$s_l d\log x_l = \theta_l d\log\mathbf{Q} + \sum_{j=1}^{L}\left(\frac{\lambda u_{ij}p_l p_j}{w} - \frac{\lambda/w}{\partial\lambda/\partial w}\theta_l\theta_j\right) d\log p_j$$

Para simplificarmos mais a exposição, vamos definir $\frac{\lambda u_{ij}p_l p_j}{w} = v_{ij}$ e $\frac{\lambda/w}{\partial\lambda/\partial w} = \phi$, o que permite que escrevamos:

$$s_l d\log x_l = \theta_l d\log\mathbf{Q} + \sum_{j=1}^{L} v_{lj} d\log p_j - \phi\theta_l\sum_{j=1}^{L}\theta_j d\log p_j$$

Pode-se mostrar que $\sum_j v_{lj} = \phi\theta_l$. Além disso, o termo $\sum_{j=1}^{L}\theta_j d\log p_j$ também é uma aproximação à variação de um índice de preços, o de Fritsch. Desta forma, a equação fica sendo:

$$s_l d\log x_l = \theta_l d\log\mathbf{Q} + \sum_{j=1}^{L} v_{ij}\left(d\log p_j - d\log P^f\right)$$

[3]Uma fonte muito útil para esta derivação é Barnett e Serletis (2009)[7].

[4]Para termos certeza disso, é só lembrar que $\sum_L s_l d\log p_l + \sum_L s_l d\log x_l = d\log w$

Uma versão alternativa – chamada versão em preços relativos – desta equação é dada por:

$$s_l d \log x_l = \theta_l d \log \mathbf{Q} + \sum_{j=1}^{L} \pi_{ij} d \log p_j$$

Sendo que $\pi_{ij} = v_{ij} - \phi\theta_i\theta_j$ e dá o efeito de substituição total entre a demanda do bem l e o preço do bem j. Para impormos as restrições tradicionais, precisamos que os coeficientes atendam às seguintes restrições:

- *Adding-Up*: $\sum_j \theta_j = 1$ e $\sum_l \pi_{lj} = 0$, para todos os l
- Homogeneidade: $\sum_j \pi_{lj} = 0$, em uma mesma equação
- Simetria da matriz de Slutsky: $\pi_{ij} = \pi_{ji}$
- Concavidade: a matriz de Slutsky precisa ser negativa semidefinida com posto $L - 1$.

As elasticidades preço compensadas e elasticidade renda são:

$$\begin{aligned} \epsilon_{ij} &= \frac{\pi_{ij}}{s_i} \\ \epsilon_w &= \frac{\theta_l}{s_l} \end{aligned}$$

2.2 Linear Expenditure System

Começaremos pelo *Linear Expenditure System*. Este modelo é de Klein e Rubin, nos seus papers de 1947-1948[32], e começa com a seguinte função de utilidade indireta[5]:

$$v(\mathbf{P}, w) = \frac{w - \sum p_k b_k}{\Pi_k p_k^{a_k}}$$

Usando a Identidade de Roy, chegamos às seguintes formas funcionais para as equações:

$$s_i = \frac{p_i b_i}{w} + a_i \left[1 - \frac{\sum_k p_k b_k}{w} \right]$$

[5]Este é um dos casos em que conseguimos calcular diretamente as funções demanda a partir de funções utilidade postuladas (e não as funções de utilidade indireta). Com as chamadas *Funções de utilidade Stone-Geary* $U = \prod_{i=1}^{n} (q_i - b_i)^{a_i}$, podemos derivar as funções demanda da forma $q_i = b_i - \frac{a_i}{p_i}(w - \sum_k p_k b_k)$, que após reorganização adequada, leva à expressão dos *shares* exposta mais adiante. Uma observação: saindo diretamente da função utilidade Stone Geary acima, a função de utilidade indireta é igual à do texto, multiplicada por um termo $\prod_k b_k$.

Este modelo, pode ser entendido como uma generalização de um modelo de demanda derivado de uma função de utilidade à la Cobb-Douglas – caso todos os $b_i = 0$, teríamos participações de cada um dos bens no dispêndio total iguais aos termos respectivos a_i.

O legal deste modelo é que os parâmetros possuem interpretações comportamentais. Uma família cujo sistema de demanda é LES começa comprando quantidades "comprometidas" de cada um dos bens $(b_1, b_2, \cdots, b_n)$, e depois dividindo o excedente, $w - \sum_k p_k b_k$ entre os bens em proporções fixas $(a_1, a_2, \cdots, a_n)$. A função dispêndio, importante para a avaliação de bem-estar de algumas medidas de política, pode ser expressa como:

$$e(\mathbf{P}, u) = \sum_k p_k b_k + u \prod_k \left(\frac{p_k}{a_k}\right)^{a_k}$$

Finalmente, as elasticidades deste sistema de equações, medidas sumárias de sensibilidade a preço e a dispêndio, são dadas por:

$$\begin{aligned}
e_{ii} &= \frac{p_i b_i (1 - a_i)}{p_i b_i + a_i \left(w - \sum_k p_k b_k\right)} - 1 \\
e_{ij} &= \frac{-a_i b_j p_j}{p_i b_i + a_i \left(w - \sum_k p_k b_k\right)} \\
e_w &= \frac{a_i w}{p_i b_i + a_i \left(w - \sum_k p_k b_k\right)}
\end{aligned}$$

Ainda que esta forma funcional seja atraente, por ser baseada na teoria econômica e facilmente implementável do ponto de vista econométrico, ele é pouco adequado para a análise de dados de famílias, em que as premissas do LES são claramente rejeitadas. Em termos de premissas, o que estamos querendo dizer é que o consumo de cada família tem de ser estritamente maior que os b_i e, além disso, as curvas renda-consumo são retas positivamente inclinadas saindo do ponto caracterizado pelo vetor $\mathbf{b}$.

Exemplo 2: Com o uso do arquivo `EVIEWS.txt` foi estimado o LES para estes bens, com os seguintes coeficientes:

	Coeficiente	Est. T	p-Valor
b_c	-0.0173	-0.5261	0.5991
a_c	0.5533	78.6766	0.0000
b_f	0.0451	1.1034	0.2704
b_{pe}	0.0147	1.5644	0.1184
b_{po}	0.0396	7.2034	0.0000
a_f	0.3197	57.8076	0.0000
a_{pe}	0.0914	15.4548	0.0000
a_{po}	0.0356		0.0000

Estes resultados indicam, em primeiro lugar, que as quantidades de subsistência para carne, frango e peixe são próximas de xero, a única estatisticamente diferente de zero é a quantidade de porco – aproximadamente 40 gramas. A partir desta quantidade básica de consumo, a renda além deste mínimo é distribuída com aproximadamente 55% indo para o gasto em carne bovina, 31% em carne de frango, 9% em peixe e o restante em carne de porco.
As elasticidades preço e cruzadas estão expostas a seguir:

	Sensibilidade com relação ao preço de				
	p_c	p_f	p_p	p_po	w
Q_c	-1.0051	-0.0125	-0.0067	-0.0156	1.0399
Q_f	0.0060	-0.9748	-0.0064	-0.0148	0.9900
Q_p	0.0058	-0.0115	-0.9384	-0.0144	0.9584

Os resultados da tabela a seguir são muito próximos aos verificados no modelo duplo log, em especial nas elasticidades preço (próprias) e nas elasticidades-renda. A principal diferença está nas elasticidades cruzadas, usualmente muito menores do que as verificadas no modelo duplo log e, no caso da carne bovina, com o sinal trocado (que deve-se ao coeficiente negativo e não significante para b_c.

Estas limitações do LES já eram conhecidas dos economistas no final dos anos 60, e um grande número de pesquisas foi dedicado a construir formas funcionais que permitissem estimar de uma forma menos restritiva as elasticidades preço e renda, permitindo que os padrões de dados observados nos informassem mais sobre estas características.

2.3 Translog

O paper de Christensen, Jorgenson e Lau (1975)[19] é uma ilustração do segundo tipo de sistema de demanda. A grande questão que eles tinham era de permitir que os dados informassem o máximo possível acerca dos padrões de substituição entre os produtos e sensibilidade à variações na renda, mas que ao mesmo tempo fosse consistente com o processo de maximização da utilidade do consumidor. A principal razão para isto é que os resultados tradicionais de microeconomia podem ser utilizados para a análise de bem-estar dos resultados de intervenções várias na economia.

Os autores do artigo começam postulando a seguinte forma funcional para a função de utilidade indireta, que os autores consideram como sendo "flexível"[6] – ou seja, para valores específicos dos parâmetros α e β, aproxima as derivadas primeira e segundas de qualquer forma funcional para a mesma função de utilidade inidreta:

$$\ln(v(\mathbf{P}, w)) = \alpha_0 + \sum \alpha_i \ln \frac{p_i}{w} + \frac{1}{2} \sum \sum \beta_{ij} \ln \frac{p_i}{w} \ln \frac{p_j}{w}$$

Após a aplicação da Identidade de Roy, chegamos às seguintes especificações para os coeficientes da função demanda na forma de participação no dispêndio de mercado:

$$\begin{aligned} s_i &= \frac{\alpha_i + \sum_k \beta_{ik} \ln \frac{p_k}{w}}{\alpha_M + \sum_k \beta_{Mk} \ln \frac{p_k}{w}} \\ \alpha_M &= \sum \alpha_k \\ \beta_{Mi} &= \sum \beta_{ki} \end{aligned}$$

A idéia de flexibilidade da forma funcional aqui colocada é que, em um ponto do espaço composto pelos preços e dispêndio, esta função é capaz de, por meio dos valores estimados dos parâmetros, replicar as quantidades demandadas, as elasticidades renda, preço e cruzadas.

Podemos entender o denominador destas frações como sendo a soma de todos os numeradores das equações. Uma normalização comum é que $\alpha_M = -1$. É importante notar que, neste caso, as participações de mercado também são uma generalização das funções utilidade derivadas da Cobb-Douglas, no sentido que, no caso de homogeneidade, em que $\sum_k \beta_{ki} = 0$ temos que a equação dos *shares* simplifica para:

$$s_i = \gamma_i - \sum_k \beta_{ki} \ln p_k$$

Neste caso, o modelo permite um efeito linear do logaritmo de cada um dos preços sobre as participações de mercado, sendo os $-\beta_{ki}$ são aproximações às elasticidades preço das participações de mercado. A versão completa, por outro lado, permite diferentes caminhos pelos quais as alterações nos preços relativos se tornam alterações nas participações de mercado.

Vamos calcular as elasticidades, e para isso iremos fazer a seguinte definição:

$$\begin{aligned} \mathbf{A_i} &= \alpha_i + \sum_k \beta_{ik} \ln \frac{p_k}{w} \\ \mathbf{B} &= \alpha_M + \sum \beta_{Mi} \ln \frac{p_j}{w} \end{aligned}$$

[6] Iremos detalhar o conceito de forma funcional flexível na parte referente à estimativa de custos e produção; no momento, é suficiente afirmar que uma forma funcional algébrica é entendida como flexível se ela consegue aproximar qualquer comportamento dos agentes econômicos de uma forma teoricamente consistente. O texto de Lau (1986)[34] é uma referência importante sobre o tema.

Com isto, podemos definir a quantidade demandada como sendo:

$$x_i = \frac{w}{p_i}\left[\frac{\mathbf{A_i}}{\mathbf{B_i}}\right]$$

Calculando a elasticidade cruzada[7]:

$$e_{ij} = \left[\frac{\beta_{ij}}{\mathbf{A_i}} - \frac{\beta_{Mj}}{\mathbf{B_j}}\right]$$

Calculando a elasticidade-preço[8]:

$$e_{ii} = \left[\frac{\beta_{ii}}{\mathbf{A_i}} - \frac{\beta_{Mi}}{\mathbf{B_i}}\right] - 1$$

[7] A demonstração é a seguinte:

$$\begin{aligned}
\frac{\partial x_i}{\partial p_j} &= \frac{w}{p_i}\left[\frac{\frac{\partial \mathbf{A}}{\partial p_j}\mathbf{B} - \frac{\partial \mathbf{B}}{\partial p_j}\mathbf{A}}{\mathbf{B}^2}\right] \\
\frac{\partial \mathbf{A}}{\partial p_j} &= \frac{\beta_{ji}}{p_j} \\
\frac{\partial \mathbf{B}}{\partial p_j} &= \frac{\beta_{Mj}}{p_j} \\
\frac{\partial x_i}{\partial p_j} &= \frac{w}{p_i p_j}\left[\frac{\beta_{ji}\mathbf{B} - \beta_{Mj}\mathbf{A}}{\mathbf{B}^2}\right] \\
\frac{\partial x_i}{\partial p_j} &= \frac{w}{p_i p_j}\frac{\mathbf{A}}{\mathbf{B}}\left[\frac{\beta_{ji}}{\mathbf{A}} - \frac{\beta_{Mj}}{\mathbf{B}}\right]
\end{aligned}$$

[8] Da mesma forma:

$$\begin{aligned}
\frac{\partial x_i}{\partial p_i} &= \frac{w}{p_i}\left[\frac{\frac{\partial \mathbf{A}}{\partial p_i}\mathbf{B} - \frac{\partial \mathbf{B}}{\partial p_i}\mathbf{A}}{\mathbf{B}^2}\right] - \frac{w}{p_i^2}\left[\frac{\mathbf{A}}{\mathbf{B}}\right] \\
&= \frac{w}{p_i^2}\left[\frac{\beta_{ji}\mathbf{B} - \beta_{Mj}\mathbf{A}}{\mathbf{B}^2}\right] - \frac{w}{p_i^2}\left[\frac{\mathbf{A}}{\mathbf{B}}\right] \\
&= \frac{w}{p_i^2}\frac{\mathbf{A}}{\mathbf{B}}\left[\frac{\beta_{ji}}{\mathbf{A}} - \frac{\beta_{Mj}}{\mathbf{B}}\right] - \frac{w}{p_i^2}\left[\frac{\mathbf{A}}{\mathbf{B}}\right]
\end{aligned}$$

Calculando a elasticidade-renda[9]:

$$e_w \quad = \quad \left[-\frac{\sum_k \beta_{ik}}{\mathbf{A_i}} + \frac{\sum_k \beta_{Mk}}{\mathbf{B_i}}\right] + 1$$

A principal vantagem deste sistema de demanda em relação ao LES, discutido acima é que, além de consistente com a maximização de utilidade, os valores dos coeficientes estimados é o que nos dá o padrão de substituição, bem como as características da curva renda-consumo. Além disso, como características importantes do padrão de comportamento dos consumidores são capturadas por meio dos parâmetros do modelo podemos, com o uso dos testes de hipótese tradicionais, testar se os valores estimados dos coeficientes são consistentes com os valores que os mesmos deveriam assumir para consistência com a racionalidade dos consumidores. Tais valores estão descritos na subseção a seguir.

2.3.1 Restrições dos Coeficientes

Como vimos anteriormente, para que os resultados da estimativa de um sistema de equações de demanda sejam teoricamente plausíveis, existem três propriedades que precisam estar presentes – a exaustão da restrição orçamentária (*adding-up*), simetria da matriz de derivadas da demanda hicksiana e homoteticidade.

A primeira das propriedades, a exaustão da restrição orçamentária, é conseguida se $\alpha_M = -1$e que os coeficientes β_{Mi} sejam iguais em todas as equações. Ou seja, $\beta^j_{Mi} = \beta_{Mi}, \forall j$. Por outro lado, para garantir a simetria da matriz de derivadas da função demanda hicksiana, temos que:

$$\beta_{ij} = \beta_{ji}$$

Para a homoteticidade, temos que os β_{Mi} tem que ser proporcionais aos α_i, da seguinte forma:

$$\beta_{Mi} = \sigma\alpha_i$$

[9]Demonstração:

$$\begin{aligned}
\frac{\partial x_i}{\partial w} &= \frac{w}{p_i}\left[\frac{\frac{\partial \mathbf{A}}{\partial w}\mathbf{B} - \frac{\partial \mathbf{B}}{\partial w}\mathbf{A}}{\mathbf{B}^2}\right] + \frac{1}{p_i}\left[\frac{\mathbf{A}}{\mathbf{B}}\right] \\
\frac{\partial \mathbf{A}}{\partial w} &= -\frac{\sum_j \beta_{ji}}{w} \\
\frac{\partial \mathbf{B}}{\partial w} &= -\frac{\sum_j \beta_{Mj}}{w} \\
\frac{\partial x_i}{\partial w} &= \frac{1}{p_i}\left[\frac{-\sum \beta_{ji}\mathbf{B} + \sum \beta_{Mj}\mathbf{A}}{\mathbf{B}^2}\right] + \frac{1}{p_i}\left[\frac{\mathbf{A}}{\mathbf{B}}\right] \\
\frac{\partial x_i}{\partial w} \quad w &= \left\{\frac{w}{p_i}\left[\frac{-\sum \beta_{ji}\mathbf{B} + \sum \beta_{Mj}\mathbf{A}}{\mathbf{B}^2}\right] + \frac{w}{p_i}\left[\frac{\mathbf{A}}{\mathbf{B}}\right]\right\}
\end{aligned}$$

Se supusermos homogeneidade – elasticidades-renda unitárias para cada produto, é a mesma coisa com $\sigma = 0$. Vimos na seção anterior que esta hipótese simplifica bastante a estimação dos coeficientes.

Exemplo 3: Com o mesmo arquivo EVIEWS.txt, foi estimado o modelo translog para os dados de carne bovina, frango, peixe e porco, sob a premissa de simetria e homogeneidade. A tabela a seguir mostra as elasticidades-preço e cruzadas (as elasticidades-dispêndio não são mostradas porque são, dada a premissa, iguais a um).

	Sensibilidade com relação ao preço de			
	p_c	p_f	p_p	p_po
Q_c	-1.1107	-0.0885	-0.0999	-0.0439
Q_f	-0.0956	-1.1062	-0.1191	-0.0496
Q_p	-0.1196	-0.1320	-1.1130	-0.0460

Podemos notar que os resultados acima são muito similares aos outros do ponto de vista de elasticidade-preço e algumas elasticidades-cruzadas, tanto com relação ao modelo duplo log quanto ao LES. No entanto, os sinais de algumas das elasticidades cruzadas têm sinais diferentes dos mostrados nos últimos modelos, o que talvez indique a inadequação das restrições impostas ao modelo.

O grande problema desta e da especificação anterior reside no fato que este é um sistema não linear nos coeficientes, o que dificulta bastante a estimação – em especial, na época em que isto foi proposto (final dos anos 70), quando o custo computacional da obtenção destas estimativas era proibitivamente alto. Um modelo que possui vantagens em relação aos anteriores neste sentido é o Sistema de Demanda Quase Ideal.

2.4 Sistema de Demanda Quase Ideal (Almost Ideal Demand System)

Este sistema – que, de acordo com os seus proponentes (Deaton e Muellbauer (1980)[22]), assim foi denominado devido ao fato de oferecer uma aproximação de primeira ordem a qualquer tipo de demanda – se baseia na seguinte função utilidade indireta:

$$v(\mathbf{P}, w) = G(\mathbf{P})[\ln w - \ln g(\mathbf{P})]$$

Sendo que a função $G(\mathbf{P})$ é homogênea de grau zero nos preços, e a $g(\mathbf{P})$ é homogênea de grau 1. A classe geral deste tipo de função de utilidade indireta é denominada PIGLOG ("Price Independent Generalized Linearity", PIGL em forma Logaritmica). Na verdade, esta condição se relaciona com a relação entre os preços relativos e a curva de Engel. Podemos recuperar a função dispêndio também – ao lembrarmos que o valor do dispêndio para um consumidor maximizador de utilidade é igual a w[10]:

$$\ln e(\mathbf{P}, u) = \frac{\ln u}{G(\mathbf{P})} - \ln g(\mathbf{P})$$

Esta é a função que se encontra no texto original de Deaton e Muellbauer (1980) [22] No caso específico da demanda AIDS, temos que[11]:

$$\begin{aligned} G(\mathbf{P}) &= \Pi_k p_k^{-\gamma_k} \\ g(\mathbf{P}) &= \alpha_0 + \sum \alpha_k \ln p_K + \frac{1}{2} \sum_k \sum_j \beta_{kj} \ln p_k \ln p_j \end{aligned}$$

A escolha destas funções específicas tinha por objetivo fornecer uma aproximação geral para os parâmetros de uma função utilidade. Aplicando a nossa amiga, a Identidade de Roy, nesta função de utilidade indireta e cozinhando vigorosamente, temos a seguinte forma para a equação demanda pelo produto na forma de *share de consumo*:

$$s_i = \alpha_i + \sum_i \beta_{ki} \ln p_k + \gamma_i \ln \left(\frac{w}{g(\mathbf{P})} \right)$$

Deaton e Muellbauer, no seu paper da *AER*, mencionam que uma alternativa quando os preços dos diferentes produtos são muito colineares, é a utilização do seguinte índice de preços de Stone (1953) no lugar da função $g(\mathbf{P})$:

$$\mathbf{P}^* = \sum_k \bar{s}_k \ln p_k$$

Em que $\bar{s}$ seria a média das participações de mercado. A partir do artigo original dos autores, surgiu uma linha de pesquisa que tinha por objetivo discutir a qualidade desta aproximação e os seus efeitos do ponto de vista das estimativas dos parâmetros. Por um lado, Buse (1994) [16] afirma que o uso do índice $\mathbf{P}^*$, por envolver uma média dos *shares* dos diferentes produtos, levaria a estimativas viesadas e não consistentes dos parâmetros se o sistema fosse estimado pelo método SUR. Além disso, o autor mostra que este problema não poderia ser resolvido pelo uso de métodos tradicionais de Variáveis Instrumentais. Buse e Chan (2000) [17] extendem este argumento para outros índices de preços – com os de Paasche e Laspeyres – no lugar de $\mathbf{P}^*$. Por outro lado, Asche e Wessells (1997) [3], mostram que, com os preços reescalonados como função dos valores em um ponto dos dados, os valores das elasticidades são iguais no modelo linearizado e no modelo completo. Independentemente desta discussão, a aproximação (conhecida por LA-AIDS)

[10]Pollak e Wales (1995)[39] demonstram que qualquer sistema de equações de demanda que seja linear na função dispêndio tem de ser gerado por uma função de utilidade indireta da forma PIGLOG.

[11]No paper original, a função $G(\mathbf{p})$ é igual à $(\log(b(p)) - \log(a(p)))/\beta_0$. A função $\ln g(\mathbf{p})$ é igual a $\log(a(p))$.

ganhou muitos adeptos, principalmente por que a estimação do sistema de equações envolve apenas equações lineares, o que facilita a implementação computacional.

As elasticidades preço e cruzadas do modelo são da seguinte forma, na versão LA-AIDS:

$$\begin{aligned} e_{ii} &= \frac{\beta_{ii} - \gamma_i s_i}{s_i} - 1 \\ e_{ij} &= \frac{\beta_{ij} - \gamma_i s_j}{s_i} \\ e_w &= 1 + \frac{\gamma_i}{s_i} \end{aligned}$$

Caso não seja adotada a linearização do índice de preços, as elasticidades-preço assumem uma forma um pouco mais complexa.

$$\begin{aligned} e_{ii} &= \frac{\beta_{ii} - \gamma_i s_i + \gamma_i^2 \ln\left(\frac{w}{g(\mathbf{P})}\right)}{s_i} - 1 \\ e_{ij} &= \frac{\beta_{ii} - \gamma_i s_i + \gamma_i \gamma_j \ln\left(\frac{w}{g(\mathbf{P})}\right)}{s_i} \end{aligned}$$

Da mesma forma que no modelo de demanda translog, a vantagem do AIDS é que as restrições necessárias para que os resultados da estimação de um destes modelos sejam consistentes com o processo de maximização da utilidade por parte do consumidor não são impostas pela forma funcional, mas podem ser testadas formalmente. Este teste formal toma a forma de testes sobre restrições aos parâmetros do sistema. Iremos detalhar quais são estas restrições na subseção seguinte.

2.4.1 Restrições dos Coeficientes

A primeira das restrições para que os resultados sejam consistentes com a maximização da utilidade do consumidor é o da simetria da matriz de Slutsky (derivadas preço e cruzadas das funções demanda hicksiana). Para isso, precisamos que os termos β cruzados sejam iguais, ou:

$$\beta_{ij} = \beta_{ji}$$

O segundo tipo de restrição que precisa ser atendido é o de Adding-up, que garante que os valores dos dispêndios previstos pelas equações do modelo sejam iguais ao dispêndio total. Esta premissa também permite que recuperarmos os coeficientes da última equação, mesmo ela não estimada pelo fato das participações no gasto necessariamente somarem 1. Estas restrições implicam:

$$\begin{aligned} \sum_i \alpha_i &= 1 \\ \sum_i \beta_{ij} &= 0 \\ \sum_i \gamma_i &= 0 \end{aligned}$$

Homogeneidade:

$$\sum_j \beta_{ij} = 0$$

Exemplo 4: Com o arquivo EVIEWS.txt, foram estimadas as elasticidades-preço e cruzada, assumindo simetria, homogeneidade e adding-up, tanto na versão linearizada quanto na versão não linear. Os resultados para o modelo linearizado com o uso do índice de preços de Stone estão a seguir:

	Sensibilidade com relação ao preço de			
	p_c	p_f	p_p	p_{po}
Q_c	-1.1163	0.0608	0.0373	0.0237
Q_f	0.1205	-0.9973	-0.0969	0.0172
Q_p	0.1732	-0.4771	-0.9135	-0.0618

Para o modelo completo, as elasticidades são:

	Sensibilidade com relação ao preço de		
	p_c	p_f	p_p
Q_c	-1.0611	0.0498	0.0496
Q_f	0.1161	-0.9905	-0.0987
Q_p	0.2795	-0.4748	-0.8945

Podemos notar, com esres resultados, que ainda que existam diferenças entre os valores das elasticidades -preço e cruzadas pelos dois tipos de modelo, tais diferenças no caso em tela não são tais que invalidem qualquer conclusão de natureza qualitativa acerca do padrão de substituição entre os produtos.

Os exemplos de aplicação dos vários métodos de estimação nos deram resultados bastante similares. Ainda que estes modelos sejam adequados – respeitando-se as limitações e hipóteses implícitas de cada um deles – para entender o processo de escolha em um contexto de teoria microeconômica neoclássica, existem algumas situações de demanda em que estes modelos não são adequados.

A primeira destas situações trata do lançamento de novos produtos. Caso o analista deseje elaborar uma previsão de qual seria a participação de mercado de um produto em particular antes do lançamento do mesmo, nenhum destes modelos seria capaz de fazê-lo. Isso se deve ao fato que as participações de mercado são funções dos efeitos cruzados e dos preços dos bens; ainda que se saiba o preço do bem a ser lançado, os efeitos cruzados somente podem ser estimados a partir dos dados – que ainda não existem.

Outra situação em que estes modelos podem não ser adequados para responder perguntas do interesse do econometrista é quando o número de produtos é grande em relação à base de dados. Podemos notar que a matriz de elasticidades preço e cruzada de um conjunto de N produtos tem dimensão $(N \times N)$, na ausência de restrições adicionais. Estes termos precisam ser estimados a partir dos dados e, em muitos contextos, o número de observações disponível para o econometrista é inferior ao número de parâmetros.

Usualmente, nestes casos, o analista toma uma de três alternativas possíveis, na suposição que dados adicionais não estejam disponíveis. A primeira delas é, com a ajuda das restrições aos coneficientes vindas da teoria microeconômica – simetria, homogeneidade e *adding-up* – reduzir o número de parâmetros estimáveis. A segunda delas é levar a cabo algum procedimento de agregação entre produtos ou sobre o processo de escolha, o que também implicaria em menos parâmetros estimáveis. Esta linha de ação será discutida mais adiante no capítulo 04.

Finalmente, o terceiro caminho tomado pelos econometristas envolve a utilização de modelos de escolha discreta, em que os consumidores derivam utilidade não das quantidades dos produtos disponíveis no mercado, mas sim das características dos produtos e características individuais idiosincráticas. Esta abordagem metodológica será detalhada no próximo capítulo.

2.5 Exercícios

1. XXXXX
2. YYYYYYYY
3. ZZZZZZ

Capítulo 3

Modelos de Escolha Discreta

Ainda que continuamos a analisar a escolha do consumidor, neste capítulo teremos um ponto de vista distinto. Diferentemente do que assumimos no capítulo anterior, em que o consumidor é capaz de fazer exatamente a escolha que dá a ele a maior utilidade, quando defronte a uma situação em que as alternativas são claramente discerníveis, aqui abriremos a porta para a possibilidade de a escolha do consumidor ter um componente probabilístico.

Além disso, aqui também serão exploradas situações em que a escolha do consumidor pode ser discreta – ou seja, ao invés de quantos quilos de carne de porco serão consumidos na próxima semana, como vimos no capítulo anterior, a decisão pode ser sobre se compra ou não um determinado produto. Neste caso, ao invés de postular um consumidor representativo e derivar especificações testáveis com no seu comportamento, como foi feito no capítulo anterior, uma linha de ação bastante promissora seria especificar uma função de utilidade para cada consumidor, com base probabilística.

Uma terceira faceta do processo de escolha do consumidor aqui postulado diz respeito ao que determina a escolha do consumidor além do seu componente probabilístico. Em decorrência do exposto no parágrafo anterior, além do componente probabilístico o determinante da escolha do consumidor são as características de cada alternativa, e não a quantidade de unidades consumidas de cada uma.

Tais características não implicam dificuldade do ponto de vista de aplicação econométrica; pelo contrário, com base em técnicas comuns em manuais de econometria[1] é possível recuperar os parâmetros da escolha do consumidor facilmente. Neste capítulo, iremos trabalhar os aspectos básicos do modelo de escolha discreta. Inicialmente, iremos detalhar quais são as diferenças entre a abordagem de escolha discreta e a abordagem dos sistemas de demanda de equações neoclássico, para a seguir discutirmos os aspectos da estimação propriamente ditos. Na segunda seção, iremos analisar a abordagem de Berry (1994)[10] para a utilização destes modelos de esccolha discreta

[1] Como por exemplo, Johnston e DiNardo (1997)[31] e Wooldridge (2009) [51]

para o caso em que temos apenas dados agregados dos produtos, e finalmente, na terceira seção serão discutidas as formas pelas quais podemos incorporar características não observadas dos consumidores e dos produtos em uma situação com dados agregados, como em Berry, Levinsohn e Pakes (1995)[9].

3.1 Modelos de Escolha Discreta.

Neste capítulo, iremos assumir que cada um dos consumidores possui uma função de utilidade específica, que atribui valores para cada um dos produtos disponíveis. No capítulo anterior, ao não assumirmos heterogeneidade entre os consumidores, implicitamente os resultados da estimação referiam-se a um "consumidor representativo", cujas preferências eram as que davam margem ao sistema de equações estimadas. É importante notar que, dependendo de como a função de utilidade específica ao consumidor é especificada, podemos definir os resultados da estimação mesmo em um modelo de escolha discreta em termos de um consumidor representativo[2].

Os modelos de escolha discreta tem por razão final entender a escolha do consumidor entre alternativas mutuamente excludentes, com base na premissa que o consumidor escolherá a que dará a ele a maior satisfação possível, com base nas características da mesma.

Para que possamos recuperar as características da estrutura de escolha do consumidor, precisamos adotar o ponto de vista do econometrista. A utilidade do consumidor não é observável, mas apenas alguns atributos dos produtos e características do tomador de decisão – sendo que muito provavelmente não todas – além das escolhas do mesmo. Assumindo a premissa de escolha discreta do consumidor, o primeiro passo é postular uma função que relaciona estes dados observados com a sua escolha, que chamaremos de $V_i(\mathbf{x_j}, \mathbf{s_j}) = V_{ij}$, sendo que $\mathbf{x_j}$ é um vetor que representa as características observadas do produto por parte do econometrista e $\mathbf{s_j}$o vetor de características não observadas.

Uma vez que alguns aspectos da utilidade do consumidor[3] não são observados, em geral $V \neq U$, em que U é a "verdadeira" utilidade do consumidor. Desta forma, podemos fazer o seguinte ajuste:

$$U_{ij} = V_{ij} + \epsilon_{ij}$$

Em que i denota o consumidor e j a alternativa. De um ponto de vista prático, o termo ϵ_{ij} captura os aspectos do produto ou do indivíduo que não são observados pelo econometrista, representado por $\mathbf{s_j}$ além do componente inerentemente aleatório desta escolha enfatizado na introdução deste capítulo. As características deste termo dependerão de como o econometrista especifica V_{ij}.

Assim, o passo seguinte é atribuir uma distribuição para este termo que reúne as características não-observáveis. A partir desta distribuição, podemos derivar a probabilidade de escolha de cada

[2]Anderson, de Palma e Thisse (1992)[1] mostram esta equivalência do ponto de vista teórico.

[3]Independentemente das características do indivíduo e do produto.

uma das alternativas, que é dada pela massa de probabilidade associada com a hipótese de que a alternativa gera mais utilidade do que qualquer outra.

Dada esta distribuição para os componentes não observáveis do modelo, $\varepsilon =< \epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{iN} >$, temos:

$$\begin{aligned}
P_{ij} &= Prob(U_{ij} > U_{ik}, \forall k \neq j) \\
&= Prob(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, \forall k \neq j) \\
&= Prob(V_{ij} - V_{ik} > \epsilon_{ik} - \epsilon_{ij}, \forall k \neq j) \\
&= Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j) \\
&= Prob(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j)
\end{aligned}$$

A penúltima igualdade é uma função de distribuição acumulada, que responde a massa de probabilidade de cada um dos $J-1$ produtos ter a diferença entre os componentes idiosincráticos – $\epsilon_{ik} - \epsilon_{ij}, \forall k \neq j$ inferior que a diferença entre as partes determinísticas entre o $J-$ésimo produto e os outros – representado por $V_{ij} - V_{ik}$. Podemos calcular esta probabilidade, usando a distribuição conjunta dos ε, por meio da seguinte integral multidimensional:

$$P_{ij} = \int_{\varepsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j) f(\varepsilon) d(\varepsilon)$$

Nesta integral, a função $I(\cdot)$ representa a função indicadora, dando o valor de um caso a desigualdade dentro do argumento seja verdadeira e zero, caso contrário. $f(\varepsilon)$ é a função densidade de probabilidade de ε, e dá a probabilidade associada com cada um dos pontos desta distribuição.

Diferentes especificações de modelos de escolha discreta surgem em resposta a diferentes especificações da variável aleatória multidimensional ε. Por exemplo, se ε for uma distribuição $N(0, \Omega)$, isso nos dá o modelo probit multinomial[4].Se ε seguir uma distribuição de valores extremos I ou Gumbel:

$$\begin{aligned}
f(\epsilon_{ij}) &= e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} \\
F(\epsilon_{ij}) &= e^{e^{-\epsilon_{ij}}}
\end{aligned}$$

Temos o modelo LOGIT Multinomial. É importante notar que, para o caso deste tipo de distribuição, a integral multidimensional que fizemos anteriormente pode ser resolvida analiticamente. O primeiro passo para entendermos isso é uma regrinhaque diz que as diferenças entre duas

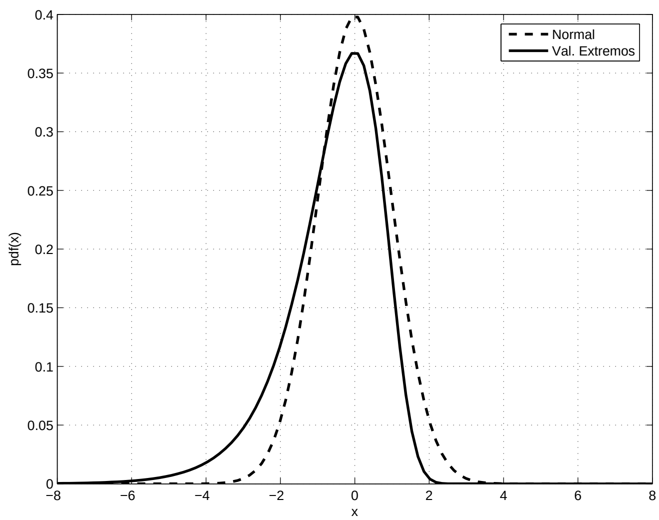
[4]Usualmente, esta integral sobre a distribuição normal multivariada é bastante difícil de ser calculada de um ponto de vista computacional. Como veremos mais adiante, por um lado existem formas de cálculo desta integral por métodos computacionais (simulação) e por outro existem distribuições que permitem a obtenção de uma forma analítica para estas probabilidades.

variáveis aleatórias que seguem esta distribuição são distribuídas de acordo com uma distribuição logística:

$$\begin{aligned} \epsilon_{ij}^* &= \epsilon_{ik} - \epsilon_{ij} \\ F(\epsilon_{ij}^*) &= \frac{\epsilon_{ij}^*}{1 + \epsilon_{ij}^*} \end{aligned}$$

A figura a seguir mostra a diferença entre as funções densidade de probabilidade de Valores Extremos I e a Normal padronizada:

Figura 3.1: Valores Extremos e Normal Padrão



Uma segunda premissa, importante para a derivação simples destas probabilidades de escolha é que os componntes idiosingráticos das utilidades – os ϵ_{ij}– são independentes e identicamente distribuídos[5]. A segunda parada é que os componentes idiosincráticos das utilidades são i.i.d.; mas antes, vamos reescrever a última das probabilidades antes da integral da seguinte forma:

$$P_{ij} = \int_{\varepsilon} I(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j) f(\varepsilon) d\varepsilon$$

[5] Para compreender melhor esta premissa, imagine que não existam características não observadas tanto em termos das pessoas quanto das alternativas. Isso significa que, no momento da escolha, quando enfrentando as alternativas, além das características das mesmas o consumidor retira algumas bolinhas de uma urna – que seriam os ϵ_{ij}. A idéia da independência aqui significa que, se para um produto o ϵ_{ij} é alto, isso não é informativo sobre que bolinha o consumidor irá tirar para qualquer outro produto.

Se o ϵ_{ij} é considerado como dado, esta função nos dá a função de distribuição acumulada para cada ϵ_{ik} avaliada em $\epsilon_{ij} + V_{ij} - V_{ik}$, o que, de acordo com a distribuição valores extremos I é igual a $\exp[-\exp[-(\epsilon_{ij} + V_{ij} - V_{ik})]]$. Como os elementos do vetor ε são independentes, isto significa que esta probabilidade conjunta – afinal de contas, vale para todos os elementos de ε exceto j – é igual a um produto das distribuições individuais:

$$P_{ij}|\epsilon_{ij} = \prod_{k \neq j} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}}$$

Evidentemente, ϵ_{ij} não é dado, desta forma a probabilidade conjunta é a integral desta equação com respeito a todos os valores de ϵ_{ij}:

$$P_{ij} = \int_{\epsilon_{ij}} \left(\prod_{k \neq j} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} \right) e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} d\epsilon_{ij}$$

Vamos cozinhar um pouco esta equação; lembrando que, para o produto j , $V_{ij} - V_{ij} = 0$, temos que a integral acima pode ser reconstruída da seguinte forma:

$$P_{ij} = \int_{\epsilon_{ij}} \left(\prod_{k} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} \right) e^{-\epsilon_{ij}} d\epsilon_{ij}$$

Podemos transformar este produtório em soma, uma vez que as bases são iguais:

$$\begin{aligned} P_{ij} &= \int_{\epsilon_{ij}} \exp\left(-\sum_{k} e^{-(\epsilon_{ij}+V_{ij}-V_{ik})} \right) e^{-\epsilon_{ij}} d\epsilon_{ij} \\ &= \int_{\epsilon_{ij}=-\infty}^{+\infty} \exp\left(-e^{-\epsilon_{ij}} \sum_{k} e^{-(V_{ij}-V_{ik})} \right) e^{-\epsilon_{ij}} d\epsilon_{ij} \end{aligned}$$

Redefinindo as variáveis de integração, tal que $e^{-\epsilon_{ij}} = t$, tal que $dt = -e^{-\epsilon_{ij}} d\epsilon_{ij}$. Note que, quando $\epsilon_{ij} \to \infty$, $t \to 0$, e quando $\epsilon_{ij} \to -\infty$, $t \to \infty$, o que faz com que os limites de integração agora sejam 0 e ∞. Usando este novo termo:

$$\begin{aligned} P_{ij} &= \int_{t=\infty}^{0} \exp\left(-t \sum_{k} e^{-(V_{ij}-V_{ik})} \right) (-dt) \\ &= \int_{t=0}^{\infty} \exp\left(-t \sum_{k} e^{-(V_{ij}-V_{ik})} \right) dt \\ &= \left. \frac{\exp\left(-t \sum_k e^{-(V_{ij}-V_{ik})}\right)}{\sum_k e^{-(V_{ij}-V_{ik})}} \right|_{0}^{\infty} \\ &= \frac{1}{\sum_k e^{-(V_{ij}-V_{ik})}} = \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}} \end{aligned}$$

A passagem da primeira para a segunda equação simplesmente implica em uma troca de posição dos limites de integração. Ao final, temos as probabilidades do modelo Logit Multinomial.

O passo seguinte é incorporar estas probabilidades de escolha em um contexto de Máxima Verossimilhança para que possamos recuperar os parâmetros relevantes da função V_{ij}. Mas antes, uma questão de identificação dos coeficientes precisa ser avaliada.

Identificação dos Coeficientes

Podemos resumir os cuidados que temos na estimação dos modelos de escolha discreta em dias afirmações. A primeira é que "apenas diferenças de utilidade são importantes", e a segunda é que "a escala da utilidade é arbitrária". Iremos falar brevemente somente da primeira das características pois, em geral, a normalização da escala da utilidade está implícita na modelagem LOGIT.

Com relação à primeira das escolhas, temos que o nível absoluto da utilidade auferida pelos consumidores em cada uma das alternativas é irrelevante para entendermos o comportamento dos mesmos. Por exemplo, se multiplicássemos a função U por um múltiplo qualquer, teríamos então as mesmas escolhas. Isso tem algumas implicações importantes. A primeira delas é que a imposição de constantes na parte observável da utilidade do consumidor implica que teremos constantes associadas com $(N-1)$ alternativas, uma vez que senão a estimação destes modelos fica indeterminada.

Outro ponto importante é que variáveis socio-demográficas, que afetam uniformemente as alternativas, não podem ser incluídas no lado determinista da função utilidade em todas as alternativas. Pelo menos um dos coeficientes associado com uma das alternativas precisa ser normalizado em zero. Isso não ocorre quando temos esta variável entrando como interação com as características específicas das alternativas.

Da mesma forma que somando uma constante a todas as alternativas não afeta a escolha do consumidor, não há alteração na escolha do consumidor se multiplicarmos a parte determinística da utilidade do mesmo por uma constante. Ou seja, se temos um modelo $U_{ij} = V_{ij} + \epsilon_{ij}$ como sendo válido, qualquer multiplicação de U_{ij} por uma constante também se constitui em um modelo válido, o que levaria a problemas de identificação dos parâmetros do nosso modelo.

Para isso, o melhor a fazer é normalizar a escala da utilidade, o que passa por padronizar a variância dos termos ϵ_{ij}. No caso em que temos o modelo logit tradicional, em que os ϵ_{ij} são independentes entre alternativas, uma opção é dividir os coeficientes por um fator igual a $\sqrt{\pi^2/6}$, em que o argumento da raiz quadrada é aproximadamente igual a 1,6. Caso não façamos esta alteração na estimação, acabamos por obter coeficientes maiores do que deveriam.

Estimação

Em geral, os procedimentos de estimação do modelo Logit Multinomial está baseado no princípio da Máxima Verossimilhança. Inicialmente, vamos supor que a amostra seja aleatória, e que tenhamos dados sobre N tomadores de decisão. A probabilidade de um indivíduo i escolher a alternativa que ele efetivamente escolheu é igual a:

$$\prod_{j \in J} (P_{ij})^{y_{ij}}$$

Em que é uma variável *dummy* supondo $y_{ij} = 1$ se o indivíduo i escolheu o produto e $y_{ij} = 0$, caso contrário. Na prática, isto significa o valor da probabilidade do produto que ele escolheu. Supondo independência das escolhas dos indivíduos, a probabilidade de observação de uma amostra igual à que temos é:

$$L(\beta) = \prod_{i \in N} \prod_{j \in J} (P_{ij})^{y_{ij}}$$

Denominamos esta probabilidade por $L(\beta)$ para representar o fato que esta probabilidade é função dos parâmetros do modelo. Em geral, os algoritmos numéricos maximizam o logaritmo neperiano desta probabilidade conjunta, o que dá:

$$\ln(L(\beta)) = LL(\beta) = \sum_{i \in N} \sum_{j \in J} y_{ij} \ln P_{ij}$$

McFadden (1974) provou que, para V_{ij} linear nos parâmetros, esta função objetivo é côncava, o que garante que os algoritmos numéricos tradicionais convirjam rapidamente para um mínimo. Em geral, também podemos dar uma interpretação de GMM ao método de estimação utilizado da seguinte forma. O vetor de parâmetros que minimiza esta função deve atender à seguinte condição de primeira ordem:

$$\frac{\partial LL(\beta)}{\partial \beta} = 0$$

Para facilitar, vamos supor que $V_{ij} = x_{ij}\beta$. Neste caso, temos o seguinte conjunto de condições de primeira ordem com respeito aos parâmetros a serem estimados:

$$\sum_{i \in N} \sum_{j \in J} (y_{ij} - P_{ij}) x_{ij} = 0$$

O que vai dentro de um parênteses seria análogo ao resíduo da regressão – a diferença entre o valor observado (que são as escolhas e estão caputradas na variável y_{ij}) e o valor previsto, que são as probabilidades de escolha. O x_{ij} que o está multiplicando seriam os regressores do modelo. Ou seja, aqui temos uma ilustração de uma característica bastante interessante: podemos pensar que estimativas por Máxima Verossimilhança são também estimativas de GMM, só que considerando as derivadas – também conhecidas como *scores* – como condições de momento.

É importante notar quem diferentemente do modelo de Regressão Linear tradicional, os efeitos marginais e/ou as elasticidades não são diretamente estimáveis como parâmetros, mas são funções deles. As fórmulas precisas para a obtenção destas informações a partir dos parâmetros estimados estão expostas na seção seguinte..

Derivadas e Elasticidades

Um dos elementos importantes da análise envolve o cálculo das elasticidades e derivadas das probabilidades de escolha com respeito aos atributos. A mudança na probabilidade de escolha da alternativa j por parte do indivíduo i em resposta a alterações em um atributo, por exemplo, x_{ij}é dada por:

$$\begin{aligned}
\frac{\partial P_{ij}}{\partial x_{ij}} &= \frac{\partial(e^{Vij}/\sum_{k\in J} e^{V_{ik}})}{\partial x_{ij}} \\
&= \frac{e^{V_{ij}}}{\sum_{k\in J} e^{V_{ik}}}\frac{\partial V_{ij}}{\partial x_{ij}} - \frac{e^{V_{ij}}}{(\sum_{k\in J} e^{V_{ik}})^2}\frac{\partial V_{ij}}{\partial x_{ij}}e^{V_{ij}} \\
&= \frac{\partial V_{ij}}{\partial x_{ij}}(P_{ij} - P_{ij}^2) \\
&= \frac{\partial V_{ij}}{\partial x_{ij}}P_{ij}(1 - P_{ij})
\end{aligned}$$

Um outro ponto importante diz respeito a como a probabilidade da escolha de uma alternativa j se altera em resposta a alterações nos atributos de outra alternativa, por exemplo, k. Podemos derivar este efeito cruzado da seguinte forma:

$$\begin{aligned}
\frac{\partial P_{ij}}{\partial x_{ik}} &= \frac{\partial(e^{Vij}/\sum_{c\in J} e^{V_{ic}})}{\partial x_{ik}} \\
&= -\frac{e^{V_{ij}}}{(\sum_{c\in J} e^{V_{ic}})^2}\frac{\partial V_{ik}}{\partial x_{ik}}e^{V_{ik}} \\
&= -\frac{\partial V_{ik}}{\partial x_{ik}}P_{ij}P_{ik}
\end{aligned}$$

Aqui surge um ponto importante. Os efeitos cruzados entre os diferentes produtos são função apenas das probabilidades de escolha dos dois produtos, multiplicada pelo efeito marginal da alteração do atributo na utilidade observável do produto k. Este fenômeno, em que os padrões de substituição dependem apenas das probabilidades de escolha dos dois produtos, é chamado de

Independência de Alternativas Irrelevantes e quer dizer que uma melhora em uma opção atrai escolhas de outras alternativas na mesma proporção. Em outras palavras; uma melhora de 10% em uma alternativa quer dizer que as outras probabilidades de escolha se reduzirão em 10%. Isto é claramente irrealista, e veremos mais adiante formas pelas quais podemos modelar o padrão de substituição entre as alternativas.

Passemos então às elasticidades. A elasticidade da probabilidade de escolha P_{ij} com respeito a alterações nos atributos do mesmo produto é igual a:

$$\begin{aligned} E_{ix_{ij}} &= \frac{\partial P_{ij}}{\partial x_{ij}} \frac{x_{ij}}{P_{ij}} \\ &= \frac{\partial V_{ij}}{\partial x_{ij}} P_{ij}(1 - P_{ij}) \frac{x_{ij}}{P_{ij}} \\ &= \frac{\partial V_{ij}}{\partial x_{ij}} (1 - P_{ij}) x_{ij} \end{aligned}$$

A elasticidade da probabilidade de escolha P_{ij} com respeito a alterações nos atributos dos outros produtos é igual a:

$$\begin{aligned} E_{ix_{ik}} &= \frac{\partial P_{ij}}{\partial x_{ik}} \frac{x_{ik}}{P_{ij}} \\ &= -\frac{\partial V_{ij}}{\partial x_{ik}} P_{ij} P_{ik} \frac{x_{ik}}{P_{ij}} \\ &= -\frac{\partial V_{ij}}{\partial x_{ij}} P_{ik} x_{ik} \end{aligned}$$

Exemplo. No arquivo `refri.txt` encontram-se 20.616 observações sobre a escolha de refrigerantes extraídas da PNAD 1995/96. Na variável select, encontra-se o indicador da escolha de um dos diferentes tipos de refrigerante, enquanto na variável preco encontra-se o preço por litro da bebida escolhida.

Código	Marca	Código	Marca
1	COCA COLA	19	GASOSA
2	PEPSI	20	REFRIGERANTES - NÃO ESPECIFICADO
3	GUARANA	21	LARANJA (EXCETO FANTA, SUKITA, POP, CRUSH)
4	FANTA LARANJA, UVA, LIMÃO	22	COLA (EXCETO COCA-COLA E PEPSI-COLA)
5	CRUSH	23	MAÇÃ (QUALQUER MARCA)
6	MINEIRINHO	24	TUTI FRUTI (QUALQUER MARCA)
7	SODA LIMONADA	25	TANGERINA (QUALQUER MARCA)
8	MIRINDA	26	CAJÚ (QUALQUER MARCA)
9	MINUANO	27	AGUA TÔNICA
10	AGUA MINERAL	28	GRAPETE
11	MATE COURO	29	PARAGUAI
12	BIDÚ	30	GOIANINHA
13	JÃO	31	REAL
14	ALOA	32	FRATELLY VITA
15	SUKITA	33	DIETÉTICO
16	POP LARANJA	34	GATORATE
17	XODÓ DA BAHIA	35	SNAPPLE
18	TUBAINA	0	Outras Bebidas

Foi estimado um modelo de escolha discreta em que a utilidade indireta condicional da escolha do refrigerante j é dada por:

$$u_j = \beta_j - \alpha p_j + \varepsilon_j$$

O valor estimado para o parâmetro α é igual a 2,24233, e algumas das elasticidades preço e cruzadas estão expostas na tabela a seguir:

	COCA COLA	PEPSI	GUARANA
COCA COLA	-1.1893	0.6679	0.0696
PEPSI	1.3507	-1.7849	0.0696
GUARANA	1.3507	0.6679	-2.3961
FANTA LARANJA, UVA, LIMÃO	1.3507	0.6679	0.0696

Como podemos notar, os efeitos cruzados do modelo logit dependem apenas das características dos produtos envolvidos. Isso faz com que nem sempre os padrões de substituição sejam plausíveis. Na seção seguinte, iremos discutir uma metodologia que não depende desta premissa e, portanto, pode nos gerar resultados mais interessantes.

3.1.1 LOGIT Aninhado

Como vimos, um dos principais fatores que motivam a escolha do modelo Logit Multinomial para a a estimação de parâmetros estruturais do lado da demanda é a sua simplicidade computacional. No entanto, uma das principais críticas que podem ser levantadas a este modelo está relacionada com a hipótese de independência de alternativas irrelevantes (*Independence of Irrelevant Alternatives*-IIA). A razão pela qual esta hipótese deve ser criticada é que ela implica uma competição igual entre os diferentes tipos de alternativas, o que claramente não é verdade em todos os tipos de situação.

O exemplo clássico para isto é o chamado exemplo do ônibus vermelho/ônibus azul. Suponha que existam dois tipos de transporte, carro e um ônibus pintado de azul, e que as pessoas escolhem seu modo de transporte de acordo com o modelo Logit. Ou seja, a utilidade das duas alternativas é dada por:

$$\begin{aligned} U_c &= V_c + \epsilon_c \\ U_{bb} &= V_{bb} + \epsilon_{bb} \end{aligned}$$

Em que bb representa as variáveis relevantes à escolha do ônibus azul e c as mesmas variáveis para a escolha do carro. Por simplicidade, vamos supor $V_{bb} = V_c = 0$, o que implica uma probabilidade de escolha das duas alternativas igual a 0,5.

Agora, imagine que uma alternativa adicional se torne disponível – um outro tipo de ônibus, exatamente igual ao primeiro, só que pintado de vermelho. Uma vez que ele é absolutamente igual ao outro, devemos supor que $V_v = 0$, o que dá uma probabilidade de 1/3 para cada alternativa. Isso fica ainda mais complexo ao analisarmos o padrão de substituição implícito – de onde vêm este 1/3 dos usuários para a nova alternativa. Por hipótese, temos que a razão entre a probabilidade de escolha do carro e do ônibus azul é $e^{V_c}/e^{V_{bb}} = 1$, e isso será válido antes e depois da nova alternativa. Estas duas condições – um terço dos usuários na nova alternaiva e as probabilidades de escolha das duas outras iguais antes e depois – implicam que cada alternativa, carro e ônibus azul, contribuem cada uma com 1/6 de probabilidade para a alternativa do ônibus vermelho. Ainda que o argumento seja correto do ponto de vista algébrico, é difícil justificar este padrão de substituição de um ponto e vista econômico. Mesmo que alguns consumidores da alternativa automóvel prefiram o ônibus vermelho, difilmente será na mesma proporção que os usuários do ônibus azul.

Esta limitação decorre do fato que os termos ϵ_{ij} são independentes entre si, hipótese utilizada na derivação das probabilidades de escolha do modelo. Podemos relaxar esta hipótese, o que dá margem a diferentes modelos. Um que retém parte da facilidade computacional do Logit Multinomial, reduzindo o problema da IIA é o chamado Logit Aninhado (*Nested Logit*). A característica principal deste modelo é que agrupa as diferentes alternativas em grupos que são mais similares entre si do que aquelas que estão fora dos respectivos grupos.

A derivação do modelo se baseia na premissa que algumas alternativas compartilham alguns componentes entre si nos seus termos representativos da parte aleatória da escolha do consumidor.

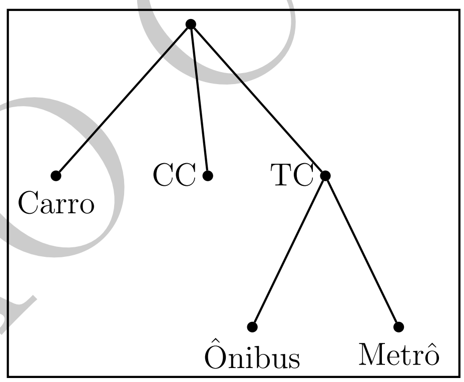
Para as derivações das probabilidades de escolha, fundamentais para a implementação do estimador usando o princípio da Máxima Verossimilhança, vamos supor um exemplo em que o consumidor tem de enfrentar quatro alternativas de transporte – o uso de um carro sozinho, o uso compartilhado de um carro, o uso de ônibus e o uso de metrô. Adicionalmente, supõe-se que duas destas alternativas - ônibus e metrô - são mais próximas entre si aos olhos do consumidor, que permite que as agrupemos em um "ninho" chamado Transporte Coletivo.

$$\begin{aligned} U_C &= V_C + \epsilon_C \\ U_{CC} &= V_{CC} + \epsilon_{CC} \\ U_O &= V_{TC} + V_O + \epsilon_{TC} + \epsilon_O \\ U_M &= V_{TC} + V_M + \epsilon_{TC} + \epsilon_M \end{aligned}$$

Podemos notar que os termos representativos da utilidade de ônibus e metrô possuem um termo comum, V_{TC} que seria a parte determinística da utilidade de se utilizar o transporte coletivo. Além disso, a parte aletória da utilidade seria o fruto de dois componentes – a parte aleatória de se usar transporte coletivo e a parte aleatória de se utilizar o modal específico.

Podemos racionalizar esta estrutura de escolha como uma escolha em dois estágios – ainda que, conceitualmente, a idéia de escolha em dois estágios aqui se baseie em elementos muito diferentes dos discutidos na parte de modelos neoclássicos de demanda.

Figura 3.2: Árvore de Escolha



Podemos derivar as probabilidades de escolha das diferentes alternativas em duas partes. Inicialmente, condicional à escolha de Transporte Coletivo, temos as seguintes probabilidades de escolha:

$$P_{O|TC} = \frac{e^{V_O/\sigma}}{e^{V_O/\sigma} + e^{V_M/\sigma}}$$
$$P_{M|TC} = \frac{e^{V_M/\sigma}}{e^{V_O/\sigma} + e^{V_M/\sigma}}$$

No nível superior, temos três opções – carro, carro compartilhado ou transporte coletivo. A probabilidade de escolha neste nível superior é dada por:

$$P_C = \frac{e^{V_C}}{e^{V_C} + e^{V_{CC}} + e^{V_{TC}+\sigma I_{TC}}}$$
$$P_{CC} = \frac{e^{V_{CC}}}{e^{V_C} + e^{V_{CC}} + e^{V_{TC}+\sigma I_{TC}}}$$
$$P_{TC} = \frac{e^{V_{TC}+\sigma I_{TC}}}{e^{V_C} + e^{V_{CC}} + e^{V_{TC}+\sigma I_{TC}}}$$

A última fórmula merece uma análise mais pormenorizada. Podemos notar que o numerador desta função possui um termo diferente dos logit tradicional, que é composto por dois termos. O primeiro deles é apenas a parte determinística da utilidade de se pegar transporte coletivo, enquanto o termo I_{TC} representaria o valor experado do máximo de utilidade das alternativas dentro do ninho. Em especial, $I_{TC} = \log[\exp\left(\frac{V_O}{\sigma}\right) + \exp\left(\frac{V_M}{\sigma}\right)]$. Podemos, a partir dos resultados acima, construir as probabilidades de escolha – não condicionais – das alternativas dentro do ninho "transporte coletivo":

$$P_O = P_{O|TC} \times P_{TC}$$
$$P_M = P_{M|TC} \times P_{TC}$$

Vamos detalhar mais o que significa o parâmetro σ. O papel do parâmetro σ é mais importante, pois significa que os parâmetros dentro da função V de todas as alternativas dentro de um mesmo ninho são padronizadas por um mesmo valor. Uma vez que, por definição, σ é limitado entre zero e um, estes coeficientes são aumentados, aumentando assim a sensibilidade a alterações nos atributos da alternativa dentro do ninho. Este parâmetro é uma função da correlação subjacente entre a parte aleatória da utilidade para os pares de alternativas dentro de um mesmo ninho, caracterizando o grau de substitutibilidade entre os parâmetros. Caso $\sigma = 0$, temos que existe perfeita correlação entre as alternativas dentro de um mesmo ninho, o que implica que a escolha deixa de ser estocástica e passa a ser determinística. Ou seja, como a correlação é perfeita, a probabilidade que a diferença entre os termos ϵ seja superior à diferença entre as partes determinísticas da utilidade do consumidor ou é zero ou é um. Caso $\sigma = 1$, temos que o modelo Logit Aninhado colapsa no modelo Logit Multinomial tradicional, pois os termos ϵ entre as diferentes alternativas deixam de ser correlacionados entre si.

Da mesma forma que no modelo Logit tradicional, os efeitos marginais e as elasticidades não são estimados diretamente como parâmetros, mas sim fórmulas calculadas a partir dos parâmetros. Devido à estrutura das probabilidades de escolha, elas são um pouco diferentes das do Logit tradicional e estão detalhadas a seguir.

Efeitos Marginais e Elasticidades

A derivação dos efeitos marginais pode ser realizada lembrando que a probabilidade de escolha de uma alternativa em particular pode ser decomposta como o produto de duas outras probabilidades – a probabilidade de escolha do ninho em que o produto se encontra, e a probabilidade de escolha do produto, condicional à escolha do ninho, como vimos anteriormente:

$$P_j = P_{j|K} \times P_K$$

Desta forma, podemos reescrever os efeitos marginais da alteração de um atributo x sobre a probabilidade de escolha:

$$\frac{\partial P_j}{\partial x_r} = \frac{\partial P_{j|K}}{\partial x_r} P_K + \frac{\partial P_K}{\partial x_r} P_{j|K}$$

Vamos generalizar as fórmulas da seção anterior da seguinte forma:

$$\begin{aligned}
P_{j|K} &= \frac{e^{V_j/\sigma}}{\sum_{l \in K} e^{V_l/\sigma}} \\
P_K &= \frac{e^{I_K \times \sigma}}{\sum_{L \in \mathbf{K}} e^{I_L \times \sigma}} \\
I_K &= \ln\left[\sum_{l \in K} e^{V_l/\sigma}\right]
\end{aligned}$$

Isto implica que teremos resultados diferentes se r for o mesmo produto j, se r e j forem dois produtos diferentes em um mesmo ninho, ou se r e j forem de ninhos diferentes.

Mesmo Produto

Se $r = j$, temos as seguintes derivações. Em primeiro lugar, para a probabilidade de escolha condicional ao grupo:

$$\begin{aligned}
\frac{\partial P_{j|K}}{\partial x_j} &= \frac{\partial V_j}{\partial x_j} \times \frac{1}{\sigma} \times \frac{e^{V_j/\sigma} \times \sum_{l \in K} e^{V_l/\sigma} - \left(e^{V_j/\sigma}\right)^2}{\left(\sum_{l \in K} e^{V_l/\sigma}\right)^2} \\
&= \frac{\partial V_j}{\partial x_j} \frac{1}{\sigma} P_{j|K}(1 - P_{j|K})
\end{aligned}$$

Para a probabilidade de escolha do ninho, temos:

$$\begin{aligned}
\frac{\partial P_K}{\partial x_j} &= \sigma \frac{\partial I_K}{\partial x_j} \frac{e^{I_K \sigma} \sum_{L \in \mathbf{K}} e^{I_L \sigma} - \left(e^{I_K \sigma}\right)^2}{\left(\sum_{L \in \mathbf{K}} e^{I_L \sigma}\right)^2} \\
&= \sigma \frac{\partial I_K}{\partial x_j} P_K(1 - P_K)
\end{aligned}$$

Finalmente, para I_K:

$$\begin{aligned}
\frac{\partial I_K}{\partial x_j} &= \frac{\partial V_j}{\partial x_j} \frac{1}{\sigma} \frac{e^{V_j/\sigma}}{\sum_{l \in K} e^{V_l/\sigma}} \\
&= \frac{1}{\sigma} \frac{\partial V_{ij}}{\partial x_j} P_{j|K}
\end{aligned}$$

Combinando as últimas duas equações, temos:

$$\begin{aligned}
\frac{\partial P_K}{\partial x_j} &= \frac{\partial V_j}{\partial x_j} P_{j|K} P_K (1 - P_K) \\
&= \frac{\partial V_j}{\partial x_j} P_j (1 - P_K)
\end{aligned}$$

E, finalmente, temos:

$$\begin{aligned}
\frac{\partial P_j}{\partial x_j} &= \frac{\partial V_j}{\partial x_j} \left[\frac{1}{\sigma} P_{j|K}(1 - P_{j|K}) P_K + P_j(1 - P_K) P_{j|K}\right] \\
&= \frac{\partial V_j}{\partial x_j} \frac{P_j}{\sigma} \left[1 - P_{j|K} + \sigma P_{j|K} - \sigma P_j\right] \\
&= \frac{\partial V_{ij}}{\partial x_j} \frac{P_j}{\sigma} \left[1 - P_{j|K}(1 - \sigma) - \sigma P_j\right]
\end{aligned}$$

Podemos notar que esta fórmula reduz-se à tradicional Logit caso $\sigma \rightarrow 1$. A elasticidade pode ser definida da seguinte forma:

$$E_{j,x_j} = \frac{\partial V_{ij}}{\partial x_j}\frac{x_j}{\sigma}\left[1 - P_{j|K}(1-\sigma) - \sigma P_{ij}\right]$$

Produtos Diferentes no mesmo ninho

Se $r \neq j$, mas os dois pertencem ao mesmo ninho, algumas derivadas parciais mudam sua forma de cálculo:

$$\begin{aligned}
\frac{\partial P_{j|K}}{\partial x_r} &= -\frac{\partial V_r}{\partial x_r}\frac{1}{\sigma}\frac{e^{V_j/\sigma}e^{V_r/\sigma}}{\left(\sum_{l\in K} e^{V_l/\sigma}\right)^2} \\
&= -\frac{\partial V_r}{\partial x_r}\frac{1}{\sigma}P_{r|K}P_{j|K}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P_K}{\partial x_r} &= \sigma\frac{\partial I_K}{\partial x_r}\frac{e^{I_K\sigma}\sum_{L\in\mathbf{K}} e^{I_L\sigma} - \left(e^{I_K\sigma}\right)^2}{\left(\sum_{L\in\mathbf{K}} e^{I_L\sigma}\right)^2} \\
&= \sigma\frac{\partial I_K}{\partial x_r}P_K(1-P_K)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial I_K}{\partial x_r} &= \frac{\partial V_r}{\partial x_r}\frac{1}{\sigma}\frac{e^{V_r}}{\sum_{l\in K} e^{V_l}} \\
&= \frac{\partial V_r}{\partial x_r}\frac{1}{\sigma}P_{r|K}
\end{aligned}$$

Combinando as duas últimas equações, nós temos:

$$\frac{\partial P_K}{\partial x_r} = \frac{\partial V_r}{\partial x_r}P_r(1-P_K)$$

Finalmente, o efeito marginal de alteração do atributo do produto r na probabilidade de escolha do produto j é dada por:

$$\begin{aligned}
\frac{\partial P_j}{\partial x_r} &= -\frac{\partial V_r}{\partial x_r}\frac{1}{\sigma}P_{r|K}P_{j|K}P_K + \frac{\partial V_r}{\partial x_r}P_r(1-P_K)P_{j|K} \\
&= \frac{\partial V_r}{\partial x_r}P_r\left[P_{j|K}(1-\frac{1}{\sigma}) - P_j\right]
\end{aligned}$$

Que também volta ao mesmo efeito cruzado do Logit tradicional caso $\sigma \rightarrow 1$. A Elasticidade, neste caso, é dada por:

$$E_{j,x_r} = \frac{\partial V_r}{\partial x_r} x_r \left[P_{j|K}(1 - \frac{1}{\sigma}) - P_j \right]$$

O terceiro caso envolve produtos que não pertencem ao mesmo ninho.

Produtos diferentes em ninhos diferentes

Supondo que o produto r seja localizado em um outro grupo de produtos, denominado M, os efeitos cruzados ficam diferenntes dos outros casos:

$$\begin{aligned}
\frac{\partial P_{j|K}}{\partial x_r} &= 0 \\
\frac{\partial P_K}{\partial x_r} &= -\sigma \frac{\partial I_M}{\partial x_r} \frac{e^{\sigma I_K} e^{\sigma I_M}}{\left(\sum_{L \in \mathbf{K}} e^{\sigma I_L} \right)^2} \\
\frac{\partial I_M}{\partial x_r} &= \frac{1}{\sigma} \frac{\partial V_r}{\partial x_r} \frac{e^{V_r/\sigma}}{\sum_{l \in M} e^{V_l/\sigma}} \\
&= \frac{1}{\sigma} \frac{\partial V_r}{\partial x_r} P_{r|M}
\end{aligned}$$

Substituindo a equação de $\frac{\partial I_M}{\partial x_r}$ na definição de $\frac{\partial P_K}{\partial x_r}$, nós temos:

$$\begin{aligned}
\frac{\partial P_K}{\partial x_r} &= -\frac{\partial V_r}{\partial x_r} P_{r|M} P_M P_K \\
&= -\frac{\partial V_r}{\partial x_r} P_r P_K
\end{aligned}$$

O que leva à seguinte expressão para o efeito cruzado:

$$\frac{\partial P_j}{\partial x_r} = -\frac{\partial V_r}{\partial x_r} P_r P_j$$

Neste caso, a fórmula é a mesma do modelo Logit tradicional, assim como a elasticidade cruzada:

$$E_{j,x_r} = -\frac{\partial V_{ir}}{\partial x_r} P_r x_r$$

Todas estas medidas de sensibilidade aos diversos atributos de diferentes produtos podem nos dar medidas de sensibilidade da demanda ao preço – tanto do bem quanto de produtos cruzados. Uma aplicação interessante destes modelos de escolha discreta é para a análise de bem-estar de mudanças de política. Este tópico será melhor analisado na seção seguinte.

3.1.2 Análise de Bem-Estar com Modelos de Escolha Discreta

Da mesma forma quer no capítulo anterior, uma aplicação interessante da modelagem de demanda com escolha discreta é na análise de bem-estar de mudanças na economia[6]. O primeiro passo é definir a demanda de mercado pelo produto j, para uma população com N indivíduos. Esta demanda pode ser escrita da seguinte forma, supondo que todos os consumidores são idênticos:

$$Q_j = NP_{ij}\tilde{x}_{ij}$$

Em que $\tilde{x}_{ij}$ denotaria a demanda do indivíduo representativo do produto (que supõe-se que seja um, mas a relevância deste termo será revelada mais adiante). Uma vez que $\tilde{x}_{ij}$ representa uma demanda individual, ela pode ser definida da seguinte forma:

$$q_j = -\left(\frac{N}{\frac{\partial v}{\partial w}}\right) P_{ij}\frac{\partial v}{\partial p_j}$$

Em que $v(\cdot)$ representa a função de utilidade indireta. Para que possamos avançar com a nossa análise, algumas hipóteses adicionais precisam ser feitas. A primeira delas é que a utilidade marginal da renda é independente dos preços de j e das características do mesmo. Dessa forma, podemos considerar $\frac{\partial v}{\partial w}$ como constante e representada por α. A segunda hipótese é que os efeitos renda de alteração de preços e características são relativamente pequenos, de forma que q_j pode representar a função demanda marshalliana. A terceira é que, para preços altos demais, o efeito das características do produto sobre a utilidade do consumidor é próximo de zero, pois nenhuma unidade está sendo consumida. Ou seja:

$$\lim_{p_j \to \infty} \frac{\partial V_{ij}}{\partial x_j} = 0$$

Um passo importante é definir qual é o efeito marginal sobre o gasto decorrente de uma pequena alteração em um dos atributos:

$$\frac{\partial e}{\partial x_j} = \left(\frac{N}{\alpha}\right)\frac{\partial}{\partial x_j}\int_{p_j^0}^{\infty} P_{ij}\frac{\partial V_{ij}}{\partial p_j}dp_j$$

Podemos fazer uma redefinição das variáveis $\omega_j = V_{ij}(\mathbf{p}, x_j)$, e a integral fica sendo:

$$\frac{\partial e}{\partial x_j} = \left(\frac{N}{\alpha}\right)\frac{\partial}{\partial x_j}\int_{\omega_j^0}^{\omega_j^1} P_{ij}(\omega_j)d\omega_j$$

[6] Uma referência clássica aqui é Small e Rosen (1981) [41]

Em que $\omega_j^0 = V_{ij}(p_j^0, \mathbf{p}, x_j)$ e $\omega_j^1 = V_{ij}(\infty, \mathbf{p}, x_j)$. Podemos adicionalmente integrar $x_j$para descobrirmos o efeito sobre o gasto decorrente de uma alteração discreta em x_j:

$$\Delta EC = -\left(\frac{N}{\alpha}\right)\int_{V_{ij}^0}^{V_{ij}^1} P_{ij} dV_{ij}$$

Esta formulação pode ser adaptada para os dois tipos principais de modelos de escolha discreta. No caso de um vetor ε com distribuição de valores extremos I, a variação no excedente do consumidor se torna:

$$\Delta EC = -\left(\frac{N}{\alpha}\right)[\ln\left(\sum_j \exp(V_{ij}^1)\right) - \ln\left(\sum_j \exp(V_{ij}^0)\right)]$$

Este insight de determinação das demandas agregadas a partir das probabilidades de escolha será retomado na próxima seção, sobre a aplicação deste tipo de modelo em situações em que os dados das escolhas individuais não estão disponíveis.

3.2 Dados Agregados

Em muitos casos, temos situações em que não é possível o acesso aos microdados da escolha do consumidor, e é disponível apenas a participação de mercado – por unidades – de cada produto. Ainda assim, é possível estender os conceitos discutidos anteriormente para lidar com esta situação. Para o caso de dados agregados, geralmente assumimos uma função de utilidade para o consumidor i da alternativa j da seguinte forma:

$$U_{ij} = \mathbf{x_j}\beta_{\mathbf{ij}} + \xi_j + \epsilon_{ij}$$

Em que $\mathbf{x_j}$ representa um vetor de características do produto j, β_{ij} representa uma sensibilidade específica do consumidor aos atributos dos produtos – que pode ser modelado como alguma forma de interação entre as características dos indivíduos e as características do produto – ξ_j representa as características não observadas pelo econometrista do produto, que dará margem à implementação econométrica deste modelo, e ϵ_{ij} representa a parte aleatória da utilidade. Vamos fazer uma partição do vetor $\mathbf{x_j}$ em dois componentes – um deles representando o preço do bem e o outro representando todas as outras características. Portanto:

$$\mathbf{x_j} = \left[\begin{array}{cc} p_j & \mathbf{\bar{x}_j} \end{array}\right]$$

Vamos, por simplicidade, assumir que a sensibilidade da utilidade do produto j a alterações nos preços seja constante entre os consumidores e igual a α, o que implica que o vetor de coeficientes $\beta_{\mathbf{ij}}$

pode ser particionado em dois componentes, $\beta_{\mathbf{ij}} = \left[\begin{array}{cc} \alpha & \tilde{\beta}_{ij} \end{array} \right]$. Como estamos aqui tratando de uma análise de escolha discreta em que não temos informações sobre as escolhas individuais, existem dois jeitos de se lidar com o fato que, potencialmente, os efeitos sobre a utilidade condicional indireta podem diferir entre os indivíduos. O primeiro deles seria ignorar completamente a heterogeneidade e supor que todos os indivíduos tivessem os mesmos valores para α e $\tilde{\beta}_{ij}$. Neste caso, supondo que tenhamos todas as características diretamente observadas pelo econometrista, podemos reescrever isso como:

$$\begin{aligned} U_{ij} &= \mathbf{x_j}\beta_{\mathbf{j}} + \xi_j + \epsilon_{ij} \\ &= V_j + \epsilon_{ij} \end{aligned}$$

Neste caso, a análise com dados agregados se torna bastante simples. Em primeiro lugar, supondo que existam N consumidores, temos que a demanda de mercado pelo produto j é dada por:

$$Q_j = N \times P_j$$

Em que P_j é a probabilidade de escolha do produto j, que é suposta idêntica para todos os consumidores. Supondo uma distribuição logística para o termo ϵ_{ij}, temos:

$$P_j = \frac{e^{V_j}}{\sum_{k \in J} e^{V_k}}$$

A participação de mercado – sendo o mercado definido em número de unidades – é dada por:

$$s_j = \frac{Q_j}{\sum_{k \in J} Q_k}$$

Isto pode ser redefinido como:

$$s_j = \frac{NP_j}{N} = P_j$$

Uma vez que a probabilidade de escolha entre os produtos tem que totalizar 100%. Agora, para que seja possível a estimação, em primeiro lugar, temos que ter alguns cuidados. O primeiro deles diz respeito aos efeitos marginais de alterações de um preço sobre as quantidades consumidas. Para entender isso melhor, suponhamos apenas dois produtos denominados A e B. Se o preço de A sobe, temos que o efeito marginal sobre a participação de mercado do produto seria de $\frac{\partial V_A}{\partial p_a} P_A P_B$. Da mesma forma, o efeito deste aumento sobre a participação de mercado de B seria exatamente de $-\frac{\partial V_A}{\partial p_A} P_A P_B$, o que daria um efeito nulo sobre os market shares (e, incidentalmente, sobre as quantidades, assumindo o tamanho do mercado constante). Isso tem algumas implicações sérias, especialmente se supusermos que, se os preços dos dois bens subirem marginalmente, não teremos efeito nenhum sobre a quantidade demandada, o que não é uma hipótese realista, de qualquer maneira.

Para lidar com este problema, é comum adotarmos a premissa que um dos bens consistiria na "não-escolha" dos bens[7]. Este bem teria sua parte determinística da utilidade normalizada em zero, o que significa que:

$$s_0 = \frac{1}{1 + \sum_{k \neq 1, \in J} e^{V_k}}$$

Enquanto para os outros bens, teríamos:

$$s_j = \frac{e^{V_j}}{1 + \sum_{k \neq 1, \in J} e^{V_k}}$$

Ou seja, podemos fazer a seguinte conta:

$$\ln(s_j) - \ln(s_0) = V_j = \mathbf{x_j} \beta_\mathbf{j} + \xi_j$$

Para a estimação dos parâmetros $\beta_\mathbf{j}$, usualmente incorporamos um termo erro ξ_j na equação. Evidentemente, as propriedades estatísticas deste termo erro ξ_j serão analisadas em profundidade, pois diferentemente da situação em que temos microdados, aqui é necessário que nos preocupemos com a relação entre o $\mathbf{x_j}$ e o termo ξ_j – a violação da hipótese de exogeneidade. Este ponto será analisado com mais cuidado no capítulo XXX mais adiante.

Caso tenhamos o modelo Logit aninhado, precisamos supor que o bem externo seja membro de um grupo composto por apenas este bem. Neste caso, temos que cada um dos bens tem a participação de mercado dada por:

$$\begin{aligned} s_j &= s_{j|K} s_K \\ s_{j|K} &= \frac{e^{V_j/\sigma}}{\sum_{l \in K} e^{V_l/\sigma}} \\ s_K &= \frac{e^{I_K \sigma}}{1 + \sum_{L \in \mathbf{K}} e^{I_L \sigma}}, K \neq 0 \end{aligned}$$

Para o bem externo, temos que:

$$s_0 = \frac{1}{1 + \sum_{L \in \mathbf{K}} e^{I_L \sigma}}$$

Ou $\ln(s_0) = -\ln(1 + \sum_{L \in \mathbf{K}} e^{I_L \sigma})$

[7] O problema aqui é que, ao eliminarmos um problema – o de assumirmos efeitos agregados sobre as quantidades – acabamos por introduzir outro, que é o da determinação do tamanho do mercado potencial (que daria o tamanho do mercado e a quantidade adquirida do produto 0).

Lembrando que podemos reescrever a equação para I_K da seguinte forma:

$$e^{I_K} = \sum_{l \in K} e^{V_l}$$

Podemos escrever a definição de s_j da seguinte forma:

$$\begin{aligned} s_j &= \frac{e^{V_j/\sigma}}{e^{I_K}} \times \frac{e^{I_K \sigma}}{1 + \sum_{L \in \mathbf{K}} e^{I_L \sigma}} \\ &= \frac{e^{V_j/\sigma} e^{I_K(\sigma-1)}}{1 + \sum_{L \in \mathbf{K}} e^{I_K \sigma}} \end{aligned}$$

Tirando o logaritmo da definição de s_j, temos:

$$\ln(s_j) = \frac{V_j}{\sigma} + I_K(\sigma - 1) - \ln(1 + \sum_{L \in \mathbf{K}} e^{I_K \sigma})$$

Aqui temos um problema, relacionado com o termo I_K. Passando o logaritmo em s_K, temos:

$$\begin{aligned} \ln(s_K) &= \sigma I_K - \ln(1 + \sum_{L \in \mathbf{K}} e^{I_K \sigma}) \\ \ln(s_K) &= \sigma I_K + \ln(s_0) \\ I_K &= \frac{\ln(s_K) - \ln(s_0)}{\sigma} \end{aligned}$$

Substituindo na fórmula acima, temos:

$$\begin{aligned} \ln(s_j) &= \frac{V_j}{\sigma} - \left(\frac{\ln(s_K) - \ln(s_0)}{\sigma} \right) (1 - \sigma) + \ln(s_0) \\ V_j &= \sigma \ln(s_j) + (1 - \sigma) \ln(s_K) - (1 - \sigma) \ln(s_0) - \sigma \ln(s_0) \\ V_j &= \sigma \ln(s_j) + \ln(s_j) - \ln(s_j) + (1 - \sigma) \ln(s_K) - \ln(s_0) \\ V_j &= \ln(s_j) - \ln(s_0) - (1 - \sigma)(\ln(s_j) - \ln(s_K)) \\ V_j &= \ln(s_j) - \ln(s_0) - (1 - \sigma) \ln(s_{j|K}) \\ \ln(s_j) - \ln(s_0) &= V_j + (1 - \sigma) \ln(s_{j|K}) \end{aligned}$$

A passagem da segunda para a terceira equação é dada por somar e subtrair $\ln(s_j)$. E quando a hipótese de igualdade dos coeficientes entre as diferentes pessoas não se mantém? Até há uns 20 anos, não existiam soluções para este tipo de situação, mas atualmente já existem formas de lidar com isto. O ponto de partida conceitual para isto é o artigo de Berry (1994) [10], que parte da seguinte decomposição do vetor $\beta_{\mathbf{ij}}$:

$$\beta_{\mathbf{ij}} = \bar{\beta}_{\mathbf{j}} + \sigma_{\mathbf{i}}\zeta_{\mathbf{ij}}$$

Ou seja, o valor que a sensibilidade da utilidade para o indívíduo i tem em relação a alterações no atributo j é composta por duas partes: uma parte "média", denotada por $\bar{\beta}_{\mathbf{j}}$, e uma parte aleatória composta por um valor retirado de uma distribuição normal padronizada ($\zeta_{\mathbf{ij}}$) e um valor referente à dispersão dos gostos entre a população $\sigma_{\mathbf{i}}$. Substituindo a equação acima na especificação da utilidade, temos:

$$\begin{aligned} U_{ij} &= \mathbf{x_j}\beta_{\mathbf{ij}} + \xi_j + \epsilon_{ij} \\ &= \mathbf{x_j}(\bar{\beta}_{\mathbf{j}} + \sigma_{\mathbf{i}}\zeta_{\mathbf{ij}}) + \xi_j + \epsilon_{ij} \\ &= \mathbf{x_j}\bar{\beta}_{\mathbf{j}} + \mathbf{x_j}\sigma_{\mathbf{i}}\zeta_{\mathbf{ij}} + \xi_j + \epsilon_{ij} \end{aligned}$$

Podemos juntar todos estes termos da seguinte forma:

$$\begin{aligned} U_{ij} &= V_{ij} + v_{ij} \\ V_{ij} &= x_j\bar{\beta}_j + \xi_j \\ v_{ij} &= \mathbf{x_j}\sigma_{\mathbf{i}}\zeta_{\mathbf{ij}} + \epsilon_{ij} \end{aligned}$$

Neste caso, o termo aleatório ϵ_{ij}necessariamente é heterocedástico, pois além da parte aleatória da utilidade, capturada pelo termo ϵ_{ij}, temos as diferenças nas avaliações dos indivíduos, que é representada por $\mathbf{x_j}\sigma_{\mathbf{j}}\zeta_{\mathbf{ij}}$. Esta heterogeneidade é o que permite que possamos nos desviar dos problemas de padrões estranhos de substituição do modelo Logit tradicional. Iremos detalhar isso um pouco mais na seção seguinte.

3.3 Características Não Observadas

Nesta seção, iremos trabalhar os modelos de escolha discreta em que existem características não observáveis, tanto dos produtos quanto dos consumidores, permitindo heterogeneidade dos mesmos e fazendo com que o mapa de elasticidades cruzadas tenha características mais realista, de acordo com a metodologia de Berry, Levinsohn e Pakes (1995)[9]. Uma vez que é baseada em uma microfundamentação bem robusta, permite que estendamos esta metodoogia para um conjunto bem amplo de situações. Vamos então começar introduzindo o modelo microeconômico subjacente:

$$U_{ij} = \mathbf{x_j}\beta_{\mathbf{ij}} + \xi_j + \epsilon_{ij}$$

A diferença em relação ao que vimos na aula anterior é que voltamos a supor que a sensibilidade do consumidor i em relação ao atributo k do produto j pode ser diferente da mesma sensibilidade no caso de outro consumidor, ou seja $\beta_{ik} \neq \beta_{dk}, \forall i, d$. A idéia aqui é modelar a heterogeneidade dos coeficientes dos indivíduos como sendo funções das características – tanto observadas quanto não observadas pelo econometrista:

$$\beta_{\mathbf{ij}} = \bar{\beta}_{\mathbf{j}} + \mathbf{z_i}\beta^{\mathbf{o}} + \mathbf{v_i}\beta^{\mathbf{u}}$$

Ou seja, este coeficiente é composto por uma sensibilidade "média", $\bar{\beta}_{\mathbf{j}}$, os efeitos que um conjunto de características individuais observadas possui sobre este coeficiente, que denotaremos $\mathbf{z}_i$ (com os efeitos sobre o coeficiente determinado por $\beta^{\mathbf{o}}$), e os efeitos que um outro conjunto de características individuais **não observadas**, denominadas $\mathbf{v}_i$, possui sobre este coeficiente (cuja sensibilidade é dada por $\beta^{\mathbf{u}}$). Substituindo esta definição dos coeficientes, temos que:

$$U_{ij} = \mathbf{x_j}\bar{\beta}_{\mathbf{j}} + \mathbf{x_j z_i}\beta^{\mathbf{o}} + \mathbf{x_j v_i}\beta^{\mathbf{u}} + \xi_j + \epsilon_{ij}$$

Podemos reescrever de forma similar à da seção anterior, separando esta especificação em uma parte correspondente á utilidade "média" da alternativa, e a parte aleatória:

$$\begin{aligned} U_{ij} &= V_{ij} + \mathbf{x_j z_i}\beta^{\mathbf{o}} + \mathbf{x_j v_i}\beta^{\mathbf{u}} + \epsilon_{ij} \\ V_{ij} &= \mathbf{x_j}\bar{\beta}_{\mathbf{j}} + \xi_j \end{aligned}$$

É interessante notar que o modelo possui interações entre as características dos consumidores e as características do produto que, na prática, resolvem o problema da independência de alternativas irrelevantes, permitindo a modelagem das inter-relações dos produtos de forma mais realista. Na maior parte dos casos, não teremos dados observados sobre as características dos consumidores que compraram cada uma das alternativas, de forma que apenas trabalharemos com características não observadas dos consumidores – ou seja, apenas os termos $\mathbf{v}_i$. O algoritmo detalhado mais adiante consiste em três etapas.

1. Especificar as participações de mercado em função dos coeficientes;

2. Recuperar os sinais dos termos ξ_j a partir dos resultados da etapa anterior;

3. Estimar os coeficientes por GMM

Estimar as Participações de Mercado

A primeira parte é recuperar as participações de mercado em função da utilidade média, V_{ij}, e dos coeficientes β_k^u. Uma vez que os termos $\mathbf{v}_i$ não são observados, iremos impor uma premissa sobre eles, que é a que eles seguem uma distribuição qualquer, enquanto que os termos ϵ_{ij} seguem a já tradicional distribuição de Valores Extremos I. Desta forma, temos que:

$$s_j(V_j, \beta) = \int \left(\frac{\exp[V_j + \mathbf{x_j v_i}\beta^{\mathbf{u}}]}{1 + \sum_{q>0} \exp[V_q + \mathbf{x_q v_i}\beta^{\mathbf{u}}]} \right) f(\mathbf{v}) d\mathbf{v}$$

Não existe uma forma analítica de resolução desta integral, de forma que teremos que usar ou métodos numéricos (por exemplo, método da quadratura) ou métodos de simulação para a obtenção do valor desta integral – e, consequentemente, o valor desta participação de mercado. Vamos falar um pouco sobre os métodos de simulação.

Em geral simulação consiste em sortear valores aleatórios de uma distribuição, calcular alguma coisa com cada um destas valores sorteados e depois tirar uma média destes cálculos. Em todos estes casos, o pesquisador quer calcular uma média da forma $\tilde{t} = \int t(\varepsilon)f(\varepsilon)d\varepsilon$. Existem vários métodos para a obtenção destas observações. No caso em questão, o primeiro passo é definir qual seria a forma da distribuição das características dos consumidores. Um caminho interessante para determinar a distribuição das características dos consumidores é utilizar micro-dados como os do Censo, da PNAD ou de outras fontes. Estes dados podem nos dar uma forma para a função $f(\mathbf{v})$.

Agora, supondo que você já saiba esta distribuição, o passo seguinte é obter um determinado número de valores sorteados desta distribuição $f(\mathbf{v})$. Vamos supor que você tenha sorteado ns valores desta distribuição. Com estes valores, podemos calcular a participação de mercado da alternativa j com a seguinte fórmula:

$$\hat{s}^{ns}(\delta_j, \beta) = \frac{1}{ns}\sum_{r=1}^{ns}\left[\frac{\exp[V_j + \sum_r \mathbf{x_j}\beta^{\mathbf{u}}\mathbf{v_{ir}}]}{1 + \sum_{q>0}\exp[V_j + \sum_r \mathbf{x_q}\beta^{\mathbf{u}}\mathbf{v_{ir}}]}\right]$$

Ou seja, pegamos os valores de $\mathbf{v}_i$ associados com os sorteios de cada característica, e calculamos a fórmula do *share* com cada um deles. Depois disso, somamos e tiramos a média. Evidentemente, a utilização de métodos de simulação aumenta a imprecisão das estimativas, que pode ser reduzida quanto maior for o valor de ns. Além disso, existem métodos, revisados em Train (2003)[48], que permitem ganhos em termos de tempo para o cálculo destas estimativas. Agora passemos á etapa seguinte de estimação.

Calculando a Utilidade Média

O passo seguinte é calcular os efeitos das características dos produtos que não são observadas pelo analista. Para isso, precisamos obter estimativas do termo V_j. Tendo este negócio, podemos obter estimativas do ξ_j, do $\bar{\beta}_{\mathbf{j}}$ e do $\beta^{\mathbf{u}}$. Em primeiro lugar, no paper BLP, os autores notam que a seguinte relação:

$$V_j^h = V_j^{h-1} + \ln[s_j] - \ln[\hat{s}_j^{ns}]$$

É um chamado *contraction mapping*, que possui um ponto fixo. Ou seja, se fizermos um processo sequencial, começando com um valor inicial para o V_j, e em cada iteração ajustando o valor de V_j no valor igual à diferença entre os logs da participação de mercado e a participação de mercado observada, acabaremos em um ponto fixo, em que não há alterações adicionais em V_j. Um bom valor inicial para fazer a recursão pode ser a participação de mercado obtida com a estimação de um modelo LOGIT multinomial como o da seção anterior. Na verdade, na programação é utilizada a seguinte versão não-linear do estimador:

$$\exp[V_j^h] = \exp[V_j^{h-1}] \times \frac{s_j}{\hat{s}_j^{ns}}$$

Desta forma, obtemos uma estimativa para o termo V_j. A partir desta estimativa, podemos construir as condições de momento, lembrando que o termo ξ_j pode ser entendida como a diferença entre a utilidade média e os valores dos coeficientes médios. O passo seguinte é construir as condições de momento.

Construindo as Condições de Momento

Tendo este valor para V_{ij}, podemos definir os erros da seguinte forma:

$$\xi_j = V_j - \mathbf{x_j}\bar{\beta}_\mathbf{j}$$

O problema é que aqui temos que estimar os $\bar{\beta}_\mathbf{j}$. Isto é análogo a concentrar a função objetivo para ficar apenas em termos dos $\beta^\mathbf{u}$. Vamos fazer isso projetando a covariância entre os $\mathbf{x_j}$ e as variáveis instrumentais, $\mathbf{Z}$, na covariância entre o V_j e os instrumentos, ou seja:

$$\bar{\beta}_j = (\mathbf{K}(\mathbf{Z^T Z})\mathbf{K^T})^{-1}\mathbf{K}(\mathbf{Z^T Z})\mathbf{L}$$

Em que $\mathbf{K} = \mathbf{x_j^T Z}$ e $\mathbf{L} = V_j^T\mathbf{Z}$. Implicitamente, esta conta significa que estamos expressando o $\bar{\beta}_j$ como função dos β^u. Supondo que tenhamos variáveis exógenas e instrumentos para os preços, podemos construir as condições de momento da seguinte forma:

$$m(\theta) = \Xi\mathbf{Z}$$

Em que Ξ é o empilhamento dos ξ_j, e $\mathbf{Z}$ é a matriz de instrumentos. Com isto, podemos construir uma função objetivo da seguinte forma:

$$q = \Xi^T\mathbf{Z}\mathbf{\Phi^{-1}}\mathbf{Z}^T\Xi$$

Em que Φ representa uma estimativa da matriz variância-covariância dos momentos das equações, ou $\Phi = E(\mathbf{Z}^T\Xi\Xi^T\mathbf{Z})$. Em geral, esta matriz de ponderação é algo mais chatinho de ser feito. Um caminho razoavelmente adequado envolve, na primeira iteração, calcular Φ como sendo homocedástica $\Phi = \mathbf{Z}^T\mathbf{Z}$. Com isto, são calculados os valores dos parâmetros e, com os valores dos parâmetros, calcular aí uma estimativa completa da matriz de variância covariância das condições de momento, $\Phi = \mathbf{Z}^T\Xi\Xi^T\mathbf{Z}$.

Neste ponto, perdemos de vista os coeficientes. Para isto, precisamos ter em mente que as etapas anteriores são repetidas todas ao longo de cada iteração. Ou seja, em cada iteração temos o seguinte roteiro. Inicialmente, são fornecidos novos valores para os coeficientes, segundo os quais são realizadas as etapas acima e obtendo, no processo, um novo valor para o vetor de coeficientes.

3.4 Exercícios

1. O papel deste exercício é dar uma idéia da relevância da normalização dos coeficientes. Portanto, vamos simular o comportamento de 500 diferentes indivíduos com a possibilidade de escolha de cinco alternativas – quatro produtos e uma "alternativa externa". Cada indivíduo possui a seguinte função utilidade:

$$U_{ij} = -\alpha p_j + \beta x_j + \epsilon_{ij}, j = 1, 2, 3, 4$$

Em que α e β são parâmetros a serem estimados, p_jé o preço do bem j, x_j é uma outra característica do mesmo bem, e ϵ_{ij} é a parte idiosincrática da escolha do consumidor – supondo que siga uma distribuição Gumbel e independente entre as alternativas. Desta forma, simule os preços dos 4 produtos (a alternativa externa pode ser normalizada de forma a $U_{i0} = \epsilon_{i0}$) da seguinte forma:

- Preços dos produtos – uniformemente distribuídos entre 0,25 e 1. Suponha que os preços dos produtos são os mesmos entre todos os consumidores.
- Característica x – normalmente distribuída, com média 1 e desvio-padrão 1.
- ϵ_{ij} – retirado de uma distribuição Gumbel (como no comando `evrnd` do MATLAB– Toolbox Estatístico)
- $\alpha = 1$ e $\beta = 2$.

A partir desta simulação e determinando qual dos produtos cada consumidor escolhe, e de posse dos preços e das características de cada produto, tente estimar os coeficientes α e β por Máxima Verossimilhança. Os valores obtidos para α são estatisticamente diferentes de um? E os valores obtidos para β, são estatisticamente diferentes de dois? Porquê sim, ou porque não?

Capítulo 4

Agregação e Orçamento em Múltiplos Estágios

4.1 Agregação e Separabilidade

Podemos notar que uma característica comum a todos os modelos acima mostrados é a sua dependência dos preços de todos os outros produtos envolvidos. Especialmente para o caso de produtos diferenciados e/ou com um número grande de produtores, podemos imaginar que o número de coeficientes a ser estimados rapidamente exaure a informação contida na amostra, por maior que ela seja. Desta forma, uma solução interessante pode ser a hipótese de separabilidade, de forma que a escolha do consumidor pode ser quebrada em pedaços menores, colocando desta forma menos exigências sobre os dados disponíveis. Em especial, vamos nos focar sobre a separabilidade das preferências.

Do ponto de vista de separabilidade, podemos assumir que um ordenamento de preferências é diretamente aditivo se ele pode ser representado por uma função da forma:

$$U(X) = T(v_i(x_i))$$

Nada garante que precisamos ter apenas um bem no vetor x_i. Neste caso, teríamos algo análogo a uma árvore de utilidade, em que são escolhidos os gastos nos níveis mais elevados primeiro e, em segundo lugar, são escolhidos nos mais baixos. Esta é a idéia de *orçamento em dois estágios*. NO primeiro estágio, a alocação é possível com o conhecimento do gasto total e dos valores de "índices de preços" definidos para cada um dos subgrupos. No segundo estágio, as despesas individuais são definidas com base nos gastos dentro do grupo e os valores dos preços de cada bem dentro do grupo. Note-se que a separabilidade, como mostramos acima, é condição necessária e suficiente para o segundo estágio do orçamento em dois estágio. Apesar de ser uma condição necessária e suficiente para o segundo estágio, no primeiro estágio as coisas são mais complicadas, o que faz com os dois conceitos não sejam simétricos; não é verdade que um implique o outro.

E como isto impacta a nossa análise? Podemos construir um sistema de demanda bem mais complexo, em que tenhamos vários "subsistemas"; alguns para cada um dos subgrupos e um sistema determinando o gasto em cada um dos grupos. Vamos fazer um exemplo, supondo que exista um estágio superior em que são determinados os consumos agregados de um produto, e em um nível inferior são determinados os gastos em cada uma das marcas de um produto. No nível inferior iremos supor que as demandas para cada produto sejam dadas pelo modelo AIDS, na sua versão linearizada:

$$s_i = \alpha_i + \sum_i \beta_{ki} \ln p_k + \gamma_i \ln \left(\frac{w}{\mathbf{P}^*} \right)$$

Vamos supor adicionalmente que no nível superior a escolha do gasto no agregado não dependa dos gastos agregados em nenhuma outra coisa, apenas do índice de preços, da renda dos consumidores e de qualquer outra coisa que o sujeito queira. Aqui, poderíamos fazer uma função demanda agregada da forma que quiséssemos, mas vamos fazer uma demanda log-linear (desconsiderando as restrições que mencionei mais acima):

$$\ln Q = \alpha_N + \beta \ln W + \Gamma \ln \mathbf{P}^* + etc$$

Neste caso, as elasticidades-preço e renda terão os componentes que mencionamos acima, mas terão um elemento adicional, que representa o efeito da mudança dos preços sobre a alocação determinada no estágio superior.

Vamos fazer a derivação desta fórmula, começando com o resultado conhecido que:

$$\ln q_i = \ln w - \ln p_i + \ln s_i$$

Uma vez que podemos escrever a elasticidade como sendo a derivada do log da quantidade como a derivada do log do preço, temos:

$$e_{ii} = \frac{\partial \ln q_i}{\partial \ln p_i} = \frac{\partial(\ln w - \ln p_i + \ln s_i)}{\partial \ln p_i} = \frac{\partial \ln w}{\partial \ln p_i} - 1 + \frac{\partial \ln s_i}{\partial \ln p_i}$$

Comecemos com o terceiro dos termos:

$$\frac{\partial \ln s_i}{\partial \ln p_i} = \frac{1}{s_i}\frac{\partial s_i}{\partial \ln p_i} = \frac{1}{s_i}\left(\beta_{ii} + \gamma_i \frac{\partial \ln w}{\partial \ln p_i} - \gamma_i \frac{\partial \ln \mathbf{P}^*}{\partial \ln p_i} \right)$$

O que, jogando de volta, nos dá:

$$e_{ii} = \frac{\partial \ln w}{\partial \ln p_i}\left(1 + \frac{\gamma_i}{s_i}\right) - 1 + \frac{1}{s_i}\left(\beta_{ii} - \gamma_i \frac{\partial \ln \mathbf{P}^*}{\partial \ln p_i} \right)$$

Com a aproximação linear de Stone (1959), temos que:

$$\frac{\partial \ln \mathbf{P}^*}{\partial \ln p_i} \simeq \bar{s}_i$$

O que simplifica mais um pouco nossa conta, deixando-a da seguinte forma:

$$e_{ii} = \frac{\partial \ln w}{\partial \ln p_i}\left(1 + \frac{\gamma_i}{s_i}\right) - 1 + \frac{1}{s_i}\left(\beta_{ii} - \gamma_i \bar{s}_i\right)$$

Agora a última derivada que temos que eliminar é o de w com relação a p_i. Esta elasticidade pode ser resolvida, ao aproximarmos w como:

$$w = \sum_i p_i q_i \simeq \mathbf{P}^* \sum_i q_i = \mathbf{P}^* Q$$

O que equivale a:

$$\ln w = \ln \mathbf{P}^* + \ln Q$$

Ou seja:

$$\frac{\partial \ln w}{\partial \ln p_i} = \frac{\partial \ln \mathbf{P}^*}{\partial \ln p_i} + \frac{\partial \ln Q}{\partial \ln p_i} = \frac{\partial \ln \mathbf{P}^*}{\partial \ln p_i}\left(1 + \frac{\partial \ln Q}{\partial \ln \mathbf{P}^*}\right)$$

O que, segundo a equação do nível superior, nos dá:

$$\frac{\partial \ln w}{\partial \ln p_i} = \bar{s}_i\left(1 + \Gamma\right)$$

Desta forma, temos as seguintes elasticidades-preço e elasticidades cruzadas:

$$\begin{aligned} e_{ii} &= \frac{\beta_{ii} - \gamma_i \bar{s}_i}{s_i} - 1 + \left(1 + \frac{\gamma_i}{s_i}\right)(1 + \Gamma)\bar{s}_i \\ e_{ij} &= \frac{\beta_{ij} - \gamma_i \bar{s}_j}{s_i} + \left(1 + \frac{\gamma_i}{s_i}\right)(1 + \Gamma)\bar{s}_j \end{aligned}$$

4.1.1 Sistema em três estágios[1]

Neste caso, temos um sistema em três níveis. No nível mais baixo (i do sub-ninho m), temos um sistema como o AIDS tradicional:

$$s_{imt} = \alpha_{im} + \gamma_i \ln\left(\frac{w_{It}}{\mathbf{P}^*}\right) + \sum_j \beta_j \ln p_{jmt} + \varepsilon_{imt}$$

[1] Agradecimento a Verônica Orellano da EESP/FGV pela demonstração

Sendo que w_{mt} representa o gasto agregado com as marcas neste ninho inferior e $\mathbf{P}^*$ o índice de preços – Stone – dos produtos neste ninho. No estágio intermediário, temos uma equação duplo log:

$$\ln Q_{mt} = \beta_m \ln w_M + \sum_{j=1}^{M} \delta_{mj} \ln \mathbf{P}_m^* + \varepsilon_{mt}$$

No nível superior, temos:

$$\ln \mathbb{Q}_t = \beta_0 + \beta_1 \ln W_t + \beta_2 \ln \Pi_t + \varepsilon_t$$

Neste caso, W é o dispêndio total e Π_t um índice de preços geral da cerveja. Vamos agora derivar as elasticidades:

Mesmo ninho inferior

Começamos pela definição de participação de mercado no ninho inferior:

$$s_{imt} = \frac{p_{imt} q_{imt}}{w_{It}}$$

Passando o log, temos:

$$\begin{aligned} \ln s_{imt} &= \ln p_{imt} + \ln q_{imt} - \ln w_{It} \\ \ln q_{imt} &= \ln s_{imt} - \ln p_{imt} + \ln w_{It} \end{aligned}$$

Tirando a derivada logarítmica desta última equação, temos uma fórmula geral para a elasticidade cruzada:

$$\eta_{ij} = \frac{\partial \ln q_{imt}}{\partial \ln p_{jmt}} = \frac{\partial \ln s_{imt}}{\partial \ln p_{jmt}} + 1[i = j] + \frac{\partial \ln w_{It}}{\partial \ln p_{jmt}}$$

Vamos analisar o primeiro dos termos:

$$\frac{\partial \ln s_{imt}}{\partial \ln p_{jmt}} = \frac{1}{s_{imt}} \frac{\partial s_{imt}}{\partial \ln p_{jmt}} = \frac{1}{s_{imt}} \left[\gamma_i \frac{\partial w_{It}}{\partial \ln p_{jmt}} - \gamma_i \frac{\partial \mathbf{P}^*}{\partial \ln p_{jmt}} + \beta_{ij} \right]$$

Jogando isso na fórmula anterior, temos:

$$\eta_{ij} = \frac{\partial w_{It}}{\partial \ln p_{jmt}} \left(1 + \frac{\gamma_i}{s_{imt}} \right) - \frac{1}{s_{imt}} \left[\beta_{ij} - \gamma_i \frac{\partial \mathbf{P}^*}{\partial \ln p_{jmt}} \right] + 1[i = j]$$

Agora precisamos resolver as duas derivadas que ainda estão perdidas aí, que são:

$$\begin{aligned} \frac{\partial w_{It}}{\partial \ln p_{jmt}} &= (1 + \delta_i) \bar{w}_{It} \\ \frac{\partial \mathbf{P}^*}{\partial \ln p_{jmt}} &= \bar{w}_{It} \end{aligned}$$

Ninhos Inferiores Diferentes

Aqui temos uma grande diferença devida à hipótese da separabilidade. Ou seja, a quantidade demandada de um produto no nível inferior é uma função dos preços "médios" nos dois segmentos, o que faz com que a sensibilidade da quantidade do produto i ao preço do produto j, sendo que o produto j pertence a um ninho inferior diferente dependa dos preços médios dos dois grupos e da importância de cada produto em cada ninho. Ou seja:

$$\epsilon_{ij} = s_{i|I} \times \epsilon_{IJ} \times s_{j|J}$$

Em que $s_{i|I}$ é a participação do produto i no ninho dele (que chamaremos de I), $s_{j|J}$ a participação de j no ninho de J e ϵ_{IJ} é a elasticidade cruzada entre os grupos de produtos I e J, calculada na equação de nível superior. Com a função detalhada acima, temos que ϵ_{IJ} é um dos δ estimados.

Usualmente, estas simplificações não vêm de graça. A imposição destas restrições usualmente não são apoiadas pelos dados, de forma que as elasticidades resultantes nestes dados tendem a ser bem diferentes das que obteríamos caso conseguíssemos estimá-las diretamente. Ou seja, este tipo de abordagem pode ser utilizado, mas é necessário que tenhamos muito cuidado com a imposição deste tipo de estrutura.

Capítulo 5

Identificação de Modelos de Demanda

5.1 Introdução

Neste capítulo, iremos discutir mais detalhadamente como conseguiríamos a identificação dos parâmetros estruturais de um modelo de demanda. Em princípio, a estimação de um modelo de demanda, com a forma geral:

$$q_i^d = f(\mathbf{P}, \mathbf{X}) + \varepsilon_i \tag{5.1}$$

Em que **P** representa um vetor de preços de todos os produtos e **X** os outros deslocadores de demanda, possui problemas de identificação como os tradicionais identificados por Wooldridge (2009) [51], como por exemplo erros de medida, a omissão de variáveis relevantes, e o viés de simultaneidade. Para lidar com estes problemas, o caminho mais comum da literatura passa pelo uso de Variáveis Instrumentais.

O grande problema aqui é como encontrar instrumentos. O que a literatura econômica afirma é que o viés de simultaneidade está presente em todos os elementos do vetor **P**, sendo que, para muitas indústrias em que os produtos são diferenciados, este vetor tem uma dimensão linha bastante grande. Isso faz com que o número de instrumentos necessário também seja muito elevado[1].

Como afirma Bresnahan (1997) [11], esta questão gerou uma ampla gama de abordagens econométricas distintas. Ainda segundo este autor, a melhor estratégia envolve a combinação de mais de uma abordagem, combinada com conhecimento do setor e das empresas envolvidas, buscando entender como o lado da oferta funciona e criando assim instrumentos válidos. Tais abordagens podem ser classificadas como:

1. Usar deslocadores de custo observáveis para cada um dos produtos de uma indústria

[1]Pela conhecida Condição de Ordem de Identificação, é necessário que tenhamos pelo menos um instrumento para cada um dos elementos do vetor **P**.

2. Ter deslocadores de custo, mas em quantidade inferior ao número de produtos de uma indústria
3. Usar micro-dados de consumidores, e supor que a varição observável entre os consumidores identifique os coeficientes da curva de demanda
4. Supor um modelo específico para produtos diferenciados, como competição *à la* Bertrand, e use as restrições entre a oferta e a demanda para identificar o modelo.
5. Fazer um argumento de componentes de variância
6. Usar os preços defasados
7. Assumir que o termo erro na forma reduzida para preços não é correlacionado com o termo erro na equação de demanda e estimar tudo isso por MQO

Cada um destes pontos será detalhado mais adiante, e suas implicações e pressupostos também o serão. A primeira das abordagens é a mais recomendada pela literatura econométrica. Imagine que, para cada um elemento k do vetor $\mathbf{P}$, temos a seguinte forma reduzida:

$$p_{kt} = \gamma_0 + \gamma_1 z_{kt} + v_t$$

Neste caso, o procedimento de Variáveis Instrumentais é especialmente adequado. Usualmente, tais instrumentos são encontrados entre os preços de insumos pagos pelas empresas, tais como preços de embalagens, mão-de-obra, energia. A principal desvantagem deste método é que dificilmente teremos um número de variáveis z igual ao número de produtos, especialmente quando o problema envolve a análise de produtos diferenciados. Em uma indústria de refrigerantes, por exemplo, não é incomum encontrarmos mercados com 20 marcas coexistindo, e a obtenção de variáveis instrumentais diferentes para cada um dos 20 preços é extremamente difícil.

O que costuma ser bem mais comum é a existência de mais de um instrumento disponível, mas ainda assim em número bastante inferior ao número de coeficientes que precisam ser identificados no sistema de demanda. Este é o caso mais comum, e costuma pedir por mais hipóteses. Uma das possibilidades é a utilização de procedimentos de agregação e separabilidade das decisões do consumidor, como visto no capítulo anterior.

Ao reduzir o número de coeficientes do modelo, um sistema de orçamento em dois – ou vários – estágios reduz o número de instrumentos necessários para a identificação dos coeficientes. Um modelo baseado no sistema de demanda neoclássico tem um número de coeficientes aproximadamente igual ao quadrado do número de produtos. Ou seja, se desejássemos estimar um sistema de equações com estes produtos, teríamos de ter dados suficientes para estimar aproximadamente 400 coeficientes, além de um número equivalente de instrumentos. Um primeiro passo é impor condições como simetria dos efeitos cruzados, homogeneidade das funções demanda, e se assumirmos múltiplos estágios o número se reduz mais ainda.

Esta pode ser uma abordagem interessante, mas é necessário cuidado. Baker e Rubinfeld (1999) [6] apontam que os resultados podem ser muito dependentes de qual tipo de hipótese é feita sobre a agregação ou sobre a simetria das matrizes de elasticidades cruzadas. Por exemplo, se um agrupamento em vários estágios acaba por colocar Guaraná Antarctica e Kuat em um mesmo grupo e Coca-Cola e Pepsi em outro, isso aumenta a chance que encontremos uma elevada elasticidade cruzada entre Guaraná Antarctica e Kuat por um lado, e diminui a chance que encontremos uma alta elasticidade cruzada entre Guaraná-Antarctica e Coca-Cola.

O terceiro dos pontos envolve a utilização de micro-dados dos consumidores e, a partir daí, assumir que os preços são exógenos à decisão do consumidor. Neste caso, as diferenças entre os consumidores é que fazem o trabalho de determinar os coeficientes de interesse. Esta presmissa está implícita na análise realizada nas seções 3.1 e no capítulo 2, especialmente para justificar o uso do Método da Máxima Verossimilhança para estimar o modelo LOGIT e o LOGIT aninhado, bem como o Método das Regressões aparentemente não-correlacionadas para estimar o sistema de equações no capítulo 2.

De uma forma geral, esta é uma boa abordagem, pois em muitos casos é possível assumir o vetor de preços enfrentado pelo agente como sendo exógeno à decisão de compra do mesmo. O grande problema desta abordagem é que, na maior parte das vezes, tal disponibbilidade de dados não é factível, sendo disponível no máximo dados agregados de venda e preços.

O passo seguinte envolve assumir um modelo específico de competição entre os produtos e, a partir daí, derivar as condições de primeira ordem[2]. Estas condições de primeira ordem determinariam a adequação de alguns instrumentos. Para que isto seja melhor compreendido, vamos supor que tenhamos uma empresa multiproduto, com o número total de produtos indexado por J. Neste caso, a função lucro desta empresa é dada por:

$$\Pi = \sum_{i \in J} q_i(\mathbf{P}) p_i - CT(\mathbf{Q^J}, w)$$

Em que $\mathbf{Q}$ representa o vetor de quantidades (tanto dos produtos oferecidos pela empresa quanto os oferecidos pelas outras empresas) enquanto o vetor $\mathbf{Q^J}$ representa as quantidades produzidas dos produtos oferecidos pela empresa. A condição de primeira ordem para o $j-$ésimo produto é dada por:

$$\frac{\partial \Pi}{\partial p_j} = q_j + \sum_{i \in J} \left(p_i - \frac{\partial CT}{\partial q_i} \right) \frac{\partial q_i}{\partial p_j} = 0$$

[2]Esta abordagem é adotada por Bresnahan, Stern e Trajtenberg (1997) [14], em seu artigo sobre computadores pessoais, por sua vez adaptado de um texto mais antigo, Bresnahan (1981) [12].

Reorganizando, e nas linhas de Bresnahan, Stern e Trajtemberg (1997)[14], temos que:

$$
\begin{aligned}
-p_j q_j &= \sum_{i \in J} \left(p_i - \frac{\partial CT}{\partial q_i} \right) \frac{\partial q_i}{\partial p_j} p_j \\
p_j &= \sum_{i \in J} \left(p_i - \frac{\partial CT}{\partial q_i} \right) \frac{\partial q_i}{\partial p_j} \frac{p_j q_i}{-q_j q_i} \\
p_j &= \sum_{i \in J} \left(p_i - \frac{\partial CT}{\partial q_i} \right) \eta_{ij} \frac{1}{-q_j}
\end{aligned}
$$

Ou seja, o que a equação a seguir mostra é que qualquer coisa que afete a somatória do lado direito da igualdade pode ser usado como um instrumento válido – e, portanto, os preços dos outros produtos de uma mesma empresa. Adicionalmente, os preços de fatores de produção, como argumentos da função CT também podem ser considerados como instrumentos válidos.

I1. Be given observable cost shifters for each of the products in an industry. Rare and lucky.

I2. Have a few supply shifters from the cost function, but far fewer than the number of endogenous prices. More common. Calls for further assumptions.

I3. Restrict the demand system so that there are only a few free elasticities, perhaps by having price have the same coefficient in each demand equation. Very popular in the context of demand for expensive commodities such as cars, houses, and so on, where price of each good can be interpreted as a lump sum "income not available for other goods."

I4. Use individual consumer micro data, and permit the observable variation between consumers to trace out much of the demand curve for products. Great idea, but calls for data.

I5. Assume a specific model of product differentiated supply, such as Bertrand competition, and use the restrictions between supply and demand to identify the model.

I6. (Variant) Use supply instruments, like the number of products in a category or their "closeness" in product space, loosely derived from a specific model a la I5. This is a popular method

I7. Make a variance-components argument. This is Hausman's approach. He observes the price of the same good in a variety of cities, and puts restrictions on the statistical process driving prices.

I8. Assume that the error in the reduced form for price is uncorrelated with the demand error and estimate by OLS.

Parte II

Custos e Produção

Capítulo 6

Formas Funcionais

Nesta aula, iremos mudar o foco da nossa análise. Até o momento, estávamos preocupados com a modelagem do comportamento do produtor; agora, nos preocuparemos com o comportamento do produtor. Nesta aula, nos preocuparemos inicialmente com a derivação das principais primitivas do comportamento do produtor: a função de produção e a função custos, para apresentar algumas formas funcionais comuns para a modelagem da função custos. Em um momento posterior, nós iremos analisar três dos principais problemas da modelagem do produtor (i) o viés de seleção de amostra, (ii) os problemas de endogeneidade entre a variável dependente e a quantidade de insumos e (iii) a consideração do fato que muitas das empresas podem estar operando abaixo da função de produção (usando a abordagem da função custo, acima da curva de custos).

Em geral, na análise da produção e custos, estamos interessados nos seguintes elementos:

- Escala: Verificação de se a empresa ou o setor exibem retornos constantes de escala, crescentes ou decrescentes;
- Substituição: O grau de substituição dos fatores de produção em resposta a alterações na quantidade produzida;
- Separabilidade: A capacidade de separação das relações de produção – ou de custos – em componentes aninhados ou aditivos.
- Progresso Técnico: Mudança na forma pela qual os fatores de produção são combinados para a produção.
- Distribuição da renda: Como as parcelas da renda se distribuem entre os fatores de produção;
- Custo Marginal: Obtenção de estimativas de custos marginais para as análises subsequentes.

Comecemos, portanto, com o primeiro dos tópicos; a derivação das funções relevantes a partir da teoria microeconômica.

6.1 Formas Funcionais – Produção e Custos

A seguir, vamos fazer uma revisão sobre as formas mais comuns de modelagem da função de produção.

Table 6.1: Funções de Produção

Forma Funcional	Fórmula	Restrições
Cobb-Douglas (Cobb e Douglas 1928)	$\ln y = a_0 + \sum_{j=1}^{J} a_j \ln z_j$	$\sum_{j=1}^{J} a_j = 1$ para homog. lin.
CES (Arrow et. al 1961)	$y^{\rho} = a_0 + \sum_{j=1}^{J} a_j z_j^{\rho}$	$a_0 = 0$ para Homog.lin.
Leontief/Linear Generalizada (Diewert 1971)	$y = a_0 + \sum_{j=1}^{J} a_j \sqrt{z_j} + \sum_{k=1}^{J} \sum_{j=1}^{J} a_{kj} \sqrt{z_k z_j}$	$a_i = 0, i = 0, \cdots J$ para Homog. Lin.
Translog (Christensen, Jorgenson e Lau (1971))	$\ln y = a_0 + \sum_{j=1}^{J} a_j \ln z_j + \sum_{k=1}^{J} \sum_{j=1}^{J} a_{kj} \ln z_k \ln z_j$	$\sum a_j = 1$ e $\sum a_{ij} = 0$para Homog. Lin.
Cobb-Douglas Generalizada (Diewert (1971))	$\ln y = a_0 + \sum_{k=1}^{J} \sum_{j=1}^{J} a_{jk} \ln((z_k + z_j)/2)$	$\sum_k \sum_j a_{jk} = 1$para H. L.
Quadrática (Lau (1974))	$y = a_0 + \sum_{j=1}^{J} a_j z_j + \sum_{k=1}^{J} \sum_{j=1}^{J} a_{kj} z_k z_j$	
Côncava Generalizada (McFadden (1974))	$y = \sum_{k=1}^{J} \sum_{j=1}^{J} z_j \phi^{kj} \left(\frac{z_k}{z_j}\right) a_{kj}$	ϕ^{kj} é uma função côncava conhecida

Todas estas funções podem ser vistas como expansões lineares em parâmetros que podem aproximar uma função arbitrária. Esta expansão pode ser vista na seguinte forma:

$$f^*(\mathbf{x}) \approx f(\mathbf{x}) = \sum_{i=1}^{N} a_i h^i(\mathbf{x})$$

Em que os a_i eram parâmetros, os h^i são funções conhecidas e os $\mathbf{x}$ são vetores de variáveis. Se algumas condições são satisfeitas para uma dada realização do vetor $\mathbf{x}^*$, podemos dizer que $f(\mathbf{x}^*)$ é uma aproximação da função verdadeira no ponto. Além disso, aproxima os valores da primeira e segunda derivadas da função também. Consideramos esta uma forma funcional flexível parsimoniosa.

Um cuidado adicional: quando estamos estimando uma função como esta com uma base de dados com um domínio extensivo – ou seja, com valores que mapeiam muito do quadrante relevante da variável $\mathbf{x}$ – é bem provável que a função obtida não será uma aproximação de segunda ordem da função de produção verdadeira em qualquer ponto. Como resultado, os efeitos de estática comparativa resultantes podem ser bem diferentes dos resultados da função verdadeira. Ou seja, podemos rejeitar uma hipótese mesmo quando a função verdadeira não rejeitaria.

Com relação à função custos, uma função muito utilizada é a Translog. Ela possui a seguinte forma funcional:

$$\ln C = a_0 + \sum_i \beta_i \ln q_i + \sum_j \gamma_j \ln r_j + \sum_j \sum_k \gamma_{jk} \ln r_j \ln r_k + \sum_h \sum_i \beta_{hi} \ln q_h \ln q_i + \sum_i \sum_j \phi_{ij} \ln q_i \ln r_j$$

Em geral, se multiplica as variáveis por 0,5 quando temos os termos quadráticos, para facilitar na hora de montar os sistemas de demanda de fatores. Vamos mostrar adicionalmente como se constrói um sistema de demandas por fatores de produção usando o lema de Sheppard:

$$\frac{1}{C}\frac{\partial C}{\partial r_i} = \gamma_i \frac{1}{r_i} + \sum_k \gamma_{ik} \frac{\ln r_k}{r_i} + \sum_l \phi_{li} \frac{\ln q_l}{r_i}$$

Pelo lema de Sheppard, $\frac{\partial C}{\partial r_i} = z_i$, ou seja:

$$\begin{aligned} \frac{z_i}{C} &= \frac{1}{r_i}\left[\gamma_i + \sum_k \gamma_{ik} \ln r_k + \sum_l \phi_{li} \ln q_l\right] \\ \frac{r_i z_i}{C} &= \left[\gamma_i + \sum_k \gamma_{ik} \ln r_k + \sum_l \phi_{li} \ln q_l\right] \end{aligned}$$

O lado esquerdo da igualdade mostra a participação da remuneração do fator de produção i no total dos custos. Para N fatores de produção, podemos colocar um sistema de $N-1$ equações, impondo as restrições relevantes, para aumentar a eficiência das estimativas.

Vamos agora falar um pouco dos problemas existentes na estimação de funções de produção e custos.

Nesta aula, iremos revisar as principais metodologias de estimação de função custo, com um foco especial na metodologia de Spady e Friedlaender (1978) sobre funções de custo hedônicas, bem como Berry, Kortum e Pakes (1996) sobre o mesmo tema. Inicialmente, vamos discutir um pouco mais sobre a anatomia do termo erro no contexto da estimação de uma função custo.

Podemos, por teoria de dualidade que vimos anteriormente, afirmar que a qualquer especificação de uma função de produção corresponde uma especificação da função custo, supondo comportamento minimizador de custos por parte das empresas envolvidas. Nos anos 70, em geral se afirmava que, na suposição de mercados competitivos para os fatores de produção, a questão da exogeneidade das variáveis independentes parecia resolvida (um ponto de vista como este foi adotado por Varian (1980, cap. 12, pp. 207-9). Desta forma, vamos nos concentrar mais nos procedimentos de análise propriamente ditos do que nas condições de identificação.

6.2 Spady e Friedlaender

A principal questão que este e o próximo artigo irão enfrentar diz respeito ao fato que, em muitos casos, os serviços oferecidos por uma mesma empresa são diferenciados – ou seja, eles dependem não apenas do número de unidades produzidas, mas também das características destes insumos. Por exemplo, um determinado número de minutos de serviço de telecomunucações pode se referir a minutos em horário comercial, e o restante fora do horário comercial. Ou ainda, em termos de Kbps trafegados em uma rede, alguns podem ser de dados e outros de voz. Consequentemente, duas empresas de telecomunicações podem possuir duas estruturas muito diferentes de produtos e de custos se uma se concentra em um conjunto de serviços e outra em outro.

Em geral, estas diferenças são levadas em conta expandindo o vetor de produtos para englobar a dimensão de qualidade (ou seja, Minutos comerciais seriam um produto e Minutos Residenciais seriam outro). O problema é quando você tem uma dimensão de características que os produtos podem tomar que, na prática, é contínua. Neste caso, exigiríamos demais dos dados.

Para resolver este problema, os autores propõem tratar o produto efetivo como sendo uma função de uma medida genérica da quantidade e das suas qualidades. Assim, estimaríamos as chamadas "funções hedônicas de custo". Os autores utilizam uma função de custos separável em qualidade, da seguinte forma:

$$C = C[\psi(y,q), w]$$

Em que $\psi(y,q)$ representa um vetor de funções que mensuram os produtos efetivos e w denota um vetor de preços de fatores de produção. Desta forma, $\psi = [\psi^1, \psi^2, \cdots, \psi^n]$, para uma empresa que oferece n produtos físicos diferentes, é um vetor em que cada elemento $\psi^i = \psi^i(y_i, q_1^i, \cdots, q_r^i)$, sendo que y_i é a quantidade física do produto i e q_r^i é a r-ésima qualidade do produto i. Em especial, eles assumem que:

$$\psi^i = y_i \times \phi(q_1^i, \cdots, q_r^i)$$

Dados estes diferentes serviços, eles estimam uma função custo translog da seguinte forma:

$$\begin{aligned}
\ln C(\psi, w) &= \alpha_0 + \sum_i \alpha_i(\ln \psi_i - \ln \bar{\psi}_i) + \sum_S \beta_s(\ln w_s - \ln \bar{w}_s) + \\
&+ \frac{1}{2}[\sum_i \sum_j A_{ij}(\ln \psi_i - \ln \bar{\psi}_i)(\ln \psi_j - \ln \bar{\psi}_j) + \\
&+ \sum_s \sum_t B_{st}(\ln w_s - \ln \bar{w}_s)(\ln w_t - \ln \bar{w}_t)] + \\
&+ \sum_i \sum_s C_{is}(\ln \psi_i - \ln \bar{\psi}_i)(\ln w_s - \ln \bar{w}_s) + \varepsilon
\end{aligned}$$

Em que as barras representam as médias das variáveis. Além disso os autores colocam no sistema as equações que representam as participações dos fatores no custo total:

$$\frac{w_s x_s}{C} = \beta_s + \sum_t B_{st}(\ln w_t - \ln \bar{w}_t) + \sum_i C_{is}(\ln \psi_i - \ln \bar{\psi}_i)$$

Até agora, o que há de novo é simplesmente a expressão das variáveis independentes em termos de desvios em relação às suas médias. Neste ponto, os autores precisamo expor a sua escolha referente à forma funcional de ψ_i. Eles utilizam uma aproximação translog para esta função, da seguinte forma:

$$\begin{aligned} \ln \psi_i &= \ln y_i + \sum_h a_h^i(\ln q_h^i - \ln \bar{q}_h^i) + \\ &+ \frac{1}{2}\sum_h \sum_l b_{hl}^i(\ln q_h^i - \ln \bar{q}_h^i)(\ln q_l^i - \ln \bar{q}_l^i) \end{aligned}$$

Esta função é substituída no sistema original e ele é estimado por Mínimos Quadrados Generalizados (no paper de Spady e Friedlaender, usaram FIML). Além disso, eles impõem as condições de simetria que estão a seguir:

$$\begin{aligned} \sum \beta_s &= 1 \\ \sum_s B_{st} &= 0, t = 1, \cdots, m; \\ \sum_s C_{is} &= 0, i = 1, \cdots, n \\ B_{ts} &= B_{st} \\ A_{ij} &= A_{ji} \end{aligned}$$

Os custos marginais de cada serviço podem ser estimados com a seguinte fórmula:

$$CMg_i = \frac{C}{y_i}\frac{\partial \ln C}{\partial \ln y_i} = \frac{C}{y_i}\left\{\alpha_i + \sum_{i=1}^{M} A_{im}(\ln y_i - \ln \bar{y}_i) + \sum_{j=1}^{M} C_{ij}(\ln w_i - \ln \bar{w}_i)\right\}$$

Adicionalmente, podemos calcular o grau de economias de escala desfrutadas pela empresa usando a seguinte fórmula:

$$S = \frac{C(y,w)}{\sum_i y_i \frac{\partial C}{\partial y_i}}$$

Se S for maior do que 1, os retornos de escala são crescentes. Se S é menor do que 1, os retornos de escala são decrescentes e se S for igual a 1, os retornos de escala são constantes.

6.3 Berry, Kortum e Pakes

Neste artigo, os autores buscam estimar uma função de custo hedônica com o objetivo de avaliar como as mudanças no ambiente econômico e regulatório afetam os custos de produção e as características econômicas na indústria automobilística. A idéia de utilizar a função de custos hedônicos surgiu do risco de se estimar incorretamente mudanças de produtividade em situações em que apenas as características dos produtos estão se alterando em resposta a mudanças na regulação.

Eles possuíam dados detalhados por planta do setor, bem como informações sobre quais produtos eram feitos em cada uma delas. No entanto, como existiam poucas informações por produto, não dava para estimar diretamente funções custos para cada produto. O foco dos autores foi, inicialmente, sobre a demanda por matérias-primas, pois eles observam que a maior parte dos custos era composto por custos de materiais. Além disso, dos três fatores de produção (capital, mão-de-obra e matérias-primas), apenas a demanda pela terceira poderia ser considerada consistente com o paradigma de minimização de custos estáticos. Finalmente, as estimações para os outros fatores de produção parecem ser ainda problemáticas....

A demanda por materiais que eles estimam é o resultado de várias etapas. Inicialmente, eles consideram uma função de produção que é condicional a um índice arbitrário de mão-de-obra e capital, podendo – ou não – ser diferente de acordo com as várias características do produto. Supondo que este índice também dependa do instante to tempo – denominado t – temos que este índice pode ser representado por $G(L, K, x, t)$, em que $x$denota as características dos produtos. Desta forma, a produção deve ser igual a um coeficiente fixo multiplicado pelo uso dos materiais. Os índices são j para o modelo de automóvel, p para planta e t para o instante do tempo.

A demanda por materiais é, desta forma, um coeficiente constante multiplicado pela produção. Os autores denotam este coeficiente por $c(x_j, \epsilon_{pt}, \beta)$, como sendo uma função das características dos produtos – denominadas x_j– , um choque de produtividade específico da planta ϵ_{pt} e um vetor de parâmetros a serem estimados. Supondo que x_j e β entrem linearmente na função, podemos escrever:

$$c = x_j\beta + \epsilon_{pt}$$

Além disso, os autores consideram um choque específico de produtividade no tempo, denominado δ_t. Isto tudo implica que a função de produção é igual a:

$$Q_{jpt} = \min\left(\frac{M}{\delta_t c(x_j, \epsilon_{pt}, \beta)}, G(L, K, x, t)\right)$$

Supondo comportamento minimizador de custos decorrentes dos custos variáveis, esta função de produção Leontief implica a seguinte demanda por matérias-primas:

$$M_{jpt} = \delta_t c(x_j, \epsilon_{pt}, \beta) Q_{jpt}$$

Esta função Leontief implica que os custos variáveis médios são constantes. No entanto, os retornos crescentes são modelados por meio da divisão do custo fixo entre diferentes unidades. Supondo que a exigência de materiais seja denotada por μ, e que, para produzir mais de um produto em uma planta, exista um custo fixo igual a Δ por produto. Sendo J_p o conjunto de modelos produzido em uma planta p, o que dá uma demanda total de fatores de produção igual a:

$$M_{pt} = \mu + J_{pt}\Delta + \sum_{j \in J(p)} M_{jpt}$$

Dividindo esta equação pela produção de cada planta, temos que:

$$\frac{M_{pt}}{Q_{pt}} = \mu \frac{1}{Q_{pt}} + \Delta \frac{J_{pt}}{Q_{pt}} + \delta_t(\bar{x}_{pt}\beta + \epsilon_{pt})$$

Sendo que

$$\bar{x}_{pt} = \sum_{j \in J_p} x_{jpt} \frac{Q_{jpt}}{Q_{pt}}$$

Exceto pelos δ_t, a equação acima poderia ser estimada por OLS. Com as dummies temporais, é possível estimar por Mínimos Quadrados Ordinários não lineares.

6.4 Evans e Heckman

O texto destes autores, parte de uma coletânea dedicada à revisar os estudos econômicos feitos no contexto da ação de separação estrutural do Bell System no começo dos anos 80 do século XX, tem por objetivo investigar uma das condições necessárias e suficientes para que uma empresa possa ser considerada um Monopólio Natural: a sub-aditividade da função custos. Para atender a este objetivo, os autores apontam para o fato que estimações de função custos baseadas em uma medida agregada de produto não necessariamente levam a resultados corretos do ponto de vista de economias de escala. Apenas sob algumas premissas – que os autores afirmam que não se verificam no contexto do Bell System no período em questão – as conclusões obtidas a partir de um modelo em que a medida de produto é um agregado das medidas de produtos individuais são válidas.

Além disso, Evans e Heckman[24][1] afirmam que, mesmo que uma estimativa de economias de escala com o uso de medidas agregadas de produto seja correta, isso não necessariamente implica que todos os produtos ou serviços componentes deste agregado também tenham economias de escala.

[1] Este artigo original deu margem a dois outros, publicados na American Economic Review (Evans e Heckman (1984 e 1986)[25, 26]). No presente texto, iremos nos ater a Evans e Heckman (1983)[24].

O primeiro passo de sua análise é assumir que a função custo da AT&T é dada por

$$C = f(L, T, r, m, w, t)$$

Em que L denotava a quantidade de chamadas locais produzidas, T a quantidade de chamadas a longa distância produzidas, w o salário médio, r o custo do capital, m o preço das matérias-primas e t um índice de mudança tecnológica. A idéia desta separação entre produtos é baseada que, à época, estes eram os principais produtos oferecidos pela empresa. Inicialmente eles começam estimando uma função translog de custos da seguinte forma:

$$\begin{aligned}
C &= \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln p_i + \sum_{k=1}^{M} \beta_k \ln q_k + \mu \ln t \\
&+ \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij} \ln p_i \ln p_j + \sum_{k=1}^{M} \sum_{l=1}^{M} \delta_{kl} \ln q_k \ln q_l \right] + \\
&+ \sum_{i=1}^{N} \sum_{k=1}^{M} \rho_{ik} \ln p_i \ln p_k + \sum_{i=1}^{M} \lambda_i \ln p_i \ln t + \sum_{k=1}^{N} \theta_k \ln q_k \ln t + \tau (\ln t)^2
\end{aligned}$$

Além disso, eles colocam um sistema associado de participações dos insumos nos custos totais, com a seguinte forma funcional:

$$S_i = \frac{p_i x_i}{C} = \alpha_i + \sum_i \gamma_{ij} \ln p_j + \sum_k \rho_{ik} \ln q_k + \lambda_i \ln t$$

Além disso, eles investigam dois pontos adicionais. O primeiro deles é se a função de produção é separável em termos dos produtos e o segundo é se a função de custos é aditiva. A primeira hipótese, separabilidade em termos dos produtos, se for verdade, implica que a decisão de montante a ser produzido pode ser separada da decisão da composção da quantidade. Se ela for válida, podemos agregar as quantidades de L e T em um único índice, da seguinte forma:

$$C = f(L, T, r, m, w, t) = f(A(L, T), r, m, w, t) = f(Q^*, r, m, w, t)$$

Sendo que $Q^* = A(L, T)$. Para que isto valha, devemos ter $\rho_{ik}\beta_l = \rho_{il}\beta_k, \forall i, k \neq l$. A segunda hipótese testável, a de aditividade, é que os custos de produzir todos os produtos em um mesmo lugar é igual à soma dos custos de produzir em unidades separadas, ou seja:

$$C = f(L, T, r, m, w, t) = C_L(L, r, m, w, t) + C_T(T, r, m, w, t)$$

Esta aditividade pode ser testada pela seguinte hipótese, na função translog: $\delta_{kl} = -\beta_k \beta_l$. Se esta restrição for válida, a empresa não possui nem economias nem deseconomias de escopo.

Finalmente, eles buscam testar a existência de economias de escopo. O teste canônico de economias de escopo foi proposto por Panzar (1989)[36] e é da seguinte forma. Seja $\hat{T}$ um subconjunto do espaço de produtos $(\hat{T}, T)$, no ponto em que a quantidade produzida é y. O grau de economias de escopo, neste caso, é dado por:

$$SC = \frac{C(y_T) + C(y_{\hat{T}}) - C(y)}{C(y)}$$

Em que y_T é a quantidade produzida do subconjunto de produtos T, $y_{\hat{T}}$ a quantidade produzida do subconjunto de produtos $\hat{T}$. A função custo apresentaria economias de escopo se SC for maior do que zero. Além disso, no contexto da análise dos autores, temos que além de economias de escala, é necessário investigar a existência de economias de escopo. A existência de economias de escopo é uma condição necessária, mas não suficiente, para a sub-aditividade na função custos. Ou seja, se não existissem as economias de escopo, duas empresas cada uma produzindo um dos serviços, o fariam a custos menores do que uma única fazendo os dois produtos.

No entanto, esta metodologia "canônica" apresenta problemas quando aplicada para a função translog. Com os coeficientes em questão, a gente somente poderia calcular os valores de $C(y_T)$ e $C(y_{\hat{T}})$ se supuséssemos os valores de algumas quantidades iguais a zero. No entanto, como a simples inspeção da fórmula acima pode nos mostrar, temos que este tipo de função não é definida nos pontos em que os argumentos são iguais a zero.

Evans e Heckman seguem um caminho diferente, testando uma hipótese mais simples, se é mais econômica a produção dos dois bens em conjunto dentro de uma empresa – no caso, a AT&T – ou em duas empresas separadas. O estudo que eles fazem é o seguinte. Considere $\mathbf{Q}_t^* = (Q_{1t}^*, Q_{2t}^*)$ como sendo o vetor de bens produzidos em um determinado ano t. Além disso, considere também o vetor $\mathbf{Q}_M = (Q_{1M}, Q_{2M}) = (\min Q_{1t}^*, \min Q_{2t}^*)$, que são os menores valores dos dois produtos disponíveis na amostra. Vamos considerar duas empresas hipotéticas, A e B, que possuem as seguintes quantidades produzidas:

$$\begin{aligned} \mathbf{Q}_{At} &= (\phi Q_{1t} + Q_{1M}, \omega Q_{2t} + Q_{2M}) \\ \mathbf{Q}_{Bt} &= ((1-\phi) Q_{1t} + Q_{1M}, (1-\omega) Q_{2t} + Q_{2M}) \end{aligned}$$

Neste caso, temos que ,ω e ϕ pertencem am intervalo $[0, 1]$, pois a quantidade observada seria igual à produção das duas empresas hipotéticas[2]. Restringindo o domínio na função aos valores observados, temos que as duas empresas, juntas, produzem:

$$\begin{aligned} Q_{1t} + 2Q_{1M} &= Q_{1t}^* \\ Q_{2t} + 2Q_{2M} &= Q_{2t}^* \end{aligned}$$

[2]A intuição dos autores é evitar conclusões para configuração de produção fora da amostra.

Finalmente, considere as seguintes definições:

$$
\begin{aligned}
C_{At}(\phi,\omega) &= C(\mathbf{Q}_{At}) \\
C_{Bt}(\phi,\omega) &= C(\mathbf{Q}_{Bt}) \\
C_t^* &= C(\mathbf{Q}_t^*)
\end{aligned}
$$

Estes são os níveis de custos para as duas empresas, A e B, com os coeficientes estimados anteriormentes, e a quantidade produzida para cada empresa hipótetica de cada um dos produtos. C_{At} e C_{Bt} são funções de ϕ e ω por $\mathbf{Q}_{At}$ e $\mathbf{Q}_{Bt}$ são construídos por intermédio destes coeficientes. A sub-aditividade vai ser obtida calculando a seguinte estatística para cada um dos produtos:

$$
Sub(\phi,\omega) = \frac{C_t^* - C_{At}(\phi,\omega) - C_{Bt}(\phi,\omega)}{C_t^*}
$$

Podemos notar que esta estatística ainda é uma função de ϕ e ω. Neste caso, o que os autores fazem é estimar o valor desta estatística para todos os parâmetros possíveis ϕ e ω[3], e se o máximo deste valor for negativo (afinal de contas, este negócio é igual a (-1) multiplicando a estatística SC anterior) e significativo[4], podemos dizer que existe economias de escopo e sub-aditividade da função de custos.

Em seu artigo, Evans e Heckman (1983) indicam que, em nenhum instante do período de tempo cujos dados permitem a análise – entre 1958 e 1977 – os resultados se mostraram significantes e negativos, rejeitando a hipótese de existência de economias de escala.

Em 1991, os resultados de Evans e Heckman se viram sob crítica de Diewert e Wales (1991)[23], que apontaram um problema com as estimativas da função translog de custos – em especial, a função de custos que os autores estimam é decrescente na quantidade produzida para um intervalo de dados da amostra, colocando em dúvida os resultados dos autores sobre sub-aditividade.

6.5 Mensuração da Eficiência

Nesta aula, iremos analisar a literatura referente à mensuração da eficiência econômica, com especial foco em duas técnicas: a primeira delas, denominada *Data Envelopment Analysis*, busca investigar a eficiência econômica com base em métodos não paramétricos, enquanto a outra metodologia, *Stochastic Frontier Analysis*, busca investigar a eficiência econômica com métodos paramétricos. Comecemos discutindo o primeiro dos elementos.

[3]Na prática, em um "grid" de pontos com diferença de 0,1 entre eles.

[4]Os autores não falam como eles derivam os erros-padrão desta estatística. No entanto, o correto seria fazer uma apliação do Método Delta discutido ao final do livro.

6.6 Métodos Não Paramétricos

A idéia básica envolvida neste tipo de estimação é o conjunto de possibilidades de produção, que consiste no envelope convexo de vetores insumo-produto. Este conjunto, segundo Farrell (1957), era representado por meio da isoquanta unitária. De acordo com esta especificação, nenhuma forma funcional específica precisava ser pré-definida.

O objetivo principal desta abordagem era definir uma superfície envoltória para todas as observações da amostra. Esta superfície é determinada pelas unidades que estejam na fronteira e que, por conseguinte, são carimbadas como eficientes. As que estão para dentro são consideradas ineficientes. Neste caso, não há espaço para elementos aleatórios, e pode ser considerada como uma técnica não estatística (pelo menos em sua versão mais simples), no sentido que as medidas de ineficiência e a fronteira eficiente são calculadas ao invés de estimadas.

Para calcular esta fronteira eficiente, é necessária a seleção de um 'grupo de pares', que consiste em uma combinação linear das diferentes empresas, que são os vértices da isoquanta unitária eficiente. Considerando um conjunto de N empresas, caracterizada por um vetor de produção com m insumos e s produtos, para cada unidade não localizada na fronteira podemos definir um vetor $\bar{\mu} = (\mu_1, \mu_2, \cdots, \mu_N)$, sendo que cada um dos μ_j representa o peso desta unidade dentro do grupo de pares. O cálculo envolvido na Análise Envoltória de Dados tem por objetivo maximizar o *efficiency score* de cada grupo, sujeito à restrição que o conjunto de pesos obtidos desta maneira seja factível para todas as empresas. Isto implica os seguinte problema de otimização:

$$\begin{aligned} TE_{CRS} &= \min_{\mu} \psi^0 \\ s.t. & \\ \sum_{j=1}^{N} \mu_j X_{ij} &\leq \psi X_i^0, \forall i = 1 \cdots m \\ \sum_{j=1}^{N} \mu_j Y_{rj} &\geq Y_r^0, \forall r = 1 \cdots s \end{aligned}$$

A solução deste problema de minimização nos dá, para o grupo analisado, um subgrupo que é o grupo de pares e que, em relação às empresas analisadas, dá pelo menos o mesmo nível de produção – segunda restrição – mas consomem apenas uma fração ψ dos insumos das empresas analisadas – primeira restrição. O objetivo final é determinar a combinação linear dos referentes, tal que, para cada empresa analisada, seja minimizado o valor de . Os *efficiency scores* são determinados pelo ψ ótimo de solução do problema. Quando estamos mencionando retornos variáveis de escala, colocamos uma condição adicional:

$$
\begin{aligned}
TE_{CRS} &= \min_{\mu} \psi^0 \\
s.t. & \\
\sum_{j=1}^{N} \mu_j X_{ij} &\leq \psi X_i^0, \forall i = 1 \cdots m \\
\sum_{j=1}^{N} \mu_j Y_{rj} &\geq Y_r^0, \forall r = 1 \cdots s \\
\sum_{j=1}^{N} \mu_j &= 1
\end{aligned}
$$

6.7 A Abordagem Paramétrica

Em termos de uma função de produção definida em termos de uma cross-section, uma fronteira paramétrica pode ser definida como:

$$Y_i = f(X_i, \beta) \times TE_i$$

Em que i representa os produtores, Y_i a quantidade produzida, X_i a quantidade dos insumos, β os coeficientes associados e TE_i a medida de eficiência técnica orientada a produto, e definida como:

$$TE_i = \frac{Y_i}{f(X_i, \beta)}$$

Na abordagem de Farrell (1957), ele adotou a forma $TE_i = \exp(-u_i)$. Note que, neste momento, não fizemos nenhuma imposição de elemento aleatório, fazendo com que esta fronteira de produção seja determinística. Com a imposição de uma forma funcional à la Cobb-Douglas, a forma funcional fica sendo:

$$\ln Y_i = \beta_0 + \sum_r \beta_r X_{ri} - u_i$$

Aigner, Lovell e Schmidt (1977) fizeram a análise econométrica canônica deste tipo de modelo, incorporando o termo u_i em um termo erro composto, $e_i = v_i - u_i$, em que v_ié uma parte aleatória independente e identicamente distribuída. Neste caso, o modelo completo fica sendo:

$$\ln Y_i = \beta_0 + \sum_r \beta_r X_{ri} + v_i - u_i$$

Supondo que o o termo de ineficiência seja com uma distribuição normal pela metade[5], podemos escrever a função verossimilhança, com a suposição que os dois termos erros são distribuidos independentemente. A partir da estimativa, podemos decompor o valor esperado da ineficiência do termo erro composto, da seguinte forma:

$$E(u_i|e_i) = \frac{\sigma\lambda}{(1+\lambda^2)}\left[\frac{\phi(\frac{e_i\lambda}{\sigma})}{\Phi(-\frac{e_i\lambda}{\sigma})} - \frac{e_i\lambda}{\sigma}\right]$$

Em que ϕ é a pdf normal, a Φ a cdf normal, $\lambda = \frac{\sigma_u}{\sigma_v}$, e $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$. Neste caso, $TE_i = 1 - E(u_i|e_i)$. Outras alternativas foram propostas, mas agora nós mudaremos o nosso foco para incorporarmos dados em painel.

6.7.1 Dados em Painel

Com a dispoibilidade de dados em painel, algumas vantagens surgem. Em primeiro lugar, não necessariamente o termo erro precisa ser exógeno em relação às quantidades dos insumos, como no caso de cross-sections. Além disso, ao adicionar informações ao longo do tempo de uma mesma empresa, isso permite que tenhamos estimativas consistentes da ineficiência. Finalmente, a terceira vantagem está no fato que, em alguns casos não é necessária a imposição de hipóteses distribucionais sobre o termo erro; simplesmente as técnicas de efeitos fixos e aleatórios nos permitem obter estimativas dos parâmetros da função de produção.

Para entender este ponto melhor, considere a função de produção como vimos anteriormente:

$$\ln Y_{it} = \beta_0 + \sum_r \beta_r \ln X_{rit} + v_{it} - u_{it}$$

Sendo que $u_{it} > 0$. Usando a transformação within, e se supusermos que o termo u_{it} seja sistemático e não varie ao longo do tempo, podemos estimar os coeficientes da função de produção:

$$(\ln Y_{it} - \ln \bar{Y}_{it}) = \sum_r \beta_r (\ln X_{rit} - \ln \bar{X}_{rit}) + v_i$$

Podemos definir os termos de ineficiência individual por meio das seguintes definições:

$$\begin{aligned} \hat{\beta}_0 &= \max(\hat{\beta}_{0i}) \\ u_i &= \hat{\beta}_0 - \hat{\beta}_{0i} \end{aligned}$$

Em que os $\hat{\beta_{0i}}$ são obtidas como:

$$\hat{\beta_{0i}} = (\ln Y_{it} - \ln \bar{Y}_{it}) - \sum_r \beta_r (\ln X_{rit} - \ln \bar{X}_{rit})$$

[5]Distribuições comumente utilizadas são a exponencial, e a truncada por baixo em zero.

Evidentemente, este tipo de análise impede que utilizemos dados que são constantes ao longo do período amostral. Uma solução para este problema é tratar a parte de ineficiência como sendo um efeito aleatório, podemos estimar o modelo da seguinte forma:

$$\ln Y_{it} = (\beta_0 - E(u_i)) + \sum_r \beta_r \ln X_{rit} + v_{it} + (u_{it} - E(u_{it}))$$

Em que tanto v_{it} e $u^*_{it} = u_{it} - E(u_{it})$ possuem média zero. Este modelo pode ser estimado por Mínimos Quadrados Generalizados. Note-se que, para a estimação da eficiência técnica, não precisamos impor nenhuma hipótese distribucional sobre o termo erro. No entanto, a estimação por efeitos aleatórios precisa, para a estimação consistente da eficiência técnica, que a dimensão tempo e cross-section da amostra tenda ao infinito. Além disso, esta estimação poe efeitos aleatórios pressupõe que a ineficiência seja exógena à escolha das quantidades de insumos. Ainda sobre este problema da consistência, a imposição de constância na eficiência técnica ao mesmo tempo que exija que a amostra tenda ao infinito é um pouco inconsistente, demandando o desenvolvimento de modelos em que a eficiência seja alterada ao longo do tempo.

A primeira linha de ação adotada foi a de Cornwell, Schmidt e Sickles, que remonta o modelo acima da seguinte forma:

$$\ln Y_{it} = (\beta_{0t} - u_{it}) + \sum_r \beta_r \ln X_{rit} + v_{it}$$

Os autores modelam os interceptos como uma função quadrática do tempo, em que o tempo está associado com características específicas dos produtores, no caso Γ:

$$u_{it} = \Gamma_{1i} + \Gamma_{2i} t + \Gamma_{3i} t^2$$

Os autores inclusive propôem inclusive uma forma de estimativa por Variáveis Instrumentais deste negócio.

A outra grande linha de ação passa pelo método da Máxima Verossimilhança. Segundo Kumbhakar (1990), supondo que a ineficiência se altere ao longo do tempo de acordo com a seguinte especificação:

$$u_{it} = \delta_t u_i$$

E se assumirmos uma distribuição half normal para o termo u_i, podemos modelar u_{it} da seguinte forma:

$$u_{it} = [1 + \exp(\gamma t + \rho t^2)]^{-1} \times u_i$$

Em que γ e ρ são parâmetros a serem estimados. Uma linha alternativa de ataque é a de Battese e Coelli (1992), em que apenas um parâmetro deve ser estimado:

$$u_{it} = [\exp(-\eta(t - T)] \times u_i$$

6.7.2 Teoria da Dualidade

Da mesma forma que podemos modelar a eficiência técnica de uma forma direta – por meio da função de produção – podemos modelar a ineficiência técnica pelo dual do problema de otimização da empresa (a sua função custos). Podemos basear a nossa escolha em vários fatores, tais como a disponibilidade de dados, características do setor e hipóteses de exogeneidade. Esta última motivação para a escolha da função custos se baseia no ponto discutido na aula anterior de que, na premissa de mercados perfeitamente competitivos para os fatores de produção, poderíamos construir um caso mais robusto para a exogeneidade das variáveis independentes da relação. Comecemos a discutir esta abordagem. Assumindo uma função Cobb-Douglas tradicional:

$$\ln Y_{it} = A + \sum_r \beta_r \ln X_{it} + (v_{it} - u_{it})$$

Podemos escrever o problema dual ao da maximização de lucros como o de minimização da seguinte função custos:

$$\ln C_{it} = \alpha + \frac{1}{\sum_r \beta_r} \ln Y_{it} + \sum_r \frac{\beta_r}{\sum_r \beta_r} \ln w_r - \frac{1}{\sum_r \beta_r}(v_{it} - u_{it})$$

Neste caso, a ineficiência tem o sinal reverso: uma empresa ineficiente e que, portanto, opera abaixo de sua função de produção, terá necessariamente que operar acima da sua curva de custos. O custo desta eficiência técnica é dado por $\frac{1}{\sum_r \beta_r} u_{it}$, que é o custo adicional além do implícito pela curva de custos. No entanto, esta função de custos derivada da Cobb-Douglas impõe restrições muito fortes sobre a separabilidade das decisões da empresa. Neste caso, uma solução interessante é o uso das formas funcionais flexíveis que vimos anteriormente. Em especial, Greene (1980) usa uma função translog com sistema associado de participações dos insumos nos custos totais:

$$\begin{aligned}
\ln C &= \beta_0 + \beta_Y \ln Y + \frac{1}{2}\gamma_{YY}(\ln Y)^2 + \\
&+ \sum_{n=1}^{N} \beta_n \ln w_n + \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \gamma_{nm} \ln w_n \ln w_m + \\
&+ \sum_{n=1}^{N} \gamma_{Nm} \ln Y \ln w_n + \varepsilon_c \\
S_n &= \beta_n + \gamma_{Yn} \ln Y + \sum_{n=1}^{N} \gamma_{mn} \ln w_m + \xi_c
\end{aligned}$$

Neste caso, supondo que as empresas sejam alocativamente eficientes (ou seja, dados preços relativos, encontram o ponto na isoquanta que minimiza os custos), ε_c mensura o custo da ineficiência técnica; no entanto, a relação entre este termo e o termo ξ_c é bastante obscura. Em um contexto em que a função de produção é homotética, e existe ineficiência alocativa (a empresa produz na isoquanta associada, mas não no ponto de mínimo custo dados os preços relativos dos fatores de produção), podemos relacionar os dois termos erros da seguinte forma: ε_c representaria o custo da ineficiência técnica e alocativa, enquanto que ξ_c representaria o custo da ineficiência alocativa. Ainda assim, as coisas não são claras, sendo que as soluções existentes claramente não são suficientes para resolver este dilema.

Capítulo 7

Identificação

7.1 Estimação de Funções de Produção e de Custos – Problemas

A estimação das funções de produção começou com o trabnalho de Cobb e Douglas (1928), que buscavam testar as implicações da teoria da distribuição baseada na produtividade marginal dos fatores. A principal crítica deste tipo de literatura é que os dados sobre fatores de produção, quando estamos falando em quantidades agregadas, são determinados simultaneamente aos valores do produto; desta forma, a função de produção não seria identificável. Vamos ilustrar este ponto mais detalhadamente. Vamos considerar a seguinte equação:,

$$q = \alpha z + \beta x + u$$

Em que q é o log da quantidade produzida, z é o log do capital (ou qualquer outra quantidade de fatores "fixos" de produção), e x o log de todos os insumos variáveis. O problema, levantado já em 1944 por Marschak e Andrews, é que não podemos tratar x e z com verdadeiramente exógenos e estimar este negócio por OLS. Mesmo que imaginemos que z foi "pré-determinado", ou seja, escolhido pela empresa em um instante anterior e fora do controle da empresa hoje, os x podem ser ajustados pela empresa.

A principal resposta a este problema foi a mudança de foco da análise para dados microeconômicos, específicos à empresa. Aparentemente, estes problemas anteriormente mencionados seriam de menor importância, pois poderíamos afirmar que muito do que está contido em u, por exemplo, o clima e pragas não seria afetado pelas escolhas do operário, e muito do que está do lado direito da equação seria determinado anteriormente. Ainda que, para alguns casos como os de fazendas, isto poderia ser verdade, claramente não o era para o caso de empresas industriais.

Vamos começar a discutir como este problema foi levado em consideração, decompondo o termo

u acima da seguinte forma:

$$u_{it} = a_{it} + e_{it} + \varepsilon_{it}$$

Em que i denota a empresa, a_{it} e e_{it} denotam os fatores que são conhecidos pela empresa e não pelo econometrista, enquanto ε_{it} denota a parte deste termo erro devido a erros de medida e de coleta de dados. Além disso, vamos supor que a_{it} seja conhecido imediatamente de forma a afetar as decisões atuais de x, enquanto que e_{it} somente depois da decisão de escolha de x (por exemplo, condições incomuns de tempo e mudanças no esforço produtivo decorrentes de alterações na demanda). Note-se que, apesar de não afetar a escolha corrente dos fatores de produção, e_{it}ainda afeta a escolha futura destes fatores de produção. Além disso, a_{it} será transmitido com atraso para a decisão de z.

Uma das formas pelas quais o problema da simultaneidade foi contornado é com a ajuda de dados em painel. Neste caso, podemos reescrever a função de produção como sendo – ignorando os erros de medida:

$$q_{it} = \alpha z_{it} + \beta x_{it} + a_i + \lambda_t + e_{it}$$

Neste caso, a_i representa efeitos individuais e λ_t os efeitos tempo. Note-se que a notação a_i, ao ignorar o subscrito t, assume que os componentes a são constantes ao longo do tempo. Supondo que os componentes λ_t estão incorporados em z_{it}, com a transformação de efeitos fixos elimina os componentes de a.

Apesar deste problema do termo a ter sido eliminado – em parte – com este trabalho, ainda assim temos problemas remanescentes. Isso se verifica no caso em que a estimação deste tipo de modelo geralmente leva a estimativas baixas e pouco significativas para o capital e baixas estimativas de retornos de escala. Um dos problemas levantados por Chamberlain em (1982) é que a colocação de efeitos fixos individuais acaba trazendo o termo e para dentro das variáveis. Vamos considerar isso da seguinte forma. A transformação "within" na equação acima mostra o seguinte:

$$(q_{it} - \bar{q}_i) = \alpha(z_{it} - \bar{z}_i) + \beta(x_{it} - \bar{x}_i) + (e_{it} - e_i)$$

Em que os termos barra são as médias do período. O problema é que, agora, temos as médias do lado direito da igualdade, o que implica que deveríamos ter x e z não apenas predeterminados em relação a e (ou seja, os seus valores correntes são independentes das realizações atuais de e; e os valores futuros de x e z são dependentes de e) mas sim estritamente exógenos.

Para resolver estes problemas, foram propostas duas alternativas: primeira diferença e/ou estimação em GMM. A primeira diferença da equação acima implica:

$$q_{it} - q_{it-1} = \alpha(z_{it} - z_{it-1}) + \beta(x_{it} - x_{it-1}) + (a_{it} - a_{it-1}) + e_{it} - e_{it-1}$$

Supondo a mesma coisa que antes, que o termo a seria constante ao longo do tempo, poderíamos estimar esta equação por Mínimos Quadrados Ordinários **se** e_{it} e e_{it-1} fossem independentes de tanto x_{it}, x_{it-1}, quanto z_{it}, z_{it-1} (pois esta diferença entre e_{it} e e_{it-1} vai cair no termo erro). Agora, se o termo a não seja constante ao longo do tempo, ou não possamos garantir a estrutura de independência entre os e e os x, esta diferença entre os x precisa ser instrumentada. A idéia dos instrumentos relevantes seria utilizar os x e z defasados pra mais de $t-1$. Esta é a essência do GMM em diferenças utilizado por Arellano e Bond (1991).

O problema é que os instrumentos disponíveis tendem a ser fracos. Em geral, as variáveis como x e z evolvem como passeios aleatórios. Neste caso, os valores defasados de x e z são pouco informativos como instrumentos para as diferenças das variáveis. Uma das formas pelas quais esta questão foi resolvida foi colocar equações adicionais em nível, de forma a aumentar o poder dos instrumentos, que é a abordagem de Blundell e Bond (1998). Na próxima aula falaremos de outras abordagens para resolver o problema da simultaneidade.

Nesta aula, iremos apresentar duas das mais modernas formas de estimação dos parâmetros de funções de produção na literatura. A primeira, devida a Olley e Pakes (1996), que busca lidar com os problemas da endogeneidade e seleção de amostra utilizando dados sobre investimentos, e a segunda, associada com Levinsohn e Petrin (2000), que lida com estes dois problemas usando dados de consumo de insumos intermediários. As duas abordagens se baseiam em bases de microdados de empresas.

7.2 Olley e Pakes (1996)

Os dois autores começam o seu estudo afirmando que, para a obtenção de estimativas consistentes dos parâmetros de funções de produção dois problemas inter-relacionados precisam ser resolvidos: um problema de seleção gerado pela relação entre a variável não observada de produtividade e a decisão de saída do mercado, e um problema de simultaneidade gerado pela relação entre produtividade e demandas de fatores.

Neste caso, os autores propõem um algoritmo para a estimação de funções de produção que leva em consideração explicitamente o viés de seleção decorrente da decisão de falência, bem como o problema de simultaneidade. O primeiro passo do algoritmo é estimar a seguinte regra de acumulação para o capital:

$$\begin{aligned} k_{t+1} &= (1-\delta)k_t + i_t \\ a_{t+1} &= a_t + 1 \end{aligned}$$

Sendo k_t o capital, δ a depreciação, e i_t o investimento em capital, e a_t a idade da empresa. Além disso, os autores assumem uma regra de evolução para a produtividade, denominada ω_t, que seguiria um processo Markoviano de ordem 1. Além disso, é assumido que a empresa maximiza o valor presente dos seus lucros, o que nos dá a seguinte equação de Bellman:

$$V_t(\omega_t, a_t, k_t) = \max\{\Phi, \sup \pi_t(\omega_t, a_t, k_t) - c(i_t) + \beta E[V_{t+1}(\omega_{t+1}, a_{t+1}, k_{t+1})|J_t\}$$

Em que Φ representa o valor terminal da planta, π_t representa os lucros da empresa como função das variáveis de estado (produtividade, idade e capital - supõe-se que o trabalho se ajuste instantaneamente às mudanças nestas variáveis), e $c(i_t)$ representa o custo do ajustamento do estoque de capital – denominado investimento. Enquanto isso,β é o coeficiente de desconto intertemporal da empresa e J_t é o conjunto de informações disponível no instante t.

O que esta relação quer dizer é que, se o valor presente dos lucros da empresa, descontados adequadamente, for menor do que o valor terminal dos ativos, ela decidirá sair do mercado e fechar a empresa. Caso isso não seja verdade, e ela decida manter-se no mercado, ela acabará por estabelecer um nível de investimento consistente com a operação continuada da empresa. Dadas as hipóteses sobre o comportamento do termo de produtividade – que, por seguir um processo Markoviano de ordem 1, incorpora os choques aleatórios – podemos gerar uma regra de saída de mercado e uma demanda por investimentos. A regra de saída é:

$$\chi_t = \begin{cases} 1 & if \omega_t \geq \underline{\omega}_t(a_t, k_t) \\ 0 & c.c. \end{cases}$$

E podemos construir uma função investimentos da seguinte forma:

$$i_t = i_t(\omega_t, a_t, k_t)$$

Segundo Pakes (1994, Teorema 27), temos que esta função, para quaisquer valores do par (a_t, k_t) é crescente em ω_t. Desta forma, definindo $h = i^{-1}$ a função inversa da anterior, podemos escrever:

$$\omega_t = h_t(i_t, a_t, k_t)$$

Ou seja, podemos escrever a produtividade como uma função do investimento, da idade e do estoque de capital. Com isso, podemos retornar à nossa função de produção. Supondo que ela seja Cobb-Douglas nos seus argumentos, podemos fazer o seguinte:

$$\ln(y_{it}) = \beta_0 + \beta_a \ln(a_{it}) + \beta_k \ln(k_{it}) + \beta_l \ln(l_{it}) + \omega_{it} + \eta_{it}$$

Substituindo a função produtividade em termos de i_t, a_t, k_t na função de produção acima, temos que:

$$\ln(y_{it}) = \beta_l \ln(l_{it}) + \phi_t(i_{it}, a_{it}, k_{it}) + \eta_{it}$$

Uma vez que a colocação da função ω_t vai absorver os logs da idade, do investimento e do estoque de capital. Ou seja:

$$\phi_t(i_t, a_t, k_t) = \beta_0 + \beta_a \ln(a_{it}) + \beta_k \ln(k_{it}) + h(i_t, a_t, k_t)$$

Aqui podemos fazer **o primeiro passo da metodologia de estimação de Olley-Pakes**. Em especial, esta metodologia consiste em fazer uma regressão de $\ln(y_{it})$ contra $\ln(l_{it})$ e um polinômio de terceira ou quarta ordem em (i_t, a_t, k_t). Ou seja, estimamos a seguinte regressão:

$$\begin{aligned} \ln(y_{it}) &= \beta_0 + \beta_l \ln(l_{it}) + \sum_{x=1}^{3} \gamma_{1x}(\ln(k_{it}))^x + \sum_{x=1}^{3} \gamma_{2x}(\ln(i_{it}))^x + \sum_{x=1}^{3} \gamma_{3x}(\ln(a_{it}))^x + \\ &+ \sum_{x=1}^{3} \gamma_{4x}(\ln(k_{it}) \times \ln(a_{it}))^x + \sum_{x=1}^{3} \gamma_{5x}(\ln(k_{it}) \times \ln(i_{it}))^x + \sum_{x=1}^{3} \gamma_{6x}(\ln(i_{it}) \times \ln(a_{it}))^x \end{aligned}$$

Com isto aqui, conseguimos estimar consistentemente o coeficiente β_l, que é o de ajustamento imediato. No entanto, precisamos ainda dos coeficientes dos outros elementos. Para estimarmos os coeficientes de β_a e β_k, precisamos das estimativas de β_l, bem como de $\hat{\phi}$ e das probabilidades de sobrevivência da empresa (ou seja, estar no mercado, dado que esteve no período anterior). Definindo esta variável por χ_{t+1}, podemos escrever um probit da seguinte forma:

$$\begin{aligned} Prob(\chi_{t+1} = 1) &= \sum_{x=1}^{3} \gamma_{1x}(\ln(k_{it}))^x + \sum_{x=1}^{3} \gamma_{2x}(\ln(i_{it}))^x + \sum_{x=1}^{3} \gamma_{3x}(\ln(a_{it}))^x + \\ &+ \sum_{x=1}^{3} \gamma_{4x}(\ln(k_{it}) \times \ln(a_{it}))^x + \sum_{x=1}^{3} \gamma_{5x}(\ln(k_{it}) \times \ln(i_{it}))^x + \sum_{x=1}^{3} \gamma_{6x}(\ln(i_{it}) \times \ln(a_{it}))^x \end{aligned}$$

A partir daí, são calculados os valores previstos da probabilidade de sobrevivência. Caso tenhamos uma base de dados balanceada, este estágio não seria necessário. Com este modelo, são estimadas as probabilidades de ocorrência daquele evento, o que denominaremos $\hat{P}_t$.

Finalmente, podemos estimar os coeficientes do capital e da idade. Em especial, isto é conseguido por meio de estimação não-linear[1] do seguinte modelo:

$$\begin{aligned} \ln(y_{it}) - \hat{\beta}_l \ln(l_{it}) &= c + \beta_a \ln(a_{it}) + \beta_k \ln(k_{it}) + \sum_{j=0}^{3-m} \sum_{m=0}^{3} \beta_{mj} \hat{h}_{t-1}^m \hat{P}_{t-1}^j + e_{it} \\ \hat{h}_{t-1} &= \hat{\phi}_{t-1} + \beta_a \ln(a_{it-1}) + \beta_k \ln(k_{it-1}) \end{aligned}$$

Finalmente, os autores calculam a produtividade da seguinte forma:

$$p_t = \exp[\ln(y_{it}) - \beta_l \ln(l_{it}) - \beta_k \ln(k_{it}) - \beta_a \ln(a_{it})]$$

Passemos, agora, ao segundo dos métodos de estimação.

[1] Usando o comando `nl` do STATA.

7.3 Levinsohn e Petrin (2000)

Os autores deste artigo levantam um ponto importante sobre os resultados de Olley e Pakes (1996): nem sempre o investimento responde integralmente aos choques de produtividade das empresas. Uma vez que o investimento é uma variável de controle sobre uma variável de estado (o estoque de capital), em geral ela é custosa de ajustar. Estes custos de ajustamento podem ser de tal ordem que tornam a inversão proposta por Olley e Pakes (1996) para a obtenção da produtividade como função dos investimentos e do capital e da idade inviável.

Em primeiro lugar, a produtividade pode possuir um componente previsível e um componente não previsível. Se parte do componente previsível, o ajustamento nas variáveis de estado (o capital) se dará específicamente sobre ela. Neste caso, o investimento somente responderia à parte não esperada do choque de produtividade. Além disso, como o trabalho é suposto que se ajuste instantaneamente, provavelmente ele se ajusta aos dois componentes, criando assim simultaneidade.

Outro cenário em que isso pode ocorrer é quando a produtividade possui um componente i.i.d. Neste caso, as expectativas sobre o futuro não são ajustadas, ainda que afetam os valores dos fatores variáveis. Além disso, como os choques de produtividades são i.i.d., os investimentos não se alteram em resposta a estes componentes da produtividade.

Para resolver este problema, Levinsohn e Petrin utilizam o valor de alguns insumos intermediários como *proxies* para resolver este problema. Em especial, considerando a seguinte função de produção:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it}^s + \beta l_{it}^u + \beta_m m_{it} + \beta_f f_{it} + \beta_e e_{it} + \omega_{it} + \eta_{it}$$

Em que as variáveis são logs de:

- y_{it} - Valor da Produção
- k_{it} - Valor do estoque de capital da planta da empresa
- l_{it}^s - Valor da mão-de-obra qualificada
- l_{it}^u - Valor da mão-de obra não qualificada
- m_{it} - Valor das matérias-primas
- f_{it} - Valor dos combustíveis
- e_{it} - Valor da eletricidade consumida

Agora, Levinsohn e Petrin fazem a hipótese que a demanda de insumos intermediários – no caso, energia elétrica – é uma função da produtividade e do estoque de capital (as duas variáveis de estado do nosso problema dinâmico):

$$e_t = e(\omega_t, k_t)$$

Como eles mostram, a função e , para qualquer valor de k_t, é crescente em ω_t, o que indica que podemos inverter a função e:

$$\omega_t = e^{-1}(e_t, k_t)$$

Fazendo esta inversão, a função de produção fica assim:

$$y_{it} = \beta_k k_{it} + \beta_l l^s_{it} + \beta l^u_{it} + \beta_m m_{it} + \beta_f f_{it} + \phi(e_{it}, k_{it}) + \eta_{it}$$

Em que:

$$\phi_{it}(e_{it}, k_{it}) = \beta_0 + \beta_k k_{it} + \omega_t(e_t, k_t)$$

Ao invés de usar aproximações polinomiais para limpar os efeitos deste negócio, Levinsohn e Petrin utilizarão um método não-paramétrico diferente para a estimação (*locally weighted least squares*). O que eles fazem é calcular o valor esperado de cada uma das variáveis, em função dos termos e_{it} e k_{it}, e aí fazer a diferença:

$$\begin{aligned} y_{it} - E(y_{it}|e_{it}, k_{it}) &= \beta_s(l^s_{it} - E(l^s_{it}|e_{it}, k_{it})) + \beta_u(l^u_{it} - E(l^u_{it}|e_{it}, k_{it})) + \\ &+ \beta_m(m_{it} - E(m_{it}|e_{it}, k_{it})) + \beta_f(f_{it} - E(f_{it}|e_{it}, k_{it})) + \eta_{it} \end{aligned}$$

Neste caso, conseguimos os coeficientes de todas as variáveis que se ajustam instantaneamente. Supondo o que os autores chamam de separabilidade, não precisamos estimar os coeficientes para o insumo intermediário cujo consumo serve de instrumento. Este coeficiente seria exatamente igual à elasticidade do produto com relação a ele. No caso de uma função Cobb-Douglas, isto implica que o coeficiente deste insumo é igual á participação do mesmo na receita:

$$\beta_e = s_e$$

Neste caso, o processo se facilita muito. Precisamos apenas obter um coeficiente para a variável capital. Isto será obtido por GMM. Dados os coeficientes estimados antes, podemos reescrever, para um valor qualquer do parâmetro β_k:

$$y_{it} - \beta_k k_{it} - \beta_l l^s_{it} - \beta l^u_{it} - \beta_m m_{it} - \beta_f f_{it} - s_e e_{it} - \beta^*_k k_{it}$$

Sabemos que este negócio é igual aos choques de produtividade mais os componentes aleatórios $\omega_{it} + \eta_{it}$. Pela suposição de que a produtividade segue um processo de Markov, temos que $\omega_{it} = E(\omega_{it}|\omega_{it-1}) + \xi_{it}$. Desta forma, temos que:

$$y_{it} - \beta_k k_{it} - \beta_l l^s_{it} - \beta l^u_{it} - \beta_m m_{it} - \beta_f f_{it} - s_e e_{it} - \beta^*_k k_{it} - E(\omega_{it}|\omega_{it-1}) = \xi_{it} + \eta_{it}$$

Tudo estaria tranquilo, se soubéssemos o valor de $E(\omega_{it}|\omega_{it-1})$; no entanto, ainda não o sabemos, de forma que temos que estimar esta parada. Considerando que o termo η_{it} é composto pela parte aleatória da função de produção, podemos afirmar que:

$$E(\omega_{it} + \eta_{it}|\omega_{it-1}) = E(\omega_{it}|\omega_{it-1})$$

Assim, podemos usar como estimativa de $\omega_{it} + \eta_{it}$ o seguinte:

$$y_{it} - \beta_k k_{it} - \beta_l l^s_{it} - \beta l^u_{it} - \beta_m m_{it} - \beta_f f_{it} - s_e e_{it} - \beta^*_k k_{it}$$

Agora, para a estimativa de ω_{it-1}, podemos fazer o seguinte:

$$\hat{\omega}_{it-1} = \hat{\phi}_{it-1} - s_e e_{it-1} - \beta^*_k k_{it-1}$$

Com estes valores, podemos utilizar mínimos quadrados ponderados locais para calcular $E(\omega_{it}|\omega_{it-1})$, que denominaremos $\hat{\Omega}$.Usando esta estimativa, podemos calcular os valores de $\xi_{it} + \eta_{it}$, para os valores candidatos de β_k. Finalmente, podemos calcular a função critério:

$$q = \left(\sum_i \sum_{t=t_0+1}^{T} (\xi_{it} + \eta_{it}) k_{it} \right)^2$$

Com isso, estimamos o coeficiente do capital.

Finalmente, um ponto importante que devemos levar em consideração é o cálculo de economias de Escala. Panzar (1989, p. 8) dá uma medida de *economias de escala tecnológicas*, definidas da seguinte forma:

$$S_t = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} x_i \frac{\partial f_j}{\partial x_i}}{\sum_j y_j}$$

Em que $\frac{\partial f_j}{\partial x_i}$ é a produtividade marginal do insumo i na produção do bem j, x_i a quantidade do fator i de produção e y_j a quantidade do produto j. Os retornos de escala tecnológicos seriam crescentes caso S_t fosse maior do que 1, decrescentes caso fosse menor do que 1, e constantes caso fossem iguais a 1.

Parte III

Conduta

Capítulo 8

Estimação de Parâmetros de Conduta

8.1 Estimação de Parâmetros de Conduta

O primeiro passo na estimação direta de conduta em modelos econômicos foi dado nos artigos de Breshanan (1982)[13] e Lau (1982)[33]. A grande pergunta que estes dois autores se fazem é se, a partir de observações sobre quantidades agregadas e preços em um dado mercado, é possível discernir entre competição perfeita e cartel? Note-se que, neste ponto, os autores não estão preocupados com a interação estratégica entre as empresas, mas sim determinar se é possível determinar, a partir dos dados observados, se os casos polares de competição e monopólio são discerníveis.

Basicamente, o autor pressupõe que as quantidades e preços observados são o resultado do equilíbrio entre uma curva de demanda e uma relação de oferta. O sunblinhado no termo relação de oferta ocorre porque apenas no caso de concorrência perfeita temos uma curva de oferta como a que entendemos tradicionalmente – ou seja, o ramo ascendente da curva de custo marginal, acima da curva de custo médio – o que implica em uma relação de oferta $P = CMg$. A forma mais geral desta relação de oferta é $RMg_p = CMg$, em que RMg_p representa a receita marginal percebida pela empresa, sendo que o percebida diz respeito ao fato que a receita marginal, em mercados oligopolizados, é o resultado não apenas das ações da empresa, mas também do que as empresas esperam que sejam as ações dos seus competidores. A diferença entre as duas formas, $P = CMg$ e $CMg = RMg_p$ daria, segundo Breshanan, uma aproximação dos graus de poder de mercado das empresas, com um extremo em que $RMg_p = RMg$ e temos o preço de monopólio, ao extremo de concorrência perfeita em que $RMg_P = P$.

Para ilustrarmos melhor o problema essencial de estimação de parâmetros de conduta, vamos assumir uma função demanda e uma função custos ambas lineares. Neste caso, a função demanda inicialmente seria igual a:

$$\begin{aligned} Q &= \alpha_0 + \alpha_1 P + \alpha_2 Y + \epsilon \\ P &= \frac{-\alpha_0}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Y + \frac{Q}{\alpha_1} \end{aligned}$$

A função custo marginal seria, portanto, igual a:

$$CMg = \beta_0 + \beta_1 Q + \beta_2 W$$

A receita marginal percebida pela empresa i seria então igual a:

$$\begin{aligned} RMg &= \frac{\partial(P \times Q)}{\partial Q} \\ &= P - Q \times \frac{\partial P}{\partial Q}\frac{\partial Q}{\partial q_i} = P - Q \times \frac{\partial P}{\partial Q} \times \lambda \\ &= P - \lambda \frac{Q}{\alpha_1} \end{aligned}$$

Em que o λ é igual a alguma medida de o quanto a empresa espera que as outras respondam à sua mudança na quantidade, o que dá o efeito sobre a quantidade agregada de mercado decorrente de um aumento inicial na quantidade produzida por uma empresa. Por exemplo, no caso em que tenhamos um monopólio, isto seria igual a um, e na competição perfeita teríamos isso sendo igual a zero. Igualando a receita marginal ao custo marginal, temos que:

$$\begin{aligned} CMg &= RMg \\ \beta_0 + \beta_1 Q + \beta_2 W &= P - \lambda \frac{Q}{\alpha_1} \\ P &= \lambda \frac{Q}{\alpha_1} + \beta_0 + \beta_1 Q + \beta_2 W \end{aligned}$$

Podemos notar que, para o caso em questão, não conseguimos identificar separadamente ois parâmetros: λ e α_1. Em outras palavras, em resposta aos choques do modelo, por exemplo, na variável Y, não conseguiremos distinguir se os efeitos sobre os preços de mercado são decorrentes da inclinação da curva de demanda ou do markup das empresas. A menos que conhecessemos diretamente os custos marginais - ou tivéssemos informações adicionais sobre a forma da função demanda – não conseguiríamos estimar separadamente estes coeficientes.

A solução proposta por Breshanan (1982)[13] e Lau (1982)[33] envolve modelar a inclinação da curva de demanda diferentemente. Em outras palavras, deveríamos modelar a sensibilidade do preço à quantidade diferentemente, permitindo que a mesma mudasse de acordo com alterações em algumas variáveis exógenas. Em especial, ele assume que a função demanda possui a seguinte forma:

$$Q = \alpha_0 + \alpha_1 P + \alpha_2 Y + \alpha_3 PY + \epsilon$$

Reescrevendo a função demanda como P em função de Q:

$$P = \frac{-\alpha_0}{\alpha_1 + \alpha_3 Y} - \frac{\alpha_2}{\alpha_1 + \alpha_3 Y} Y + \frac{Q}{\alpha_1 + \alpha_3 Y}$$

Montando a relação de oferta, temos que:

$$P = \lambda \frac{Q}{\alpha_1 + \alpha_3 Y} + \beta_0 + \beta_1 Q + \beta_2 W$$

Neste caso, conseguimos identificar adequadamente os coeficientes envolvidos, pois o pacote de três coeficientes, α_1, α_3 e λ pode ser obtido por duas variáveis diferentes. Como afirma Breshanan (1982, p. 92)[13], as rotações da curva de demanda em torno do ponto de equilíbrio sempre revelam o poder de mercado. A seguir, vamos discutir algumas das limitações da estimação de parâmetros de conduta.

Considerando o que vimos anteriormente, um caminho interessante seria a estimação de uma equação como a que vimos anteriormente, e obtendo estimativas para os coeficientes relevantes. No entanto, temos algumas considerações de ordem prática. A primeira delas diz respeito à forma pela qual deveríamos modelar estes parâmetros de conduta em setores que, além de oligopolizados, são também caracterizados por produtos diferenciados. A crítica à aplicação deste modelo de Breshanan neste caso será abordada mais adiante.

A segunda crítica diz respeito ao fato que, mesmo que consideremos os produtos dos diferentes ofertantes como essencialmente homogêneos, temos que a estimação do parâmetro de conduta, sendo realizada com dados de diferentes empresas, impõe uma restrição muito grande, que é a igualdade dos parâmetros de conduta – também conhecidos como variações conjecturais – entre as diferentes empresas. Para isto, vamos supor que tenhamos informações sobre cada uma das empresas, além de preços de mercado. Aqui podemos investigar se o modelo de Cournot também pode ser verificado com os dados. Neste caso, supondo j empresas, temos a seguinte função demanda (assumida linear):

$$\begin{aligned} P_t &= \alpha - \beta_1 (\sum_j q_{jt}) + \beta_2 w_t + \varepsilon_t^D \\ Q_t &= \sum_j q_{jt} \end{aligned}$$

Além disso, temos a seguinte função de receita marginal, derivada do problema de maximização da empresa com relação à quantidade:

$$\begin{aligned}
RT_i &= Pq_i \\
RMg_i &= P + \frac{\partial P}{\partial Q}\frac{\partial Q}{\partial q_i} \\
&= P + P\frac{\partial P}{\partial Q}\frac{Q}{P}\frac{q_i}{Q}\frac{\partial Q}{\partial q_i} \\
&= P\left(1 - \frac{q_i}{Q\left|\epsilon_D\right|}\frac{\partial Q}{\partial q_i}\right) \\
&= P\left(1 - \frac{q_i}{Q\left|\epsilon_D\right|}\left(1 + \sum_{j\neq i}\frac{\partial q_j}{\partial q_i}\right)\right)
\end{aligned}$$

Por outro lado, supondo uma função custo marginal linear nos parâmetros:

$$CMg_i = \beta_0 + \beta_1 q_i + \beta_2 z_i$$

Podemos chegar em uma relação de oferta:

$$P = \frac{\beta_0 + \beta_1 q_i + \beta_2 z_i}{\left(1 - \frac{q_i}{Q|\epsilon_D|}\left(1 + \sum_{j\neq i}\frac{\partial q_j}{\partial q_i}\right)\right)}$$

Evidentemente, isso tem que ser feito para cada uma das empresas, gerando um sistema de $J+1$ equações. Applebaum (1982)[2] desenvolve raciocínio nessas linhas para estender a metodologia de Bresnahan e Lau para o caso com muitas empresas. Se você estimar cada um dos coeficientes deste sistema de equações na suposição que os coeficientes são diferentes entre as empresas, a menos que exista um número descomunal de equações, ao estimar os termos $\frac{\partial q_j}{\partial q_i}$ individualmente não há graus de liberdade suficientes

O problema é que, para alguns casos, é razoável supor que estas variações conjecturais sejam diferentes entre as empresas. Uma solução parcial para este problema foi levantada por Gollop e Roberts (1979)[27]. Em primeiro lugar, os autores derivam uma relação de oferta a partir dos insumos, com a seguinte função translog de produção:

$$\ln q_i = \alpha_0 + \sum_k \alpha_k \ln X_k + \sum_m \sum_k \gamma_{mk} \ln X_m \ln X_k$$

A função de lucro desta empresa é dada por:

$$\pi_i = P(Q)q_i - \sum_k w_k X_k$$

As derivadas desta função com relação aa cada um destes fatores de produção são dadas por:

$$\begin{aligned}\frac{\partial \pi_i}{\partial X_k} &= \frac{\partial P}{\partial Q}\frac{\partial Q}{\partial q_i}\frac{\partial q_i}{\partial X_k}q_i + P\frac{\partial q_i}{\partial X_k} - w_k = 0\\ &= P\frac{\partial q_j}{\partial X_k}\left(1+\left(\sum_j Q\frac{\partial P}{\partial Q}\frac{P}{P}\frac{\partial q_j}{\partial q_i}\frac{q_i}{Q}\right)\right) - w_k\end{aligned}$$

Lembrando que podemos reescrever a elasticidade da demanda pelos produtos de uma empresa como uma função do market share e da elasticidade agregada de mercado[1]:

$$\frac{1}{\epsilon_j^D} = -\frac{q_j}{P}\frac{\partial P}{\partial q_j} = -\frac{q_j}{Q}\frac{Q}{P}\frac{\partial P}{\partial Q}\frac{\partial Q}{\partial q_j} = \frac{q_j}{Q\epsilon^D}$$

Com isto, podemos reescrever a fórmula em termos da elasticidade da demanda de mercado:

$$\frac{\partial \pi_i}{\partial X_k} = P\frac{\partial q_j}{\partial X_k}\left(1+\left(\frac{q_i}{Q\epsilon^D}\sum_j \frac{\partial q_j}{\partial q_i}\right)\right) - w_k$$

Dada a função de produção dos autores, temos que:

$$\begin{aligned}\frac{\partial q_j}{\partial X_k} &= \frac{\partial \ln q_j}{\partial \ln X_k}\frac{X_k}{q_j} = \frac{X_k}{q_j}\left(\alpha_k + \sum_k \gamma_{mk}\ln X_k\right)\\ &= M_{jk}\frac{X_k}{q_j}\end{aligned}$$

O que, jogando na condição de primeira ordem, nos dá o seguinte:

$$\frac{w_k X_k}{P q_i} = M_{ik}\left(1+\left(\frac{q_i}{Q\epsilon^D}\sum_j \frac{\partial q_j}{\partial q_i}\right)\right)$$

Do lado esquerdo da igualdade, temos o peso do fator k na receita total da empresa i. O problema está do lado direito da igualdade. Obtendo independentemente uma estimativa da elasticidade e do termo M_{ik}, temos que estimar os $\frac{\partial q_j}{\partial q_i}$. Neste caso, temos exatamente J coeficientes a serem estimados por equação, o que é pouco razoável. Além disso, e o que é pior, como estes coeficientes estão todos multiplicando a mesma coisa, não teríamos como identificá-los separadamente. Note-se que isso não necessariamente seria um problema. Se estivéssemos interessados em estimar o grau de dependência oligopolística, estimar a soma das variações conjecturais. No entanto, se

[1] Na suposição que as outras empresas não se mexam em resposta à sua ação

estivermos interessados em modelar a sensibilidade da resposta de algumas empresas às alterações na produção de uma empresa em especial, é necessário ir além.

Para isto, os autores começam classificando as empresas por classes de tamanho, separando as mesmas em diferentes categorias. Neste sentido, serão identificadas apenas as respostas de cada classe de tamanho a alterações na quantidade produzida de uma empresa. Neste caso, a somatória na fórmula acima se transforma:

$$\frac{w_k X_k}{P q_i} = M_{ik}\left(1 + \left(\frac{q_i}{Q\epsilon^D}\sum_{s_i \in S}\left(\sum_{j \in s_i}\frac{\partial q_j}{\partial q_i}\right)\right)\right)$$

Supondo que estejamos interessados na resposta relativa de cada um destes grupos, podemos reescrever o parênteses interno da seguinte forma:

$$\sum_{j \in s_i}\frac{\partial q_j}{\partial q_i} = \left(\sum_{j \in s_i} q_j\right)\frac{\partial \ln(\sum_{j \in s_i} q_j)}{\partial q_i} = \left(\sum_{j \in s_i} q_j\right)\beta_{s_i}$$

A última passagem decorre do fato que a resposta do grupo é um parâmetro a ser estimado. Neste caso, temos o seguinte:

$$\frac{w_k X_k}{P q_i} = M_{ik}\left(1 + \left(\frac{q_i}{Q\epsilon^D}\sum_{s_i \in S}\left(\sum_{j \in s_i} q_j\right)\beta_{s_i}\right)\right)$$

Neste caso, já teríamos apenas S grupos a serem estimados, o que facilitaria bem mais a estimação, com um problema. Aqui estamos supondo que cada uma das empresas tenha uma conjectura idêntica, o que dá problemas decorrentes do fato que fizemos uma separação arbitrária dos grupos. Ou seja, empresas de tamanhos parecidos, mas alocadas em diferentes grupos, acabariam com conjecturas diferentes.

Gollop e Roberts (1979)[27] vêm com uma solução interessante para este dilema, reescrevendo cada um dos β_{s_i} como função das conjecturas das empresas limites de cada um dos grupos:

$$\begin{aligned}
\beta_{s_i} &= \beta_{s_i}^{-} w_i + \beta_{s_i}^{+}(1 - w_i) \\
\frac{w_k X_k}{P q_i} &= M_{ik}\left(1 + \left(\frac{q_i}{Q\epsilon^D}\sum_{s_i \in S}\left(\sum_{j \in s_i} q_j\right)\left(\beta_{s_i}^{-} w_i + \beta_{s_i}^{+}(1 - w_i)\right)\right)\right)
\end{aligned}$$

Por exemplo, esta equação para dois grupos fica da seguinte forma:

$$\frac{w_k X_k}{P q_i} = M_{ik}\left(1 + \frac{q_i}{Q\epsilon^D}\sum_{j \in G_1} q_j\left(\beta_1^{-} w_i + \beta_1^{+}(1 - w_i)\right) + \frac{q_i}{Q\epsilon^D}\sum_{j \in G_2} q_j\left(\beta_2^{-} w_i + \beta_2^{+}(1 - w_i)\right)\right)$$

Esta relação vale para cada um dos K insumos, o que implica em um sistema de K equações – somadas a uma equação de demanda e/ou uma função de produção.

8.2 Críticas

A crítica principal do artigo de Corts (1998)[20] diz respeito à interpretação do parâmetro de conduta. No mundo real, o "grau de competitividade" poderia ser entendido *como se* as empresas no mercado jogam um jogo de variações conjecturais, sendo que a conjectura estimada se aplica às expectativas de resposta dos seus competidores. Isto é incorreto se as respostas, na margem, forem diferentes das respostas médias.

Para obtermos estimativas deste parâmetro de conduta, existem dois estágios, que inclusive ilustrarão interpretações interessantes do parâmetro de conduta. O primeiro deles envolve a estimação do parâmetro propriamente dito que pode ser obtido de uma das três formas:

- O custo marginal é constante e temos um deslocador paralelo da demanda
- O custo marginal não é constante, mas temos um instrumento que desloca a inclinação da curva de demanda
- Se fizermos hipóteses externas sobre a forma da curva de demanda e sobre o mecanismo de formação de preços.

O segundo passo é mapear este valor estimado na margem preço-custo ajustada à elasticidade[2]. Teoricamente, este mapeamento é da seguinte forma:

$$P_i = CMg_i - \lambda_i \frac{\partial P_i}{\partial Q} q_i$$

Reorganizando a equação acima, e supondo firmas idênticas, temos:

$$\begin{aligned}\tilde{\lambda}_i &= \frac{P_i - CMg_i}{-P' q_i} \\ &= \frac{P - CMg_i}{P} N \varepsilon_D\end{aligned}$$

Esta reorganização implica que podemos entender este "parâmetro de conduta" como sendo um Índice de Lerner ajustado à elasticidade; ou uma medida da margem preço-custo que é normalizada tanto pelo preço quanto pela elasticidade de demanda – para distinguir os mercados em que a margem é alta porque a demanda é menos elástica dos mercados em que a margem é alta por aspectos colusivos. Ou seja, uma média da margem preço custo de cada empresa.

[2]Note que, abaixo, o que está multiplicando a inclinação da curva de demanda é a quantidade da produção da empresa i, e não a quantidade total. Neste caso, a interpretação do parâmetro de conduta é similar, mas a definição é diferente.

Agora vamos ver o que nós estamos estimando diretamente com a ajuda da técnica de parâmetros de conduta ou variações conjecturais. Vamos supor o seguinte sistema:

$$
\begin{aligned}
P &= \alpha_0 + \alpha_1 Y + \alpha_2 Q + \epsilon_t \\
P &= \beta_0 + \beta_1 w + \beta_2 q_i + \xi_t
\end{aligned}
$$

A última equação é a relação de oferta e corresponde à uma função custo marginal da seguinte forma: $CMg_i = c_0 + c_1 w$. Podemos mostrar graficamente que isto pressupõe uma curva de custo marginal horizontal, que permite a identificação do parâmetro de conduta mesmo quando estamos sem um termo que rotaciona a curva de demanda. Vamos supor que estimamos o sistema acima por MQ2E, o que dá a seguinte forma funcional para o parâmetro β_2:

$$
\begin{aligned}
\hat{\beta}_2 &= (\hat{q}_i{}^T \mathbf{M_w} \hat{q}_i)^{-1} (\hat{q}_i{}^T \mathbf{M_w} P) \\
&= (\hat{q}_i{}^T \mathbf{M_w} \hat{q}_i)^{-1} (\hat{q}_i{}^T \mathbf{M_w} a_1 Y + \hat{q}_i{}^T \mathbf{M_w} a_2 Q + \hat{q}_i{}^T \mathbf{M_w} e)
\end{aligned}
$$

Uma vez que estamos em um equilíbrio simétrico, temos que $Q = Nq_i$. Substituindo isso na equação acima, podemos reescrever a fórmula da seguinte forma:

$$\hat{\beta}_2 = (\hat{q}_i{}^T \mathbf{M_w} \hat{q}_i)^{-1} (\hat{q}_i{}^T \mathbf{M_w} a_1 Y + \hat{q}_i{}^T \mathbf{M_w} e) + Na_2$$

Em que $\mathbf{M_w}$ representa a matriz projetora no espaço ortogonal ao vetor w, ou $\mathbf{M_w} = (\mathbf{I} - w^T (w^T w)^{-1} w)$Para investigarmos a consistência deste estimador precisamos tirar o limite desta expressão para o tamanho da amostra tendendo ao infinito. Logo, o *plim* será:

$$\hat{\beta}_2 = (\hat{q}_i{}^T \mathbf{M_w} \hat{q}_i)^{-1} (\hat{q}_i{}^T \mathbf{M_w} a_1 Y) + Na_2$$

Podemos então dizer que, assintoticamente, este parâmetro $\hat{\beta}_2$ é igual a:

$$\hat{\beta}_2 = \frac{a_1}{\gamma} + Na_2$$

Em que $\gamma = (Y_i^T \mathbf{M_w} Y_i)^{-1} (Y_i \mathbf{M_w} q_i)$ é a projeção linear de q_i em Y . A grande moral da história aqui é que isto implica que a estimativa do parâmetro de conduta é:

$$\hat{\lambda} = \frac{\hat{\beta}_2}{-\hat{\alpha}_2} = -\frac{a_1}{a_2 \gamma} - N$$

O que isto significa? Significa que o parâmetro de conduta estimado $\hat{\lambda}$ é uma função de apenas os parâmetros de demanda e do coeficiente γ que é a sensibilidade da quantidade de equilíbrio q_i ao deslocador de demanda Y_i. Dadas estas premissas, o parâmetro de conduta estimado é completamente determinado pelo grau que as quantidades de equilíbrio respondem ao lado da demanda. Podemos escrever $\hat{\beta} = \frac{d(P-CMg)/dY}{dq/dY}$. E como isso se relaciona com os aspectos de oferta deste modelo?

Vamos supor que a quantidade ótima de uma empresa seja determinada por uma regra que é linear em termos dos deslocadores de demanda – um equilíbrio como o do Modelo de Cournot poderia jutificar uma regra como essa. Neste caso, $\gamma = \frac{dq^*}{dY}$ e o parâmetro de conduta que vimos pode ser entendido como a 'inclinação' da margem preço-custo com relação à flutuações na quantidade de equilíbrio devidas à alterações nos deslocadores de demanda. Mais formalmente, é igual a:

$$\begin{aligned}\hat{\lambda} &= \frac{1}{-P'}\frac{dP}{dq^*} = \frac{1}{-\alpha_2}\beta_2 \\ &= -\frac{d(P-CMg_i)}{P'dY}\frac{dY}{dq^*}\end{aligned}$$

Ou seja, o parâmetro de conduta mede a resposta Marginal da quantidade de equilíbrio do preço com respeito à quantidade, sendo que estas alterações são induzidas pela demanda (por causa da premissa de constância dos custos marginais). Note que este negócio é diferente da conduta média; se os dois não baterem, as interpretações têm de ser ajustadas.

Capítulo 9

Outras Abordagens

Nesta aula, iremos fechar a discussão da modelagem de conduta apresentando duas linhas alternativas às que apresentamos anteriormente. A primeira delas, baseada no artigo de Panzar e Rosse (1987)[37] tem uma vantagem no fato que, para a sua implementação não é necessário possuir dados sobre a quantidade de produto ou de insumos, apenas dados – específicos à empresa – sobre os preços dos insumos enfrentados pela empresa, bem como dados sobre os preços dos outros insumos utilizados pela empresa.

A segunda das metodologias, empregada por Baker e Breshanan (1985, 1988)[4, 5] busca determinar a existência de poder de mercado apenas com a estimação da curva de demanda residual da empresa. Neste caso, a necessidade para o cálculo de elasticidades cruzadas é eliminada. Passemos à análise dos diferentes modelos.

9.1 Estatística Panzar-Rosse

Vamos começar pela derivação da estatística de Panzar e Rosse para o caso de uma empresa monopolística. Comecemos considerando a seguinte função receita:

$$R(q) = p(q) \times q$$

Além disso, vamos considerar a seguinte função custos:

$$C(q) = C(q, \mathbf{w})$$

Em que q era a quantidade produzida, e $\mathbf{w}$ representa o vetor de preço dos fatores. Além disso, podemos considerar a função demanda como dependendo de deslocadores exógenos, como a renda, ou alíquotas específicas de impostos. Como estas variáveis não são relevantes para a análise que se segue, não vamos representá-las para fins de economia notacional. Vamos formalizar mais esta função custo como a solução para o seguinte problema de otimização condicionada;

$$\begin{aligned} C(q, \mathbf{w}) &= \min_{\mathbf{w},q} \mathbf{w}L \\ suj.a& \\ f(L) &\geq q \end{aligned}$$

Dada esta função custo, temos que a função lucro pode ser construída da seguinte forma:

$$\pi(q, \mathbf{w}) = R(q) - C(q, \mathbf{w})$$

Considere

$$q(h) = \arg\max_{q} \pi[q, (1+h)\mathbf{w}]$$

Para algum $h \geq 0$. Isto significa que esta é a quantidade que mazimiza o lucro dado uma pequena oscilação nos preços dos insumos. Vamos definir também $R(h) = R(q(h))$. Portanto:

$$R(h) - C(q(h), (1+h)\mathbf{w}) \geq R(0) - C(q(0), (1+h)\mathbf{w}) \tag{9.1}$$

Supondo uma função custo homogênea de grau 1 nos preços dos insumos, esta regra pode ser escrita como:

$$R(h) - (1+h)C(q(h), \mathbf{w}) \geq R(0) - (1+h)C(q(0), \mathbf{w}) \tag{9.2}$$

Da mesma forma que a equação (9.1), podemos a partir do comportamento da empresa, definir:

$$R(0) - C(q(0), \mathbf{w}) \geq R(h) - C(q(h), \mathbf{w}) \tag{9.3}$$

Esta equação pode ser reescrita como:

$$(1+h)R(0) - (1+h)C(q(0), \mathbf{w}) \geq (1+h)R(h) - (1+h)C(q(h), \mathbf{w}) \tag{9.4}$$

Somando as equações (9.1) e (9.4), podemos escrever:

$$\begin{aligned} R(h) - (1+h)C(q(h), \mathbf{w}) + (1+h)R(0) - (1+h)C(q(0), \mathbf{w}) &\geq \\ R(0) - (1+h)C(q(0), \mathbf{w}) + (1+h)R(h) - (1+h)C(q(h), \mathbf{w}) & \\ R(h) + (1+h)R(0) &\geq R(0) + (1+h)R(h) \end{aligned}$$

Podemos reescrever esta última desigualdade da seguinte forma:

$$\begin{aligned} -h(R(h) - R(0)) &\geq 0 \\ \frac{R(h) - R(0)}{h} &\leq 0 \end{aligned}$$

Podemos reescrever esta razão da seguinte forma, uma vez que $q(h)$ depende dos preços dos fatores $(1+h)\mathbf{w}$:

$$\frac{R((1+h)\mathbf{w}) - R(\mathbf{w})}{h} \leq 0$$

O limite desta razão com $h \to 0$, chegamos à estatística ψ, a estatística Panzar-Rosse.

A derivada da função $R(h)$ é dada por:

$$\begin{aligned}
\frac{\partial R}{\partial h} &= \mathbf{w} \bigtriangledown_{\mathbf{w}} R(h) \leq 0 \\
\left.\frac{\partial R}{\partial h}\right|_{h\to 0} &= \mathbf{w} \bigtriangledown_{\mathbf{w}} R(q) \leq 0 \\
\psi = \frac{1}{R(q)} \left.\frac{\partial R}{\partial h}\right|_{h\to 0} &= \frac{1}{R(q)} \times \mathbf{w} \bigtriangledown_{\mathbf{w}} R(q) \leq 0
\end{aligned}$$

Como a quantidade demandada é igual à quantidade produzida do monopolista, estamos em um monopólio, e temos a conclusão do Teorema 1 de Panzar e Rosse (1987); a soma das elasticidades dos preços dos fatores com relação à receita total não pode ser positiva[1]. Este resultado intuitivamente quer dizer que, em resposta a uma elevação nos custos em $h\%$, as receitas totais se reduzem. Alternativamente, definindo $k = 1 + h$ e $\tilde{R}(k) = R(k-1)$, podemos ter:

$$\psi = \frac{1}{R(q)} \left.\frac{\partial R}{\partial h}\right|_{h\to 0} = \frac{1}{\tilde{R}(1)} \left.\frac{\partial \tilde{R}}{\partial k}\right|_{k\to 0}$$

Ou seja, também é a elasticidade da receita com um aumento dado nos preços dos fatores, mantidos os preços relativos dos mesmos constantes. Se obtivermos um $\hat{\psi} > 0$, temos indicação que a empresa não se comporta como um monopólio. Agora vamos supor que a empresa esteja em um mercado como em concorrência monopolística.

Vamos supor que existam n empresas idênticas, e que cada uma das empresas enfrenta uma função demanda inversa percebida da forma $p(q,n)$, sendo que $\frac{\partial p}{\partial q} < 0$, $\frac{\partial p}{\partial n} < 0$ e $\frac{\partial}{\partial n}\left(-\frac{p}{q\frac{\partial p}{\partial n}}\right) \geq 0$. Esta última hipótese é que a elasticidade-preço da demanda enfrentada por uma única empresa não cai quando o número de empresas aumenta. A segunda condição implica que a sensibilidade da receita com relação ao número de competidores é negativa (ou seja, mais competidores, menor receita). Finalmente, vamos assumir que o mercado está em equilíbrio de longo prazo; ou seja, existe entrada e saída de empresas até que o lucro por empresa seja zero.

[1] Na verdade, o que Panzar e Rosse afirmam acerca deste resultado é que ele caracteriza um equilíbrio em que o um jogador toma suas decisões independentemente de outras empresas.

Dadas estas condições, temos que as equações que definem o equilíbrio na indústria são:

$$\begin{aligned}
\frac{\partial R(q,n)}{\partial q} - \frac{\partial C(q,\mathbf{w})}{\partial q} &\equiv 0 \\
R(q,n) - C(q,\mathbf{w}) &\equiv 0
\end{aligned}$$

Diferenciando as duas equações com respeito a w:

$$\begin{aligned}
\frac{\partial^2 R(q,n)}{\partial q^2}\frac{\partial q}{\partial w} + \frac{\partial^2 R(q,n)}{\partial q \partial n}\frac{\partial n}{\partial w} - \left[\frac{\partial^2 C(q,\mathbf{w})}{\partial q^2}\frac{\partial q}{\partial w} + \frac{\partial^2 C(q,\mathbf{w})}{\partial q \partial w}\right] &\equiv 0 \\
\frac{\partial R(q,n)}{\partial q}\frac{\partial q}{\partial w} + \frac{\partial R(q,n)}{\partial n}\frac{\partial n}{\partial w} - \left[\frac{\partial C(q,\mathbf{w})}{\partial q}\frac{\partial q}{\partial w} + \frac{\partial C(q,\mathbf{w})}{\partial w}\right] &\equiv 0
\end{aligned}$$

Reorganizando as equações:

$$\begin{aligned}
\left[\frac{\partial^2 R(q,n)}{\partial q^2} - \frac{\partial^2 C(q,\mathbf{w})}{\partial q^2}\right]\frac{\partial q}{\partial w} + \frac{\partial^2 R(q,n)}{\partial q \partial n}\frac{\partial n}{\partial w} &= \frac{\partial^2 C(q,\mathbf{w})}{\partial q \partial w} \\
\left[\frac{\partial R(q,n)}{\partial q} - \frac{\partial C(q,\mathbf{w})}{\partial q}\right]\frac{\partial q}{\partial w} + \frac{\partial R(q,n)}{\partial n}\frac{\partial n}{\partial w} &= \frac{\partial C(q,\mathbf{w})}{\partial w}
\end{aligned}$$

Na primeira equação, o coeficiente de $\frac{\partial q}{\partial w}$ é negativo devido às condições de segunda ordem de maximização de lucros (o que está no colchete é a derivada segunda da função lucro). Na segunda equação, pelas condições de primeira ordem do problema de maximização de lucros da empresa, temos que os colchetes são iguais a zero. Desta forma, podemos representar o sistema acima na forma matricial:

$$\begin{pmatrix} \frac{\partial^2 R(q,n)}{\partial q^2} - \frac{\partial^2 C(q,\mathbf{w})}{\partial q^2} & \frac{\partial^2 R(q,n)}{\partial q \partial n} \\ 0 & \frac{\partial R(q,n)}{\partial n} \end{pmatrix} \begin{pmatrix} \frac{\partial q}{\partial w} \\ \frac{\partial n}{\partial w} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 C(q,\mathbf{w})}{\partial q \partial w} \\ \frac{\partial C(q,\mathbf{w})}{\partial w} \end{pmatrix}$$

Vamos definir

$$D = \left[\frac{\partial^2 R(q,n)}{\partial q^2} - \frac{\partial^2 C(q,\mathbf{w})}{\partial q^2}\right]\frac{\partial R(q,n)}{\partial n} > 0$$

O determinante da matriz à esquerda. Desta forma, o sistema fica sendo;

$$D\begin{pmatrix} \frac{\partial q}{\partial w} \\ \frac{\partial n}{\partial w} \end{pmatrix} = \begin{pmatrix} \frac{\partial R(q,n)}{\partial n} & -\frac{\partial^2 R(q,n)}{\partial q \partial n} \\ 0 & \frac{\partial^2 R(q,n)}{\partial q^2} - \frac{\partial^2 C(q,\mathbf{w})}{\partial q^2} \end{pmatrix} \begin{pmatrix} \frac{\partial^2 C(q,\mathbf{w})}{\partial q \partial w} \\ \frac{\partial C(q,\mathbf{w})}{\partial w} \end{pmatrix}$$

O que permite que determinemos a sensibilidade da quantidade produzida a alterações no fator de produção. Cozinhando fortemente este negócio, temos que, para uma empresa em concorrência monopolística, isto aqui implica que $\psi \leq 1$. Da mesma forma, para uma empresa em concorrência perfeita, eles afirmam que $\psi = 1$. Este resultado se justifica pelo fato que, com uma elevação nos custos de todos os fatores de produção em $h\%$, teremos entrada e saída de empresas até que a receita total suba na mesma porcentagem, de forma que o efeito é unitário.

Além disso, como mostrado em Shaffer (1982), podemos mostrar que, para elasticidade-preço da demanda constante e custo linear na quantidade produzida, temos que:

$$\begin{aligned} \psi &= 1 - \frac{1}{m} \\ m &= \frac{p - CMg}{p} = \frac{1}{\varepsilon} \end{aligned}$$

No contexto de modelos específicos de oligopólio, coisas adicionais podem ser feitas. Supondo, mais uma vez, que a função custo marginal é constante na quantidade produzida, no caso de conjecturas equivalentes (ou seja, as empresas seguem um comportamento oligopolizado), temos que $\psi \leq 0$. Passemos então à análise do último dos modelos de conduta que abordaremos neste curso: a análise da elasticidade da demanda residual.

9.2 Demanda Residual

A idéia desta abordagem é propôr procedimentos econométricos para a estimação de um sistema de demanda que será enfrentado pelas partes que serão objeto de fusão, baseada somente em dados pré-fusão. O elemento chave para isto é a agregação dos efeitos de todas as outras firmas em um único parâmetro. Formalmente, começaremos com um sistema de demanda de produtos diferenciados com n empresas. Após manipulação do sistema, eliminamos os preços de todas as empresas, exceto as duas diretamente envolvidas em uma fusão. Isto reduz a dimensionalidade do sistema, para um tamanho que o torna gerenciável. Passemos então à descrição da metodologia propriamente dita.

Vamos supor que duas empresas estejam contemplando a fusão. Podemos representar a função demanda inversa de cada uma das empresas da seguinte forma:

$$\begin{aligned} P_1 &= h_1(q_1, q_2, \tilde{q}, Y, \eta_1) \\ P_2 &= h_2(q_1, q_2, \tilde{q}, Y, \eta_2) \end{aligned}$$

Em que $\tilde{q}$ representa a produção das $n-2$ empresas não diretamente envolvidas, Y um vetor de deslocadores de demanda, η_1 e η_2 são vetores de parâmetros da função demanda das duas empresas. Podemos, adicionalmente, representar a função demanda inversa das $n-2$ empresas remanescentes do mercado da seguinte forma:

$$\tilde{P} = \tilde{h}(q_1, q_2, \tilde{q}, Y, \tilde{\eta})$$

Agora, vamos especificar o comportamento das $n-2$ empresas remanescentes. Independentemente da hipótese comportamental, podemos assumir que elas agirão de forma a igualar a receita marginal percebida (ou seja, envolvendo as conjecturas das empresas sobre o comportamento das outras) igual ao seu custo marginal. Ou seja, para estas $n-2$ empresas a receita marginal é da seguinte forma:

$$\tilde{MR} = \tilde{P} + \tilde{g}(q_1, q_2, \tilde{q}, Y, \tilde{\eta})\tilde{q}$$

Sendo que a função $\tilde{g}$ é uma representação da inclinação da demanda $\tilde{h}$ percebida pelas $n-2$ empresas. Da mesma forma, podemos escrever a função custo marginal da seguinte forma:

$$\tilde{CMg} = \tilde{j}(\tilde{q}, W, \tilde{\beta})$$

Ou seja, a relação de preços é dada pela seguinte igualdade:

$$\tilde{MR} = \tilde{P} + \tilde{g}(q_1, q_2, \tilde{q}, Y, \tilde{\eta})\tilde{q} = \tilde{CMg} = \tilde{j}(\tilde{q}, W, \tilde{\beta})$$

Resolvendo o sistema composto pela equação de demanda e a igualdade entre receita marginal e custo marginal, podemos expressar a quantidade produzida pelas $n-2$ firmas da seguinte forma:

$$\tilde{q} = \tilde{e}(q_1, q_2, Y, W, \tilde{\eta}, \tilde{\beta})$$

Com esta equação, podemos escrever uma forma reduzida do sistema para as duas equações que estão se unindo, da seguinte forma;

$$\begin{aligned}
P_1 &= h_1(q_1, q_2, \tilde{e}(q_1, q_2, Y, W, \tilde{\eta}, \tilde{\beta}), Y, \eta_1) \\
P_2 &= h_2(q_1, q_2, \tilde{e}(q_1, q_2, Y, W, \tilde{\eta}, \tilde{\beta}), Y, \eta_2) \\
P_1 &= r_1(q_1, q_2, Y, W, \eta_1, \tilde{\eta}, \tilde{\beta}) \\
P_2 &= r_2(q_1, q_2, Y, W, \eta_2, \tilde{\eta}, \tilde{\beta})
\end{aligned}$$

A análise empírica se baseia na estimação do sistema composto pelas duas equações do sistema. Uma versão muito comum envolve a estimação de uma versão log linear destas equações, da seguinte forma:

$$\begin{aligned}
\log P_1 &= \eta_{10} + \eta_{11}^{PR} \log q_1 + \eta_{21}^{PR} \log q_2 + \Gamma_1 \log y + \Delta_1 \log w + v_1 \\
\log P_2 &= \eta_{20} + \eta_{12}^{PR} \log q_1 + \eta_{22}^{PR} \log q_2 + \Gamma_2 \log y + \Delta_2 \log w + v_1
\end{aligned}$$

O significado dos sobrescritos PR em cada uma das elasticidades diz respeito ao fato que estas elasticidades são *residuais* pois levam em consideração as ações das $n-2$ empresas envolvidas. Além disso, o termo *parcial* se aplica pelo fato que o comportamento da empresa parceira na fusão ainda não foi especificada[2]. Em especial, podemos relacionar as elasticidades parciais residuais com as elasticidades tradicionais com a seguinte equação:

$$\begin{aligned} \eta_{11}^{PR} &= \eta_{11} + \sum_{j>2} \eta_{j1} \times \varepsilon_{j1} \\ \eta_{21}^{PR} &= \eta_{21} + \sum_{j>2} \eta_{j2} \times \varepsilon_{j2} \\ \eta_{12}^{PR} &= \eta_{12} + \sum_{j>2} \eta_{j1} \times \varepsilon_{j1} \\ \eta_{22}^{PR} &= \eta_{22} + \sum_{j>2} \eta_{j2} \times \varepsilon_{j2} \end{aligned}$$

Sendo que η_{j1} e η_{j2} são as elasticidades cruzadas – inversas – das demandas da empresa j com relação aos preços de 1 e 2, e ε_{j1} e ε_{j2} são as elasticidades da função melhor resposta da empresa j com respeito às ações da empresa 1 e 2, respectivamente.

Para resolvermos o problema da identificação dos coeficientes decorrentes da simultaneidade, Baker e Breshanan puxam os seus instrumentos da relação de oferta:

$$\begin{aligned} CMg_1(q_1, W, W_1, \beta_1) &= P_1 + q_1 t_1(q_1, q_2, Y, W, \eta_1, \tilde{\eta}, \tilde{\beta}) \\ CMg_2(q_1, W, W_1, \beta_2) &= P_2 + q_2 t_2(q_1, q_2, Y, W, \eta_2, \tilde{\eta}, \tilde{\beta}) \end{aligned}$$

Eles utilizam as variáveis deslocadoras de custos da empresa W_1 e W_2 como instrumentos para as quantidades q_1 e q_2. Uma nota interessante é que, mesmo que não tivéssemos instrumentos para q_1e q_2, os resultados seriam viesados no sentido de negar a hipótese de existir poder de mercado.

9.2.1 Elasticidade da Demanda Residual e o Markup

Qual é a relação desta elasticidade com o famoso índice de Lerner? Baker e Breshanan mostram esta relação, supondo que a demanda residual seja calculada para uma empresa apenas. Supondo que ela seja líder à la Stackelberg, de forma que:

$$\frac{P_1 - CMg_1}{P_1} = -\eta_1^R$$

Da mesma forma, se as empresas estiverem em um equilíbrio consistente em conjecturas, a equação acima vale, com a diferença que $\eta_1^R = \eta_{11} + \sum_{j \notin 1} \eta_{j1} \hat{\varepsilon_{j1}}$, em que $\hat{\varepsilon}_{j1}$é uma conjectura consistente.

[2]Uma observação importante é que podemos escrever a função demanda residual somente; para tanto, precisamos apenas redefinir o comportamento das outras empresas apenas para deixar livre a primeira empresa.

9.3 Resíduo de Solow

Uma última forma de se estimar o poder de mercado – ou, no mínimo, checar se há violações da hipótese de competição perfeita – foi a desenvolvida por Hall (1988)[30] e Hall (1986)[29] para a indústria manufatureira norte-americana. O ponto de partida dos artigos é o cálculo do resíduo de Solow de uma função de produção:

$$\theta = \Delta \ln\left(\frac{Q}{K}\right) - \tilde{\alpha}\Delta \ln\left(\frac{L}{K}\right) - \tilde{\beta}\Delta \ln\left(\frac{M}{K}\right)$$

Em que Q é o valor da produção, K o estoque de capital, L o estoque de mão de obra, M a quantidade de matéria-prima utilizada, $\tilde{\alpha} = \frac{wL}{pQ}$ é o share da remuneração da mão-de-obra no valor da produção e $\tilde{\beta} = \frac{rM}{pQ}$ a participação da remuneração das matérias-primas no valor da produção. Sob a premissa de competição perfeita e retornos constantes de escala, Hall mostra que este termo θ calculado não deve ser correlacionado com nenhuma variável que seja afetada pelos choques de produtividade nem ser causada por variáveis que causam choques de produtividade. Para fazer esta estimativa, o autor desenvolve um teste baseado na seguinte regressão:

$$\theta = \phi_0 + \phi_1 I + \varepsilon$$

E testando se $\phi_1 = 0$, para uma ampla gama de instrumentos – preços do petróleo, gastos miltares, oferta de moeda ou população. Nestes artigos, Hall também apresenta uma forma alternativa de estimação da margem preço-custo marginal, a partir da seguinte equação:

$$\Delta \ln\left(\frac{Q}{K}\right) = \mu\left[\tilde{\alpha}\Delta \ln\left(\frac{L}{K}\right) + \tilde{\beta}\Delta \ln\left(\frac{M}{K}\right)\right]$$

Em competição perfeita, teríamos que $\mu = 1$, e fora de competição teríamos que $\mu = \frac{p}{CMg}$. No entanto, não temos como interpretar isso adequadamente a menos que tenhamos informações sobre as elasticidades de demanda, conforme Shapiro (1987)[43].

Capítulo 10

Estimação de Modelos de Conduta

10.1 Modelos de Simulação

Vamos agora estender a metodologia discutida anteriormente para que possamos entender uma das mais importantes aplicações de métodos empíricos em Organização Industrial: a elaboração de modelos de simulação para os efeitos unilaterais de uma fusão. Vamos supor que existam n empresas, cada uma com um produto. Originalmente cada uma das empresas tem a seguinte função lucro:

$$\pi_i = (p_i - CMg_i) \times q_i(\mathbf{p})$$

Maximizando este negócio, temos a seguinte CPO:

$$q_i(\mathbf{p}) + (p_i - CMg_i) \times \frac{\partial q_i}{\partial p_i} = 0$$

Reorganizando:

$$\frac{p_i - CMg_i}{p_i} = \frac{-q_i(\mathbf{p})}{\frac{\partial q_i}{\partial p_i}} = -\frac{1}{\epsilon_{ii}(\mathbf{p})}$$

Agora vamos supor que a empresa 1 e a empresa 2 decidam se unir, formando apenas uma empresa. Neste caso, o ente resultante da fusão deve ter a seguinte condição de primeira ordem para o produto da ex-empresa 1:

$$q_1(\mathbf{p}) + (p_1 - CMg_1^p) \times \frac{\partial q_1(\mathbf{p})}{\partial p_1} + (p_2 - CMg_2^p) \times \frac{\partial q_2(\mathbf{p})}{\partial p_1} = 0$$

Na verdade, temos um efeito adicional, decorrente do fato que a decisão de precificação do produto 1 deve levar em consideração também os efeitos que esta precificação tem sobre o produto

2. Multiplicando os dois lados da equação por p_1 e dividindo pela receita total da indústria $\sum_{k=1}^{N} q_k(\mathbf{p})p_k$, temos:

$$\frac{p_1 q_1}{\sum_k q_k p_k} + \frac{p_1 q_1}{\sum_k q_k p_k}\frac{(p_1 - CMg_1^p)}{p_1}\frac{\partial q_1}{\partial p_1}\frac{p_1}{q_1} + \frac{p_2 q_2}{\sum_k q_k p_k}\frac{(p_2 - CMg_2^p)}{p_2}\frac{\partial q_2}{\partial p_1}\frac{p_1}{q_2} = 0$$

Definindo o share de valor como $s_1 = \frac{p_1 q_1}{\sum_k q_k p_k}$ e definindo as elasticidades preço e elasticidade cruzadas, reescrevemos a equação acima da seguinte forma:

$$s_1 + s_1\frac{p_1 - CMg_1^p}{p_1}\epsilon_{11} + s_2\frac{(p_2 - CMg_2^p)}{p_2}\epsilon_{21} = 0$$

Para as outras empresas j, podemos reescrever estas CPO da seguinte forma:

$$s_j + +s_j\frac{p_j - CMg_j^p}{p_j}\epsilon_{jj} = 0$$

Neste caso, podemos reescrever este equilíbrio pós-fusão da seguinte forma:

$$\begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_n \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \epsilon_{21} & 0 & \cdots & 0 \\ \epsilon_{12} & \epsilon_{22} & 0 & \cdots & 0 \\ 0 & 0 & \epsilon_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \epsilon_{nn} \end{pmatrix} \begin{pmatrix} s_1\frac{p_1 - CMg_1^p}{p_1} \\ s_2\frac{p_2 - CMg_2^p}{p_2} \\ s_3\frac{p_3 - CMg_3^p}{p_3} \\ \vdots \\ s_n\frac{p_n - CMg_n^p}{p_n} \end{pmatrix} = 0$$

Em forma matricial, podemos escrever esta equação como $\mathbf{s}+\mathbf{E}\mathbf{w} = \mathbf{0}$. Supondo que os analistas tenham estimativas para as elasticidades e os shares de valor após a fusão, podemos obter a nova matriz $\mathbf{w}$ da seguinte forma:

$$\mathbf{w} = -\mathbf{E}^{-1}\mathbf{s}$$

Podemos recuperar os novos markups multiplicando elemento por elemento da matriz $\mathbf{w}$ pelo inverso da matriz $\mathbf{s}$, da seguinte forma:

$$\frac{p_i - CMg_i}{p_i} = \frac{w_i}{s_i}$$

Se supusermos que os custos marginais são os mesmos antes e depois, podemos derivar as variações nos preços ao consumidor.

Capítulo 11

Estrutura Vertical

Parte IV

Modelos de Entrada

Capítulo 12

Modelos de Entrada

Nesta aula, iremos mudar o foco de nossa análise, para investigarmos a questão da entrada e saída das empresas em um determinado mercado. Em especial, esta discussão se filia à questão de como entender a estrutura de mercado e a relação entre esta e a competição dentre deste mesmo mercado. Em particular, uma questão que pode ser enfrentada é: como o número e organização das empresas neste mercado, o tamanho das mesmas e dos competidores em potencial, bem como a forma das linhas de produtos das empresas afeta a competição e os lucros das mesmas.

A discussão sobre este tema variou bastante ao longo do tempo. Nos anos 50 a 70, as análises econométricas eram voltadas para como as variáveis tais como lucro das empresas, gastos com P&D, e preços variavam em mercados concentrados ou pouco concentrados. A maior parte destes trabalhos assumiam que a estrutura de mercado era exógena à decisão sobre estas variáveis. Nos anos 80, os modelos teóricos de OI eram focados em entender como o comportamento estratégico poderia influenciar a estrutura de mercado. Tais modelos consideravam a estrutura de mercado como o resultado de um jogo em duas etapas, sendo que a primeira envolvia a decisão de entrada ou saída, enquanto que a segunda etapa modelava a competição entre as empresas que decidiram pela entrada.

Do ponto de vista empírico, a maior disponibilidade de dados dos censos industriais permitiu documentar ricos padrões de entrada e saída das empresas, bem como mudanças na estrutura do mercado. Nesta aula, iremos descrever como a análise empírica mais moderna sobre este tema se baseia, utilizando modelos de teoria dos jogos para construir modelos econométricos de entrada, saída e concentração de mercado. Em especial, iremos discutir as previsões que estes modelos fazem sobre os efeitos sobre a estrutura de mercado das seguintes variáveis:

- Tamanho e irrecuperabilidade dos custos fixos não recuperáveis;
- Sensibilidade dos lucros das empresas à entrada e saída de competidores;
- O grau de substitutibilidade dos produtos

- Expectativas dos potenciais entrantes sobre a competição após a entrada propriamente dita;
- A existência e a eficiência dos entrantes em potencial
- A endogeneidade dos custos fixos e dos custos irrecuperáveis.

Vamos começar a nossa análise com uma estrutura bem simples de estudo, na seção seguinte.

12.1 Modelos de Escolhas Estratégicas Discretas

Vamos começar a ilustrar como é esta metodologia com um problema bem simples. Vamos supor que existam dados sobre as decisões de entrada das firmas – supostas homogêneas – e queremos estimar os custos fixos de produção das mesmas. Além disso, vamos supor que se observa um grande número de mercados regionais distintos e que existem dados sobre a demanda de mercado e os custos dos insumos das empresas. Se estes mercados fossem competitivos – o que implicaria que as decisões de entrada dos diferentes agentes são independentes – poderíamos modelar esta entrada das empresas como um modelo de escolha discreta. No entanto, quando estamos falando de mercados oligopolizados, as expectativas das empresas acerca do comportamento dos seus competidores também afetam as decisões de entrada, e vice-versa. Para que possamos saber até que ponto os maiores custos fixos afetam a concentração de mercado ou se o comportamento dos competidores afeta os lucros, precisamos de modelos estatísticos para decisões interdependentes.

Para isto, vamos começar com um modelo bem simples, para investigar como os custos fixos afetam a decisão de entrada de duas empresas. Vamos definir $a_i = 1$ como sendo o evento de entrada no mercado da empresa i, e $a_i = 0$ o evento de não entrada da empresa i. Podemos representar a decisão estratégica das empresas com a ajuda da seguinte matriz de *payoffs*:

Figura 12.1: Payoffs das duas empresas

		Empresa 1	
		$a_1 = 0$	$a_1 = 1$
Empresa 2	$a_2 = 0$	Π^1_{00},Π^2_{00}	Π^1_{10},Π^2_{10}
	$a_2 = 1$	Π^1_{01},Π^2_{01}	Π^1_{11},Π^2_{11}

Supondo que os dados observados venham de um jogo sem repetição e simultâneo, e que possamos restringir a nossa atenção aos equilíbrios de Nash em estratégias puras, as estratégias de equilíbrio dos agentes podem ser representadas pelas seguintes condições:

$$a_1^* = \begin{cases} 1 & se\, \pi_1^* \geq 0 \\ 0 & se\, \pi_1^* < 0 \end{cases}$$

$$a_2^* = \begin{cases} 1 & se\, \pi_2^* \geq 0 \\ 0 & se\, \pi_2^* < 0 \end{cases}$$

Sendo que π^* representa os lucros adicionais para a empresa decorrentes da decisão de entrar. Em especial, $\pi_1^* = (1 - a_2)(\Pi_{10}^1 - \Pi_{00}^1) + a_2(\Pi_{11}^1 - \Pi_{01}^1)$ e $\pi_2^* = (1 - a_1)(\Pi_{01}^2 - \Pi_{00}^2) + a_1(\Pi_{11}^2 - \Pi_{10}^2)$. Uma vez que as empresas somente incorrem em custos fixos se elas decidirem pela entrada, estas diferenças de lucros apresentam os custos fixos como termos separados. Para a estimação econométrica, precisamos especificar a distribuição da parte variável dos lucros (o ganho de lucros além do custo fixo que as empresas auferem no caso da entrada propriamente dita). Seguindo Bresnahan e Reiss (1991), podemos escrever estes lucros incrementais da entrada da seguinte forma:

$$\begin{aligned} \pi_1^* &= X_1\beta_0^1 + a_2\Delta_1^1 - \varepsilon^1 \\ \pi_2^* &= X_2\beta_0^2 + a_1\Delta_1^2 - \varepsilon^2 \end{aligned}$$

Podemos entender estas equações da seguinte forma. Normalizando os lucros na eventualidade da não entrada em zero, temos que $\pi_1^* = \Pi_{10}^1 + a_2(\Pi_{11}^1 - \Pi_{10}^1)$. O primeiro dos termos representa os lucros que a empresa 1 desfrutaria caso fosse monopolista neste mercado, que podemos assumir que é uma função de elementos observáveis, alguns parâmetros e componentes aleatórios, dando $\Pi_{10}^1 = X_1\beta_0^1 - \varepsilon^1$. Além disso, o termo entre parênteses representa a diferença nos lucros da empresa 1 decorrente da entrada da empresa 2. Este elemento podemos deixar como sendo um parâmetro a ser estimado e denotado por Δ_1^1.

Estas considerações fazem com que possamos especificar um modelo econométrico e estimar os parâmetros relevantes. No entanto, ainda temos dois problemas, um de natureza teórica e outro de natureza econométrica. O problema de natureza econométrica é que, dado que consideramos Δ_1^1 como um parâmetro a ser estimado, se não pusermos nenhuma restrição sobre o suporte da distribuição do termo ε, não conseguiremos identificar adequadamente os parâmetros β. O problema de natureza teórica é que nem sempre os valores são tais que garantem a existência e a unicidade do equilíbrio (ou seja, pode haver um equilíbrio múltiplo, em que apenas uma das empresas entra).

A solução para estes problemas passou pelo fato de se considerar os resultados não únicos como observacionalmente equivalentes. Esta foi a abordagem de Bresnahan e Reiss (1990). Esta restrição implica que o modelo econométrico deixa de ser um modelo especificamente voltado para a decisão de entrada, mas sim um modelo para a previsão do número de entrantes em um determinado mercado. No caso de um duopólio, a função de verossimilhança contém as seguintes assertivas sobre as probabilidades de ocorrência:

$$\begin{aligned} \Pr(a_1 = 0, a_2 = 0) &= \Pr(X\beta_0^1 < \varepsilon^1, X\beta_0^2 < \varepsilon^2) \\ \Pr(a_1 = 1, a_2 = 1) &= \Pr(X\beta_0^1 + \Delta_1^1 > \varepsilon^1, X\beta_0^1 + \Delta_1^2 > \varepsilon^2) \\ \Pr(a_1 = 1 \vee a_2 = 1) &= 1 - \Pr(a_1 = 0, a_2 = 0) - \Pr(a_1 = 1, a_2 = 1) \end{aligned}$$

A distribuição dos termos Δ_1^i determinaria, então, a forma funcional específica para estas probabilidades. Evidentemente, esta distribuição deve respeitar as restrições sobre os *payoffs* dos jogadores e permitir a identificação dos parâmetros chave. Uma forma utilizada por Bresnahan e Reiss (1990 e 1991) foi modelar este negócio permitindo heterogeneidades não observáveis entre os jogadores. Por exemplo, estes autores usam $\Delta_1^i = g(Z\gamma^i) + \eta^i$, em que $g(\cdot)$ é uma função definida no ramo negativo dos números reais e η^i é uma variável aleatória com um limite superior de zero. Primeiro colocaremos um modelo mais geral, para depois abordar especificamente a forma pela qual estes artigos chegaram aos seus resultados, de forma a nos familiarizar com as suas premissas.

12.1.1 Um Modelo Geral de firmas homgêneas

Como vimos anteriormente, para evitarmos o problema da identificação dos modelos econométricos, bem como a questão de como modelar os equilíbrios múltiplos, a solução é considerar os resultados múltiplos como um único elemento e, depois, modelar a questão sobre o número de empresas que entram. Iremos, em especial, discutir o que poderemos aprender sobre as primitivas do modelo quando observamos um vetor de quantidades $N_1, N_2, \cdots, N_T$, que entraram em T mercados diferentes. Para isto, precisamos relacionar o N_T observado com os lucros não observados no mercado T. Dados N_i entrantes no mercado T , temos que os lucros de cada uma delas é dado por:

$$\pi_i(N_i) = V(N_i, x_i, \theta) - F_i$$

Em que $V(\cdot)$ representa os lucros variáveis totais e F o custo fixo. Vamos assumir que os custos fixos, que também não são observados pelo econometrista são distribuídos de acordo com $\Phi(F_i|x_i, \omega)$. Com esta função lucro, podemos ligar as decisões de entrada ao número observado de empresas. Para as N^* empresas que entraram, temos:

$$V(N^*, x, \theta) - F > 0$$

Enquanto que para qualquer entrante potencial, temos que:

$$V(N^* + 1, x, \theta) - F < 0$$

Combinando estas duas desigualdades, temos um limite para os custos fixos:

$$V(N^*, x, \theta) \geq F > V(N^* + 1, x, \theta)$$

Estes limites permitem que possamos estimar os parâmetros da função lucro variável e os custos fixos a partir da observação do vetor x e do número de empresas:

$$\begin{aligned}\Pr(V(N^*, x) &\geq F|x) - \Pr(V(N^* + 1, x) > F|x) = \\ &= \Phi(V(N^*, x, \theta)|x) - \Phi(V(N^* + 1, x, \theta)|x)\end{aligned}$$

Supondo amostras independente eidenticamente distribuídas, temos que a amostra possui uma função de verossimilhança "ordenada" da seguinte forma:

$$LL(\theta|\{x, N^*\}) = \sum_t (\ln(V(N_t^*, x_t)) - \ln(V(N_t^* + 1, x_t)))$$

A pergunta, para a qual veremos algumas respostas mais adiante, é como especificar esta função de lucros variáveis. Uma delas seria montar a função $V(\cdot)$ de tal forma que ela torne a estimação simples e, ao mesmo tempo, atenda a restrição que a função seja não crescente em N. A segunda abordagem é baseada em especificar a função $V(\cdot)$diretamente a partir de premissas de mercado e hipóteses sobre o jogo após a entrada. Vamos ver dois exemplos deste tipo de abordagem nos papers a seguir.

12.1.2 Bresnahan e Reiss (1990)

Estes autores, em seu artigo, fazem a suposição de que os resultados que equivalem a equilíbrios múltiplos são observacionalmente equivalentes. Neste caso, eles se focam no número de empresas – que, neste caso, podem ser zero, uma ou duas empresas. A função de lucro de cada empresa é dada por:

$$\Pi_i^N = V_i^N \times S(Y) - F_i^N$$

Sendo que a empresa i pode ser monopolista $N = M$, ou duopolista $N = D$. A função V_i representa os "lucros variáveis por consumidor", ou seja, a margem entre preço e custo marginal, multiplicada pela função demanda individual. O termo $S(Y)$ seria uma medida do tamanho do mercado, enquanto que F_i^N é uma medida dos custos fixos. A modelagem é refinada ao assumirmos uma parte não observável para os lucros variáveis por consumidor, bem como para os custos fixos, o que deixa a função lucro da seguinte forma:

$$\begin{aligned}
\Pi_i^N &= [\bar{V}_i^N + \eta_i^N] \times S(Y) - \bar{F}_i^N + \varepsilon_i^N \\
&= \bar{V}_i^N \times S(Y) - \bar{F}_i^N + \xi_i^N \\
&= \bar{\Pi}_i^N + \xi_i^N
\end{aligned}$$

Acredito que as similaridades entre a forma colocada anteriormente e a que desenvolvemos agora estejam bastante claras. Caso não tenhamos elementos não observáveis sobre os lucros variáveis por consumidor, o termo ξ_i^N fica igual ao termo ε_i^N, de forma que podemos estimar este negócio por um modelo PROBIT ordenado. A diferença entre o modelo PROBIT multinomial (ou seja, a extensão do modelo PROBIT para o caso de mais de dois resultados) para o modelo PROBIT multinomial reside no fato que a ordem dos valores que a variável representativa dos lucros possui relevância para a análise. Caso tenhamos elementos não observáveis sobre os lucros variáveis, chegamos ao modelo PROBIT ordenado, só que o termo ξ_i^N é inerentemente heterocedástico, pois o termo ε_i^N ali presente está multiplicado pela função $S(Y)$. Neste caso, as funções de probabilidade ficariam sendo:

$$\begin{aligned} P_0 &= 1 - \Phi(\bar{\Pi}^M/\sigma_\xi) \\ P_2 &= \Phi(\bar{\Pi}^D/\sigma_\xi) \\ P_1 &= 1 - P_0 - P_2 \end{aligned}$$

Em que uma restrição deveria ser imposta para que $\bar{\Pi}^M \geq \bar{\Pi}^D$. Com relação às precisas formas funcionais utilizadas, elas foram:

$$S(Y) = TOWNPOP + \lambda_1 OPOP10 + \lambda_2 NGRW70 + \lambda_3 PGRW70 + \lambda_4 OCTY$$

Em que $TOWNPOP$ representa a população da cidade, $OPOP10$ representa o montante de demanda das pessoas em volta da cidade, $NGRW70$ e $PGRW70$ são as taxas de crescimento negativas e positivas no tamanho da população entre 1970 e 1980. $OCTY$ é a fração dos residentes da cidade que comutam para fora da cidade. Com relação à função lucros variáveis por consumidor, temos a seguinte função:

$$V^N = \theta^M + \theta^D D + Z\theta_Z + W\theta_W$$

Em que os θ são parâmetros a serem estimados, D é uma *dummy* com valor de um caso exista mais de uma empresa no mercado, e W e Z são variáveis que explicariam variações na demanda e nos custos em cada uma das regiões. Os autores colocam variáveis tais como renda média dos consumidores, a mediana da idade e dos anos de estudo, bem como o salário médio no varejo. Com relação aos custos fixos, os autores colocam a seguinte especificação:

$$F^N = \gamma^N + \gamma^D D + \gamma_W RETWAGE + \gamma_L LANDVAL$$

Em que $RETAWGE$ é o salário médio do varejo na região e $LANDVAL$ é o valor médio do acre.

As variáveis relevantes para o estudo são os valores de S^De S^M, o mínimo tamanho de mercado que é viável a sustentação de duas e uma empresa, respectivamente. Além disso, outro elemento importante é $\frac{V^D}{V^M}$ mensura a fração pela qual os lucros variáveis por consumidor caem com a entrada da segunda firma. Por exemplo, se os duopolistas vendessem produtos iguais, esta razão deveria ser igual a 0,5. Além disso, outra coisa interessante envolve a razão $\frac{F^D}{F^M}$.

12.1.3 Breshanan e Reiss (1991)

O segundo dos artigos em que temos a aplicação e o desenvolvimento desta metodologia também é de Breshanan e Reiss, publicado no JPE em 1991. Este artigo é uma extensão da metodologia anterior para o caso de um oligopólio. Ele faz a mesma premissa de equivalência entre os diferentes equilíbrios múltiplos em estratégias puras em um jogo de entrada,assumindo a seguinte função para os lucros de uma entrante em potencial:

$$\Pi_N = S(\mathbf{Y}, \lambda) \times V_N(\mathbf{Z}, \mathbf{W}, \alpha, \beta) - F(\mathbf{W}, \gamma) + \varepsilon$$

A verossimilhança deste negócio é dada por:

$$\begin{aligned} P(N &= 0) = P(\Pi_1 < 0) = 1 - \Phi(\bar{\Pi}_1) \\ P(N = n) &= P(\bar{\Pi}_n > 0 \wedge \bar{\Pi}_{n+1} < 0) = \Phi(\bar{\Pi}_n) - \Phi(\bar{\Pi}_{n+1}) \\ P(N \geq n^{max}) &= \Phi(\bar{\Pi}_{n^{max}}) \end{aligned}$$

A forma da função de lucros variáveis por consumidor é dada por

$$V_N = \alpha_1 + \mathbf{X}\beta + \sum_{n=2}^{N} \alpha_n$$

Neste paper, eles calculam os "limites de entrada", ou seja, o menor tamanho de mercado necessário para sustentar exatamente N empresas, que é dado por $S_N^* = \frac{\bar{F}}{V_N}$, bem como o limite de entrada por empresa, que seria $\frac{S_N^*}{N}$.

Referências Bibliográficas

[1] Anderson, S., de Palma, A., and Thisse, J.-F. *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, 1992.

[2] Appelbaum, E. The estimation of the degree of oligopoly power. *Journal of Econometrics 19*, 2-3 (August 1982), 287–299.

[3] Asche, F., and Wessells, C. R. On price indices in the almost ideal demand system. *American Journal of Agricultural Economics 79*, 4 (November 1997), 1182–1185.

[4] Baker, J. B., and Baresnahan, T. F. The gains from merger or collusion in product-differentiated industries. *Journal of Industrial Economics 33*, 4 (June 1985), 427–44.

[5] Baker, J. B., and Bresnahan, T. F. Estimating the residual demand curve facing a single firm. *International Journal of Industrial Organization 6*, 3 (1988), 283–300.

[6] Baker, J. B., and Rubinfeld, D. L. Empirical methods in antitrust litigation: Review and critique. *American Law and Economics Review 1*, 1-2 (Fall 1999), 386–435.

[7] Barnett, W., and Serletis, A. The differential approach to demand analysis and the rotterdam model. WORKING PAPERS SERIES IN THEORETICAL AND APPLIED ECONOMICS 200902, University of Kansas, Department of Economics, Jan. 2009.

[8] Barten, A. P. The systems of consumer demand functions approach: A review. *Econometrica 45*, 1 (January 1977), 23–51.

[9] Berry, S., Levinsohn, J., and Pakes, A. Automobile prices in market equilibrium. *Econometrica 63*, 4 (1995), 841–890.

[10] Berry, S. T. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*, 2 (Summer 1994), 242–262.

[11] Bresnahan, T. The apple-cinnamon cheerios war: Valuing new goods, identifying market power, and economic measurement. Unpublished Manuscript, July 1997.

[12] Bresnahan, T. F. Departures from marginal-cost pricing in the american automobile industry : Estimates for 1977-1978. *Journal of Econometrics 17*, 2 (November 1981), 201–227.

[13] Bresnahan, T. F. The oligopoly solution concept is identified. *Economics Letters 10*, 1-2 (1982), 87–92.

[14] Bresnahan, T. F., Stern, S., and Trajtenberg, M. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *The RAND Journal of Economics 28* (1997), S17–S44.

[15] Brown, A., and Deaton, A. S. Surveys in applied economics: Models of consumer behaviour. *Economic Journal 82*, 328 (December 1972), 1145–1236.

[16] Buse, A. Evaluating the linearized almost ideal demand system. *American Journal of Agricultural Economics 76*, 4 (November 1994), 781–793.

[17] Buse, A., and Chan, W. H. Invariance, price indices and estimation in almost ideal demand systems. *Empirical Economics 25*, 3 (2000), 519–539.

[18] Carlton, D. W., and Perloff, J. M. *Modern Industrial Organization (4th Edition) (Adison-Wesley Series in Economics)*. Addison Wesley, May 2004.

[19] Christensen, L. R., Jorgenson, D. W., and Lau, L. J. Transcendental logarithmic utility functions. *The American Economic Review 65*, 3 (1975), 367–383.

[20] Corts, K. S. Conduct parameters and the measurement of market power. *Journal of Econometrics 88*, 2 (November 1998), 227–250.

[21] Davidson, R., and Mackinnon, J. G. *Estimation and Inference in Econometrics*. Oxford University Press, USA, December 1992.

[22] Deaton, A., and Muellbauer, J. An almost ideal demand system. *The American Economic Review 70*, 3 (1980), 312–326.

[23] Diewert, W., and Wales, T. Multiproduct cost functions and subadditivity tests : A critique of the evans and heckman research on the u.s. bell system. UBC Departmental Archives 91-21, UBC Department of Economics, 1991.

[24] Evans, D. S., and Heckman, J. J. *Multiproduct Cost Function Estimates and Natural Monopoly Tests for the Bell System*, 1 ed. Elsevier, Amsterdam, 1983, ch. 10, pp. 253–282.

[25] Evans, D. S., and Heckman, J. J. A test for subadditivity of the cost function with an application to the bell system. *American Economic Review 74*, 4 (September 1984), 615–23.

[26] Evans, D. S., and Heckman, J. J. A test for subadditivity of the cost function with an application to the bell system: Erratum. *American Economic Review 76*, 4 (September 1986), 856–58.

[27] Gollop, F. M., and Roberts, M. J. Firm interdependence in oligopolistic markets. *Journal of Econometrics 10*, 3 (August 1979), 313–331.

[28] Greene, W. H. *Econometric Analysis*. Prentice Hall, August 2002.

[29] Hall, R. E. Market structure and macroeconomic fluctuations. *Brookings Papers on Economic Activity 17*, 1986-2 (1986), 285–338.

[30] Hall, R. E. The relation between price and marginal cost in u.s. industry. *Journal of Political Economy 96*, 5 (October 1988), 921–47.

[31] Johnston, J. *Econometric Methods*. McGraw Hill Higher Education, Boston, 1997.

[32] Klein, L. R., and Rubin, H. A constant-utility index of the cost of living. *The Review of Economic Studies 15*, 2 (1947), 84–87.

[33] Lau, L. J. On identifying the degree of competitiveness from industry price and output data. *Economics Letters 10*, 1-2 (1982), 93–99.

[34] Lau, L. J. Functional forms in econometric model building. In *Handbook of Econometrics*, Z. Griliches and M. D. Intriligator, Eds., vol. 3 of *Handbook of Econometrics*. Elsevier, 1986, ch. 26, pp. 1515–1566.

[35] Mas-Colell, A., Whinston, M. D., and Green, J. R. *Microeconomic Theory*. Oxford University Press, June 1995.

[36] Panzar, J. C. Technological determinants of firm and industry structure. In *Handbook of Industrial Organization*, R. Schmalensee and R. Willig, Eds., vol. 1 of *Handbook of Industrial Organization*. Elsevier, 1989, ch. 1, pp. 3–59.

[37] Panzar, J. C., and Rosse, J. N. Testing for 'monopoly'; equilibrium. *Journal of Industrial Economics 35*, 4 (June 1987), 443–56.

[38] Pindyck, R. S., and Rubinfeld, D. L. *Microeconomics*, 5th ed. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2005.

[39] Pollak, R., and Wales, T. J. *Demand System Specification and Estimation*. Oxford University Press, USA, City, 1995.

[40] Reiss, P. C., and Wolak, F. A. Structural econometric modeling: Rationales and examples from industrial organization. In *Handbook of Econometrics*, J. Heckman and E. Leamer, Eds. Elsevier, 2007.

[41] Rosen, H. S., and Small, K. A. Applied welfare economics with discrete choice models. NBER Working Papers 0319, National Bureau of Economic Research, Inc, May 1981.

[42] Scherer, F. M. *Industrial market structure and economic performance*, 3rd ed ed. Houghton Mifflin, Boston, 1990. Incluye referencias bibliográficas e índice.

[43] Shapiro, M. D. Measuring market power in u.s. industry. NBER Working Papers 2212, National Bureau of Economic Research, Inc, 1987.

[44] Shy, O. *Industrial Organization: Theory and Applications*, vol. 1 of *MIT Press Books*. The MIT Press, January 1996.

[45] Stigler, G. J. The early history of empirical studies of consumer behavior. *Journal of Political Economy 62*, 2 (1954), 95.

[46] Theil, H. The information approach to demand analysis. *Econometrica 33*, 1 (1965), pp. 67–87.

[47] Tirole, J. *The Theory of Industrial Organization*. The MIT Press, January 1988.

[48] Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2003.

[49] Varian, H. *Intermediate microeconomics: a modern approach*. WW Norton & Co., 2006.

[50] Varian, H. R. *Microeconomic Analysis*. W. W. Norton & Company, February 1992.

[51] Wooldridge, J. *Introductory Econometrics*. South-Western Cengage Learning, Mason, 2009.

[52] Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*, 298 (1962), 348–368.

Apêndice A

Fundamentos de Microeconomia

Nesta aula, iremos detalhar alguns pontos da teoria do consumidor que serão importantes para a análise econométrica subseqüente. Em especial:

- Relação entre o problema primal e o problema dual da escolha do consumidor.
- Relação entre as funções derivadas dos dois problemas.

Tais relações são importantes, na medida em que nos permitem entender muitas das derivações que veremos quando da análise do lado da demanda que será discutida nos capítulos relevantes mais adiante. O texto básico para esta aula foi o Mas-Collell, Whinston e Green [35], em especial os capítulos 3 e 5, ainda que o livro do Varian (1985)[50], caps. 7, 8 e 10, sejam referências úteeis.

A.1 O problema Primal

O problema que chamaremos de "problema primal" de escolha do consumidor é o chamado **Problema de Maximização da Utilidade do Consumidor.** O primeiro passo deste problema é a existência de uma função de utilidade contínua a partir de preferências, completas e transitivas, o que é estabelecido na Proposição 3.C.1 da página 47 do MWG. Dado isso, temos que o problema primal é

$$\begin{array}{rl} Max & u(\mathbf{x}) \\ t.q. & \mathbf{p}\cdot\mathbf{x} \leq w \end{array}$$

Ou seja, o consumidor escolhe uma cesta de consumo no conjunto orçamentário Walrasiano $B_{p,w} = \{x \in \mathbb{R}^L_+ : \mathbf{p}\cdot\mathbf{x} \leq w\}$ para maximizar a sua utilidade. Em palavras, o conjunto orçamentário Walrasiano é aquele conjunto delimitado pela reta de restrição orçamentária, especialmente comum em livros de microeconomia da graduação, como por exemplo, o Varian "baby"[49] e o Pindyck e Rubinfeld [38]. Dado que existe uma solução para este problema (Prop. 3.D.1, MWG), vamos detalhar algumas funções associadas com a solução deste problema. Esta solução do problema, denominada $x^*(\mathbf{p}, w)$ é chamada **função de demanda marshalliana**.

Antes de qualquer coisa, é importante contextualizar melhor o que entendemos por função demanda marshalliana, pois na verdade, a resposta implícita pelo problema de maximização e incorporada em x^* é um vetor, em que cada elemento seria a função demanda marshalliana de cada um dos itens componentes do vetor x.

A primeira destas funções é a chamada **função de utilidade indireta**. Basicamente, para cada solução do problema primal, existe uma relação entre os vetores de preços e renda e o nível de x de solução do problema – que denotaremos x^* (esta relação é a nossa querida função demanda Marshalliana). Com base nesta relação, podemos substituir $(\mathbf{p}, w)$ de volta na função utilidade, fazendo que exista uma função utilidade $u(x^*(\mathbf{p}, w)) \equiv v(\mathbf{p}, w)$. Esta função v é denominada **função de utilidade indireta.** Esta função possui as seguintes propriedades:

- Homogênea de grau zero nos seus argumentos
- Estritamente crescente em w e não decrescente em p_l, $\forall l$.
- Quasiconvexa: o conjunto $\{(\mathbf{p}, w) : v(\mathbf{p}, w) \leq \bar{v}\}$ é convexo para qualquer $\bar{v}$.
- Contínua nos seus argumentos.

Então, temos as seguintes relações decorrentes do problema primal: função utilidade, função demanda marshalliana e função utilidade indireta. Vamos agora conversar sobre o problema dual, na próxima seção.

A.2 O problema Dual

O problema "dual" ao problema de maximização da utilidade dos consumidores é o **Problema de Minimização dos Gastos do consumidor.** Supondo $\mathbf{p} \gg 0$ e $\bar{u} > u(\mathbf{0})$, o problema é

$$\begin{array}{rl} Min_{x \geq 0} & \mathbf{p} \cdot \mathbf{x} \\ t.q. & u(\mathbf{x}) \geq \bar{u} \end{array}$$

Ou seja, conseguir o menor custo da cesta de consumo sujeita a um determinado nível de utilidade. Da mesma forma que no caso anterior, existe uma relação entre o nível de consumo associado com a solução do problema e as variáveis condicionantes do problema $(\mathbf{p}, u)$. Esta relação é a nossa querida função demanda hicksiana, e é representada por $h(\mathbf{p}, u)$.

Da mesma forma que no caso da função demanda marshalliana, a função demanda hicksiana retorna um vetor de demandas, correspondentes a cada um dos elementos da matriz x. Jogando esta função demanda na função que está se querendo minimizar, temos a chamada **função dispêndio**, que representaremos por $\mathbf{p}\cdot h(\mathbf{p},u) \equiv e(\mathbf{p},u)$, ressalvada a conformabilidade dos vetores $\mathbf{p}$ e $h(\mathbf{p},u)$, para dar um elemento 1×1. Esta função dispêndio possui as seguintes propriedades:

- Homogênea de Grau Um em $\mathbf{p}$.
- Estritamente Crescente em u e não decrescente em $p_l \forall l$.
- Côncava em $\mathbf{p}$
- Contínua em $\mathbf{p}$ e u

A importância desta função de utilidade indireta está no fato que, em geral, os sistemas de demanda como os que veremos mais adiante em geral são derivados com uma premissa sobre a forma da função de utilidade indireta e, a partir daí, utilizamos uma relação entre os dois problemas para derivar as funções propriamente ditas de demanda, pois em muitos casos isto fica muito mais simples do que passar diretamente da função de utilidade direta para as demandas. Em resumo, temos todos os elementos do problema dual: função demanda hicksiana e função dispêndio. Vamos agora trabalhar as relações entre estas abordagens.

A.3 Relações entre o Problema Primal e o Dual

Vamos agora detalhar mais aprofundadamente as relações entre os dois problemas e entre as equações derivadas dos dois problemas. De acordo com a proposição MWG 3.E.1, se $w > 0$ e $u > u(0)$, as soluções dos dois problemas são iguais (o que poderíamos dizer que são as soluções não triviais do problema). A primeira relação que iremos fazer é entre a demanda Hicksiana e a função dispêndio. Dadas as premissas vistas anteriormente, para todos os valores de p e de u, temos que a função demanda Hicksiana é o vetor de derivadas da função dispêndio com respeito aos preços:

$$\begin{aligned} h(\mathbf{p},u) &= \bigtriangledown_p e(\mathbf{p},u) \\ h(\mathbf{p},u) &= \begin{bmatrix} h_1(\mathbf{p},u) \\ \vdots \\ h_k(\mathbf{p},u) \end{bmatrix} = \begin{bmatrix} \frac{\partial e}{\partial p_1} \\ \vdots \\ \frac{\partial e}{\partial p_k} \end{bmatrix} = \nabla_p e(\mathbf{p},u) \end{aligned}$$

Além disso, temos as seguintes propriedades da função demanda hicksiana:

- $\bigtriangledown_p h(\mathbf{p}, u) = \bigtriangledown_p^2 e(\mathbf{p}, u)$, ou

$$\begin{bmatrix} \frac{\partial h_1}{\partial p_1} & \cdots & \frac{\partial h_1}{\partial p_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_k}{\partial p_1} & \cdots & \frac{\partial h_k}{\partial p_k} \end{bmatrix} = \begin{bmatrix} \frac{\partial e^2}{\partial p_1^2} & \cdots & \frac{\partial e^2}{\partial p_1 p_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial e^2}{\partial p_1 p_k} & \cdots & \frac{\partial e^2}{\partial p_k^2} \end{bmatrix}$$

- $\bigtriangledown_p h(\mathbf{p}, u)$ é uma matriz negativa semidefinida e simétrica

- $\bigtriangledown_p h(\mathbf{p}, u) \cdot \mathbf{p} = 0$

A intuição decorrente do último ponto decorre do fato que, uma vez que $h(\mathbf{p}, u) \cdot \mathbf{p} = w$, e temos que w está fixo, as variações nas demandas hicksianas decorrentes de um pequeno aumento de preços tem que ser tais que, tanto antes quanto depois desta alteração nos preços, o consumidor continua gastando o mesmo montante w de reais. A relação seguinte é entre a função demanda hicksiana e a função demanda walrasiana. A principal forma pela qual estas duas equações se relacionam é pela Equação de Slutsky. Formalmente, se $u(\cdot)$ é uma função de utilidade contínua representando uma relação de preferências localmente não saciada e estritamente convexa definida no conjunto de consumo $X = \mathbb{R}_+^k$, para todos $(\mathbf{p}, w)$ e $u = v(\mathbf{p}, w)$, temos que

$$\frac{\partial h_l(\mathbf{p}, u)}{\partial p_k} = \frac{\partial x_l(\mathbf{p}, w)}{\partial p_k} + \frac{\partial x_l(\mathbf{p}, w)}{\partial w} x_k(\mathbf{p}, w)$$

Esta é a chamada **Equação de Slutsky .** Em forma matricial:

$$\bigtriangledown_p h(\mathbf{p}, u) = \bigtriangledown_p x(\mathbf{p}, w) + (\bigtriangledown_w x(\mathbf{p}, w))(x(\mathbf{p}, w))^T$$

Esta proposição implica que a matriz de derivadas de preços $\bigtriangledown_p h(\mathbf{p}, u)$ das demandas hicksianas é igual à matriz

$$S(\mathbf{p}, w) = \begin{bmatrix} s_{11}(\mathbf{p}, w) & \cdots & s_{1k}(\mathbf{p}, w) \\ \vdots & \ddots & \vdots \\ s_{k1}(\mathbf{p}, w) & \cdots & s_{kk}(\mathbf{p}, w) \end{bmatrix}$$

Em que cada casela é igual a $s_{lk}(\mathbf{p}, w) = \frac{\partial x_l}{\partial p_k} + \frac{\partial x_l}{\partial w} x_k$, e é chamada por isso a Matriz de Substituição de Slutsky. Uma observação importante para o que se segue é que ela é diretamente calculável a partir dos valores da função demanda Marshalliana. Por definição, a proposição 3.G.2 do MWG garante que $S(\mathbf{p}, w)$ deve possuir as três propriedades: deve ser **negativa semidefinida, simétrica e satisfazer** $S(\mathbf{p}, w) \cdot \mathbf{p} = \mathbf{0}$**.**

A relação seguinte que iremos estudar é entre as funções demanda Marshalliana e as funções de utilidade indireta. Esta relação, denominada *Identidade de Roy*, pode ser derivada mais formalmente da seguinte forma. Supondo que exista uma função utilidade contínua que representa uma relação de preferências localmente não saciada e estritamente convexa e que a função de utilidade indireta é diferenciável em $(\bar{p}, \bar{w}) \gg 0$. Então

$$x(\bar{\mathbf{p}}, \bar{w}) = -\frac{1}{\bigtriangledown_w v(\bar{p}, \bar{w})} \cdot \bigtriangledown_p v(\bar{\mathbf{p}}, \bar{w})$$

Ou seja, para cada produto:

$$x_l(\bar{\mathbf{p}}, \bar{w}) = -\frac{\partial v(\bar{\mathbf{p}}, \bar{w})/\partial p_l}{\partial v(\bar{\mathbf{p}}, \bar{w})/\partial w}$$

A prova desta afirmação está na Proposição 3.G.4 no MWG, e nos mostra um ponto interessante: a obtenção da função demanda Marshalliana é muito mais fácil partindo da função de utilidade indireta do que da função de utilidade direta. Para derivar $x(\mathbf{p}, w)$ da função de utilidade indireta, nada além do que derivadas é necessário; fazer o mesmo a partir da função de utilidade direta leva a um sistema de condições de primeira ordem em muitos casos complexo. A figura a seguir ilustra a relação entre as duas abordagens do problema de escolha do consumidor.

Finalmente, um ponto interessante é que em muitos casos estaremos interessados em preferências com a propriedade que o caminho de expansão da renda seja linear em algum intervalo de renda. Pela Identidade de Roy, podemos notar que funções de utilidade indireta da forma $v(\mathbf{p}, w) = a(\mathbf{p}) + b(\mathbf{p})w$ possuem esta propriedade.

A identidade de Roy também nos permite, a partir de uma determinada forma funcional, qual seria a função de utilidade indireta consistente com isto – COLOCAR DEMONSTRAÇÃO DA P. 127 DO LIVRO DO VARIAN.

A.3.1 Excedente do consumidor

Um elemento importante na teoria microeconômica aplicada é a avaliação de como medidas de política econômica têm efeitos sobre o bem-estar do consumidor. Nessa seção, o foco é em desenvolver conceitos que, com a ajuda dos modelos de demanda anteriormente descritos, possam gerar medidas quantitativas destes efeitos. Neste ponto, a análise aqui irá seguir de perto Varian (1992) [50], capítulo 10.

Um ponto de partida para a análise poderia ser a própria função de utilidade indireta, que geralmente é postulada como forma de se chegar às equações estimáveis econometricamente. No entanto o uso das funções de utilidade indireta, ainda que correto de um ponto de vista estritamente teórico, não gera resultados diretamente interpretáveis. Isso explica-se se nos lembrarmos que a função de utilidade tem um papel estritamente ordinal, racionalizando as escolhas do consumidor. Portanto, o uso de medidas monetárias dos efeitos de bem estar se tornou o caminho mais comum na análise dos efeitos de medidas de política econômica.

Tais medidas podem ser diretamente traçadas à seguinte pergunta: suponha que o vetor de preços se altere de $\mathbf{p}^0$, para $\mathbf{p}^1$, com a renda sendo diferente em cada um destes momentos, w^0 e w^1. Qual seria a renda que, em um terceiro vetor de preços $\mathbf{q}$, que faria o consumidor ficar indiferente à cada situação? Esta pergunta pode ser respondida com a ajuda da função dispêndio e e da função de utilidade indireta v:

$$\begin{aligned} e_0 &= e(\mathbf{q}, v(\mathbf{p^0}, w^0)) \\ e_1 &= e(\mathbf{q}, v(\mathbf{p}^1, w^1)) \end{aligned}$$

Se $e_0 > e_1$, isso significa que a renda necessária para ficar indiferente a $\mathbf{q}$, no segundo momento, é menor do que no primeiro. Portanto, o indivíduo estaria em pior situação depois. A grande pergunta é qual é o papel do vetor de preços $\mathbf{q}$. Este vetor de preços é arbitrário, e o mais comum é usar uma das duas situações – 0 ou 1 – o que dá margem aos nossos conhecidos conceitos de Variação Equivalente e Variação Compensatória:

$$\begin{aligned} EV &= e(\mathbf{p}^0, v(\mathbf{p}^1, w^1)) - w^0 \\ CV &= w^1 - e(\mathbf{p^1}, v(\mathbf{p^0}, w^0)) \end{aligned}$$

Em palavras, a Variação Equivalente quer dizer a diferença entre o quanto de renda é necessário para compensar a diferença na utilidade decorrente da mudança. Neste caso, a base é composta pelos preços no período inicial. A variação compensatória, por outro lado, é o quanto de renda adicional é necessário para compensar pela variação de preços. E como estes dois resultados se relacionam? Varian (1992) [50], p. 167, mostra que a variação do excedente do consumidor necessariamente está limitada entre estes dois conceitos.

A.4 Teoria Microeconômica: Produção e Custos

Vamos começar revendo a literatura microeconômica sobre produção e custos. Vamos começar pela definição do conjunto de todos os vetores, $\mathbf{y}$, que caracterizam as possibilidades tecnológicas de uma empresa (por conveniência, os termos negativos deste vetor são os insumos). Para facilitar, iremos partir este vetor nos seus componentes, $\mathbf{q}$, que denota os produtos e $\mathbf{z}$, que denota os insumos.

Restringindo a dimensão do vetor $\mathbf{q}$ a um, podemos construir uma relação que nos dá o maior valor de $\mathbf{q}$ que pode ser alcançado dado uma configuração do vetor $\mathbf{z}$. Esta relação é denominada *função de produção*, e pode ser escrita da seguinte forma:

$$f(\mathbf{z}) = \max_{\mathbf{q}}\{q : \mathbf{z} \in V(\mathbf{y})\}$$

Em que $V(\mathbf{y})$ denota o conjunto de todas as combinações possíveis de $\mathbf{y}$. Esta função de produção possui as seguintes propriedades:

- Domínio: $f(\mathbf{z})$ é uma função definida nos conjuntos dos reais.
- Monotonicidade: Uma elevação nos insumos não pode reduzir os produtos:

$$\mathbf{z}' \geq \mathbf{z} \Rightarrow f(\mathbf{z}') \geq f(\mathbf{z})$$

- Continuidade: $f(\mathbf{z})$é contínua (upper hemicontinous à la Mas-Collell);
- Concavidade: f é quasi-côncava, o que assegura taxas marginais de substituição técnica decrescentes.
- Existência de derivada primeira e segunda (geralmente se assume isso).

No caso em que temos mais de um elemento no vetor $\mathbf{q}$, podemos utilizar a chamada *função de transformação* para descrever a produção eficiente em $\mathbf{q}$. Esta função representaria o maior valor que poderia ser alcançado por um dos elementos de $\mathbf{q}$, mantendo os outros elementos de $\mathbf{q}$ constantes, bem como os elementos de $\mathbf{z}$. Denotando os outros elementos de $\mathbf{q}$ por $\hat{\mathbf{q}}$, podemos escrever:

$$F(\hat{\mathbf{q}}, \mathbf{z}) = \max_{q_1}\{q_1 : (\mathbf{q}, \mathbf{z}) \in V(\mathbf{y})\}$$

Além das regras tecnológicas, a empresa precisa de regras de comportamento. Supondo vetores de preços de fatores de produção, denominados $\mathbf{r}$, iremos assumir do ponto de vista da empresa um comportamento minimizador de custos. Denominando $\mathbf{\Omega_n^*}$, o quadrante positivo e sendo $\mathbf{r}$ definido neste quadrante, temos que a função custo é:

$$C(\mathbf{q}, \mathbf{r}) = \min\{\mathbf{r^T z} : \mathbf{z} \in V(\mathbf{y})\}$$

Supondo as propriedades usuais para a função de produção, podemos dizer que a função custos possui as seguintes propriedades:

- Domínio: $C(\mathbf{q}, \mathbf{r})$ é uma função definida no quadrante positivo dos números reais e definido para todos os preços positivos dos fatores de produção e quantidades produzidas. Em especial, $C(\mathbf{0}, \mathbf{r}) = 0$.
- Monotonicidade: $C(\mathbf{q}, \mathbf{r})$ é sempre não decrescente na quantidade produzida e não-decrescente nos preços dos fatores. Além disso, tende ao infinito quando a quantidade produzida tende ao infinito.
- Continuidade: $C(\mathbf{q}, \mathbf{r})$ é lower hemicontinous em $\mathbf{q}$ e contínua em $\mathbf{r}$.
- Concavidade: $C(\mathbf{q}, \mathbf{r})$ é côncava em $\mathbf{r}$.

- Homogeneidade: $C(\mathbf{q}, \mathbf{r})$ é homogênea linear em $\mathbf{r}$.
- Diferenciabilidade: $C(\mathbf{q}, \mathbf{r})$ é diferenciável duas vezes em $\mathbf{r}$.

Esta última premissa implida duas propriedades importantes:

1. Lema de Shepard:
$$\frac{\partial C}{\partial r_i} = z_i$$
2. Simetria:
$$\begin{aligned} \frac{\partial C^2}{\partial r_i \partial r_j} &= \frac{\partial C^2}{\partial r_j \partial r_i} \\ \frac{\partial z_i}{\partial r_j} &= \frac{\partial z_j}{\partial r_i} \end{aligned}$$

A primeira propriedade é útil, pois pode gerar sistemas de demanda de fatores associados com a função custo, o que aumenta a eficiência das estimativas dos parâmetros. A segunda propriedade é interessante, na medida que pode permitir reduzir o número de parâmetros a serem estimados.

Apêndice B

Estimação de Sistemas de Equações

Como mencionado nos primeiros capítulos, um elemento importante para a modelagem estrutural é saber como modelar de forma conjunta equações capazes de representar partes da estrutura em que as ações dos agentes se desenrolam. Neste apêndice, iremos discutir os aspectos econométricos envolvidos na estimação dos modelos de sistemas de equações lineares e não lineares[1]. Além disso, iremos discutir como impomos restrições lineares e não lineares para os coeficientes – por exemplo, para a imposição de restrições consistentes com a racionalidade do consumidor – e, finalmente, discutiremos as formas de estimarmos funções não lineares dos coeficientes estimados (como, por exemplo, elasticidade preço e elasticidades cruzadas).

B.1 Estimação de Sistemas de Equações

A estimação de sistemas de equações de demanda, qualquer que seja a especificação, possui uma estrutura comum:

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}_1(\beta_1) + \epsilon_1 \\
\mathbf{y}_2 &= \mathbf{X}_2(\beta_2) + \epsilon_2 \\
&\vdots \\
\mathbf{y}_M &= \mathbf{X}_M(\beta_M) + \epsilon_M
\end{aligned}$$

Existem M equações, em que as variáveis em cada uma delas possuem T observações. Escolhemos por denotar a parte independente das equações de demanda por $\mathbf{X}(\beta)$ para enfatizar

[1]Esta é uma revisão sucinta, e detalhamentos maiores dos assuntos aqui expostos estão em Greene (2007)[28], em especial o capítulo 10 – Sistema de Equações de Regressão – e 13 – Modelos de Equações Simultâneas. e Davidson e MacKinnon (1992)[21].

o fato que, na maior parte das vezes, estaremos lidando com uma função não linear envolvendo os regressores $\mathbf{X}$ e os coeficientes β, e que esta função pode ser diferente – envolvendo diferentes coeficientes, por exemplo – em cada uma da equações. Evidentemente, caso o modelo seja mais simples, por exemplo linear, estas relações se simplificam bastante. Por exemplo, caso o modelo seja linear, podemos expressar este sistema na forma do modelo de regressão generalizado:

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_M} \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \tag{B.1}$$

Iremos, neste texto, enfatizar a similaridade das técnicas de estimação e a sua base comum no Método Generalizado dos Momentos. Para começarmos, vamos introduzir uma notação adicional. Seja $\mathbf{y_\bullet}$ o vetor de dimensão $MT \times 1$ obtido a partir da "vetorização" dos pequenos vetores $\mathbf{y}_i$. Da mesma forma, iremos denotar $\mathbf{X_\bullet}(\beta)$ o vetor obtido a partir do "empilhamento" dos lados direitos da igualdade, e $\epsilon_\bullet$ o vetor obtido a partir do "empilhamento" dos resíduos de cada uma das regressões. Uma ilustração útil desta formulação pode ser obtida a partir da equação (B.1) acima, da seguinte forma:

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_M} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1}\beta_\mathbf{1} \\ \mathbf{X_2}\beta_\mathbf{2} \\ \vdots \\ \mathbf{X_M}\beta_\mathbf{M} \end{bmatrix} + \begin{bmatrix} \varepsilon_\mathbf{1} \\ \varepsilon_\mathbf{2} \\ \vdots \\ \varepsilon_\mathbf{M} \end{bmatrix}$$

Ou:

$$\mathbf{y_\bullet} = \mathbf{X_\bullet}(\beta) + \epsilon_\bullet$$

A partir desta estrutura básica, vamos começar com o método mais óbvio de estimação – o Método dos Mínimos Quadrados Ordinários (OLS).

B.1.1 OLS

Neste caso, podemos escrever a i-ésima equação do sistema da seguinte forma:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i(\beta) + \epsilon_i \\ E(\epsilon_i \epsilon_i^T) &= \sigma_{ii}\mathbf{I}_T \end{aligned}$$

Em que $\mathbf{I}_T$ denota a matriz identidade de dimensão $T \times T$. Além disso, precisamos da seguinte condição:

$$E(\epsilon_i \epsilon_j^T) = \mathbf{0}, \forall j \neq i$$

Em que $\mathbf{0}$ denota a matriz zero de dimensão $T \times T$ – Ou seja, não há correlação entre os erros da equação i e os da equação j. Neste caso, podemos construir a matriz variância-covariância do sistema empilhado da seguinte forma:

$$E(\epsilon_\bullet \epsilon_\bullet^T) = \mathbf{V} = \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{22}\mathbf{I_T} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{MM}\mathbf{I}_T \end{bmatrix}$$

Neste caso, podemos estimar os coeficientes do sistema se estimarmos equação por equação por OLS, caso não tivéssemos restrições dos coeficientes entre as equações. A estimação se daria por meio da minimização da função critério – que no caso seria a soma dos quadrados dos resíduos, ponderada pela inversa da matriz de variância-covariância dos resíduos:

$$SSR_\bullet = (\mathbf{y}_\bullet - \mathbf{X}_\bullet(\beta))^T(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})(\mathbf{y}_\bullet - \mathbf{X}_\bullet(\beta))$$

Em que $\mathbf{\Sigma}$ representa a matriz de variância-covariância dos resíduos, que no caso em questão possuem a seguinte forma:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} = \mathbf{0} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} = \mathbf{0} & \cdots & \sigma_{MM} \end{bmatrix}$$

No caso de OLS linear nos coeficientes, em uma forma parecida com a (B.1), temos que a minimização da soma dos quadrados dos resíduos dá uma forma analítica fechada, que seria o seguinte:

$$\hat{\beta} = (\mathbf{X^T}(\mathbf{\Sigma^{-1}} \otimes \mathbf{I_T})\mathbf{X})^{\mathbf{-1}}\mathbf{X^T}(\mathbf{\Sigma^{-1}} \otimes \mathbf{I_T})\mathbf{y}$$

Geralmente são estimados oa membros desta matriz com os resíduos da regressão, ou seja - $\hat{\sigma}_{11} = (\hat{\epsilon}_\mathbf{1}^\mathbf{T}\hat{\epsilon}_1)$/Com isto, temos como estimativa de variância de β:

$$Va\hat{r}(\beta) = (\mathbf{Xd^T}_\bullet(\beta)(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Xd}_\bullet(\beta))^{-1}$$

Em que $\mathbf{Xd}_\bullet(\beta)$ é uma matriz de derivadas parciais de $\mathbf{X}_\bullet(\beta)$ com relação a cada um dos coeficientes, e $\otimes$ denota o Produto Kronecker. No caso de modelos lineares, esta função simplifica para:

$$Va\hat{r}(\beta) = (\mathbf{X^T}(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X})^{-1}$$

Vamos agora relaxar a premissa de independência dos erros das equações, no Método de Regressões Aparentemente Não Correlacionadas de Zellner (1962)[52].

B.1.2 SUR

A diferença entre este método e o anterior reside na estrutura de correlação entre os erros das equações. Neste caso, temos que a matriz $\mathbf{\Sigma}$ tem a seguinte estrutura:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix}$$

O que nos dá a seguinte matriz de variância-covariância dos resíduos empilhados:

$$E(\epsilon_\bullet \epsilon_\bullet^T) = \mathbf{V} = \mathbf{\Sigma} \otimes \mathbf{I}_T$$

Temos, neste caso, a mesma função critério a ser minimizada:

$$SSR_\bullet = (\mathbf{y}_\bullet - \mathbf{X}_\bullet(\beta))^T (\mathbf{\Sigma}^{-1} \otimes \mathbf{I})(\mathbf{y}_\bullet - \mathbf{X}_\bullet(\beta))$$

A única diferença entre este e o da semana anterior é a existência de correlação entre os resíduos de equações. A idéia desta técnica é que, mesmo que tenhamos as variáveis independentes e as variáveis dependentes sendo completamente diferentes entre as equações, a correlação entre os resíduos da regressão pode ser diferente, de forma que estimar isto por Mínimos Quadrados Ordinários não dá certo. A matriz de variância-covariância dos cieficientes neste caso é similar também:

$$\hat{Var}(\beta) = (\mathbf{Xd^T}_\bullet(\beta)(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Xd}_\bullet(\beta))^{-1}$$

Algumas observações importantes para o caso do SUR. Quando os regressores são os mesmos em todas as equações, e não há restrições entre os coeficientes de diferentes equações, SUR é igual à aplicação de OLS equação por equação. Neste caso, e considerando a formulação da equação (B.1) acima, temos:

$$\hat{\beta} = \begin{bmatrix} \sigma_{11}(X^TX)^{-1} & \sigma_{12}(X^TX)^{-1} & \cdots & \sigma_{1M}(X^TX)^{-1} \\ \sigma_{21}(X^TX)^{-1} & \sigma_{22}(X^TX)^{-1} & \cdots & \sigma_{2M}(X^TX)^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}(X^TX)^{-1} & \sigma_{M2}(X^TX)^{-1} & \cdots & \sigma_{MM}(X^TX)^{-1} \end{bmatrix} \begin{bmatrix} (X^TX)\sum_{l=1}^M \sigma^{1l} b_l \\ (X^TX)\sum_{l=1}^M \sigma^{2l} b_l \\ \vdots \\ (X^TX)\sum_{l=1}^M \sigma^{Ml} b_l \end{bmatrix}$$

Sendo que $\mathbf{b_l} = (\mathbf{X^TX})^{-1}\mathbf{X^Ty_l}$. Em que $\sigma^{\mathbf{jj}}$ representa o bloco jj da matriz $\mathbf{\Sigma^{-1}}$, enquanto que $\sigma_{\mathbf{jj}}$ representa o mesmo bloco, só que da matriz $\mathbf{\Sigma}$. Neste caso, qualquer linha deste produto matricial – por exemplo, $\mathbf{k}$, pode ser escrita como:

$$\hat{\beta}_{\mathbf{K}} = \mathbf{b_1}\left(\sum_{\mathbf{j=1}}^{\mathbf{M}} \sigma_{\mathbf{kj}}\sigma^{\mathbf{j1}}\right) + \mathbf{b_2}\left(\sum_{\mathbf{j=1}}^{\mathbf{M}} \sigma_{\mathbf{kj}}\sigma^{\mathbf{j2}}\right) + \cdots \mathbf{b_M}\left(\sum_{\mathbf{j=1}}^{\mathbf{M}} \sigma_{\mathbf{kj}}\sigma^{\mathbf{jM}}\right)$$

Uma vez que os termos com subscrito são da matriz $\mathbf{\Sigma}$ e os termos com sobrescrito são elementos da matriz $\mathbf{\Sigma^{-1}}$, o produto destas coisas vão ser tal que $\sigma_{\mathbf{ij}}\sigma^{\mathbf{ji}}$ vai ser zero todas as vezes que $\mathbf{j} \neq \mathbf{i}$, e um, caso contrário. Este monte de somatórios se elimina de forma que $\hat{\beta}_{\mathbf{k}} = \mathbf{b_k}$, ou seja, é igual a OLS equação por equação.

Agora já discutimos a existência de correlação contemporânea entre os resíduos das regressões, vamos discutir os casos em que há autocorrelação e heterocedasticidade nas variâncias de cada uma das equações. Iremos ver o Método dos Mínimos Quadrados Generalizados Factíveis.

B.1.3 Mínimos Quadrados Generalizados Factíveis

Neste caso, a estrutura é bastante similar, o que mudaremos é a forma pela qual a nossa já queria matriz $\mathbf{\Sigma}$ é calculada. Neste caso, teríamos uma situação em que cada um dos elementos da matriz $\mathbf{\Sigma}$ seria calculada por um método como o de Newey-West, em que σ_{ii} seria uma soma ponderada das variâncias, das covariâncias entre os resíduos entre t e $t-1$, t e $t-2$, e assim por diante.

Passemos agora aos sistemas de equações em que temos a existência de um problema potencial de endogeneidade. Comecemos com o Método dos Mínimos Quadrados a Dois Estágios (TSLS).

B.1.4 Mínimos Quadrados a Dois Estágios

Neste caso, temos uma diferença em relação aos casos anteriores, em que temos entre os regressores uma – ou mais de uma – variável que se considera como sendo determinada conjuntamente com a variável dependente. No entanto, neste caso ainda não nos preocupamos com a estrutura de covariânia entre os resíduos das diferentes equações. Para entender isso melhor, iremos seguir a abordagem de Davidson e Mackinnon (1993 [21], cap. 18), e representar o sistema – linear nos coeficientes – como sendo:

$$\mathbf{Y\Gamma} = \mathbf{XB} + \mathbf{U}$$

Sendo que $\mathbf{Y}$ representa uma matriz $T \times M$ de variáveis endógenas, $\mathbf{X}$ uma matriz $T \times k$ de variáveis exógenas, $\mathbf{B}$ uma matriz $k \times M$ de coeficientes, $\mathbf{\Gamma}$ uma matriz $M \times M$ de coeficientes e $\mathbf{U}$ uma matriz $T \times M$ de termos erro. Um problema é que, se deixarmos $\mathbf{\Gamma}$ sem nenhuma restrição, teremos um problema sério de identificação (imaginem se em todas as equações, além de estimarmos os coeficientes do lado direito da igualdade, ainda tivéssemos que estimar um coeficiente para a variável dependente – os coeficientes obtidos não seriam invariantes à multiplicação por uma constante). Portanto, uma normalização importante é que os elementos $\mathbf{\Gamma_{ii}}$ – ou seja, os coeficientes da variável dependente da i-ésima equação na i-ésima equação – sejam iguais a um (esta normalização é arbitrária – um monte de livros de econometria de graduação, quando apresentam este assunto, supõem que a mesma variável endógena tem o coeficiente de um em todas as equações). Para que este modelo possa ser estimado, ou seja, possamos obter $\mathbf{Y}$ de forma única em função das variáveis $\mathbf{X}$, precisamos inverter $\mathbf{\Gamma}$ e pós-multiplicá-la pelo outro lado da igualdade, assim:

$$\mathbf{Y} = \mathbf{XB\Gamma^{-1}} + \mathbf{U\Gamma^{-1}} \tag{B.2}$$

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V} \tag{B.3}$$

Ainda, seguindo a abordagem de Davidson e Mackinnon [21], a primeira equação matricial, (B.2) é a forma reduzida restrita e a segunda é a forma reduzida irrestrita. A restrição decorre do fato que $\mathbf{\Pi} = \mathbf{B\Gamma^{-1}}$. Nosso interesse, evidentemente, reside em obter estimativas para $\mathbf{B}$ e $\mathbf{\Gamma}$. E aqui temos que tomar cuidado. Mesmo com a restrição sobre os elementos da matriz $\mathbf{\Gamma}$, necessária para que consigamos identificação, ela possui $M^2 - M$ elementos a serem estimados. Além disso, a matriz $\mathbf{B}$ possui Mk elementos. Por outro lado, a matriz $\mathbf{\Pi}$ possui Mk elementos apenas, bem menos que os $M^2 - M + Mk$ do pacote $\mathbf{B\Gamma^{-1}}$, o que significa que teremos que impor pelo menos $M^2 - M$ restrições em $\mathbf{B}$ ou em $\mathbf{\Gamma}$ para que possamos estimá-los.

Para entendermos melhor esta questão de identificação, vamos olhar o sistema equação por equação. Como vimos, precisamos de restrições nos coeficientes de $\mathbf{B}$ e $\mathbf{\Gamma}$ para conseguir isso. Usualmente, estas restrições são as chamadas restrições de exclusão - ou seja, alguns coeficientes são iguais a zero, ou seja, a variável não pertence àquela equação.

Vamos pegar o vetor $\mathbf{Y}$ e particioná-lo da seguinte forma:

$$\mathbf{Y} = \begin{bmatrix} y & Y^1 & Y^2 \end{bmatrix}$$

Em que $\mathbf{y}$ é a variável dependente da primeira equação, $\mathbf{Y^1}$as variáveis endógenas do sistema não excluídas da equação por restrições de exclusão, $\mathbf{Y^2}$ as variáveis exógenas do sistema excluídas daquela equação. Analogamente, podemos expressar a matriz de variáveis exógenas do sistema como sendo:

$$\mathbf{X} = \begin{bmatrix} X^1 & X^2 \end{bmatrix}$$

Com isso, podemos reescrever o sistema como sendo:

$$\begin{bmatrix} \mathbf{y} & \mathbf{Y^1} & \mathbf{Y^2} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{\Gamma_{02}} \\ -\gamma_1 & \mathbf{\Gamma_{12}} \\ \mathbf{0} & \mathbf{\Gamma_{22}} \end{bmatrix} = \begin{bmatrix} X^1 & X^2 \end{bmatrix} \begin{bmatrix} \beta_1 & \mathbf{B_{12}} \\ \mathbf{0} & \mathbf{B_{22}} \end{bmatrix} + \mathbf{U}$$

A primeira equação fica sendo, então:

$$\mathbf{y} = \mathbf{Y^1}\gamma_\mathbf{1} + \mathbf{X^1}\beta_\mathbf{1} + \mathbf{u^1} = \mathbf{y} = \mathbf{Z}\delta + \mathbf{u^1}$$

Para que possamos obter estimativas por Mínimos Quadrados a Dois Estágios, precisamos de instrumentos para $\mathbf{Y^1}$. As variáveis coletadas em $\mathbf{X^1}$, sendo exógenas, são instrumentos para elas mas as variáveis em $\mathbf{Y}^1$ precisam de instrumentos. Para que seja possível a identificação, portanto, o número de variáveis existente em $\mathbf{X^2}$tem que ser maior ou igual ao número de variáveis em $\mathbf{Y^1}$– a chamada *condição de ordem* para a identificação.

Passemos ao método dos Mínimos Quadrados a Dois Estágios. A idéia aqui é usar, como instrumentos para $\mathbf{Y^1}$, todas as variáveis constantes na matriz $\mathbf{X}$:

$$\mathbf{\hat{Y}^1} = \mathbf{X}[(\mathbf{X^T X})^{-\mathbf{1}}\mathbf{X^T Y^1}]$$

Com isso, o estimador de MQ2E para δ fica sendo:

$$\hat{\delta}_{1MQ2E} = \begin{bmatrix} (\mathbf{\hat{Y}^1})^{\mathbf{T}}\mathbf{\hat{Y}^1} & (\mathbf{\hat{Y}^1})^{\mathbf{T}}\mathbf{X^1} \\ (\mathbf{X^1})^{\mathbf{T}}\mathbf{\hat{Y}^1} & (\mathbf{X^1})^{\mathbf{T}}\mathbf{X^1} \end{bmatrix}^{-1} \begin{bmatrix} (\mathbf{\hat{Y}^1})^{\mathbf{T}}\mathbf{y} \\ (\mathbf{X^1})^{\mathbf{T}}\mathbf{y} \end{bmatrix}$$

Podemos também escrever de uma forma mais simplificada, útil quando utilizamos programas como o MATLAB para estimar:

$$\begin{aligned} \hat{\delta}_{\mathbf{1_{MQ2E}}} &= [(\mathbf{Z^{1T}X})(\mathbf{X^T X})^{-\mathbf{1}}(\mathbf{X^T Z^1})]^{-\mathbf{1}}(\mathbf{Z^{1T}X})(\mathbf{X^T X})^{-\mathbf{1}}(\mathbf{X^T Z^1})\mathbf{y} \\ \mathbf{\hat{Z}^1} &= \mathbf{X}(\mathbf{X^T X})^{-\mathbf{1}}(\mathbf{X^T Z^1}) \end{aligned}$$

A variância assintótica dos coeficientes é dada por:

$$Var(\delta_{1_{MQ2E}}) = \sigma_{11}[(\mathbf{Z^{1T}X})(\mathbf{X^T X})^{-\mathbf{1}}(\mathbf{X^T Z^1})]^{-\mathbf{1}}$$

Aqui podemos trabalhar com a segunda das condições necessárias para a identificação, a chamada <u>condição de posto</u>. Em especial, esta condição de posto pode ser entendida como que a matriz $[(\mathbf{Z^{1T}X})(\mathbf{X^T X})^{-\mathbf{1}}(\mathbf{X^T Z^1})]^{-\mathbf{1}}$ não seja singular. Vamos começar supondo um modelo linear, em que a i - ésima equação possui a seguinte forma:

$$\mathbf{y}_i = \mathbf{z}_i\delta + \epsilon_i$$

Em que $\mathbf{z}_i = [\mathbf{Y_i}|\mathbf{x_i}]$. Ou seja, juntamos as variáveis que são consideradas como endógenas e as que são consideradas como exógenas. Supondo que $\mathbf{x}_i$ inclua todas as variáveis exógenas e que estas variáveis exógenas sejam suficientes para o atendimento das condições de posto e condições de ordem do sistema de equações, temos que podemos obter a estimativa dos coeficientes do sistema por meio das chamadas "condições de ortogonalidade":

$$\mathbf{m}(\delta) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\mathbf{y}_i - \mathbf{z}_i\delta)$$

A idéia é escolher os coeficientes de forma a minimizar este valor para torná-lo o mais próximo de zero possível. Supondo que tenhamos mais instrumentos do que variáveis endógenas no sistema

– ou seja, ele é sobre-identificado – este sistema nunca dá exatamente zero. Para resolver isso, geralmente se escolhe o vetor de coeficientes que minimiza a seguinte função objetivo:

$$q = \mathbf{m}(\delta)^T W^{-1} \mathbf{m}(\delta)$$

Na verdade isto aqui é uma forma quadrática, com uma ponderação dada pela variância dos resíduos. Se os resíduos da equação são homoscedásticos e não autocorrelacionados, a matriz $W = (\mathbf{X}^T\mathbf{X})$ é a mais adequada. Passemos a seguir para os métodos que consideram a estrutura de correlação entre as equações do sistema.

B.1.5 Mínimos Quadrados a Três Estágios

Vamos agora começar a discutir os métodos de estimação que consideram a informação de todos os resíduos do sistema. O primeiro destes modelos é o método dos Mínimos Quadrados a Três Estágios, em que supomos que a estrtura de covariância entre os resíduos das equações é a seguinte:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix}$$

O que nos dá a seguinte matriz de variância-covariância dos resíduos empilhados:

$$E(\epsilon_{\bullet}\epsilon_{\bullet}^T) = \mathbf{V} = \mathbf{\Sigma} \otimes \mathbf{I}_T$$

Com esta estrutura de resíduos, podemos calcular os estimadores de Mínimos Quadrados a Três Estágios da seguinte forma:

$$\beta_{3SLS} = (\hat{\mathbf{Z}}(\mathbf{\Sigma^{-1}} \otimes \mathbf{I})\mathbf{Z})^{-\mathbf{1}}\hat{\mathbf{Z}}(\mathbf{\Sigma^{-1}} \otimes \mathbf{I})\mathbf{Zy}$$

Neste caso, a matriz $\mathbf{\Sigma}$ é calculada com os resíduos de Mínimos Quadrados a Dois Estágios calculados equação por equação. Na verdade esta é uma aplicação do caso mais geral do Método Generalizado dos Momentos com equações não-lineares, que é o que veremos mais adiante.

Finalmente, uma observação adicional, de acordo com Davidson e Mackinnon (1993, 517): quando as equações do sistema são exatamente identificadas, as estimativas dos parâmetros são exatamente iguais às dos Mínimos Quadrados a Dois Estágios.

B.1.6 GMM em Sistema

Finalmente, passamos ao método mais amplo de todos, o Método Generalizado dos Momentos para o sistema de equações. a idéia neste caso começa com o seguinte conjunto de condições de ortogonalidade para cada uma das equações do sistema:

$$\mathbf{m}_i(\delta) = \mathbf{X}^T(\mathbf{Y}_i - \mathbf{Z}_i(\beta))$$

Em que $\mathbf{X}$ denota o conjunto de instrumentos utilizado. A partir destas condições de ortogonalidade, podemos definir a seguinte função a ser minimizada:

$$q = \sum_{i=1}^{M}\sum_{j=1}^{M}((\mathbf{Y_i} - \mathbf{Z_i}(\beta))^T\mathbf{X})(\mathbf{W})^{ij}(\mathbf{X}^T(\mathbf{Y_i} - \mathbf{Z_i}(\beta)))$$

Nesta equação, o único termo ainda não visto é o $(\mathbf{W})^{ij}$, que seria a inversa da covariância entre os momentos da equação i e os momentos da equação j. Podemos expressar esta função da seguinte forma:

$$\mathbf{W} = E(\mathbf{m}(\delta)\mathbf{m}(\delta)^T)$$

Ou ainda, podemos abrir o sistema da seguinte forma:

$$q = \begin{bmatrix} (\mathbf{y}_1 - \mathbf{Z}_1(\beta))^T\mathbf{X} \\ \vdots \\ (\mathbf{y}_M - \mathbf{Z}_M(\beta))^T\mathbf{X} \end{bmatrix}^T \begin{bmatrix} \sigma_{11}(\mathbf{X}^T\mathbf{X}) & \cdots & \sigma_{1M}(\mathbf{X}^T\mathbf{X}) \\ \vdots & \ddots & \vdots \\ \sigma_{M1}(\mathbf{X}^T\mathbf{X}) & \cdots & \sigma_{MM}(\mathbf{X}^T\mathbf{X}) \end{bmatrix} \begin{bmatrix} \mathbf{X}^T(\mathbf{y}_1 - \mathbf{Z}_1(\beta)) \\ \vdots \\ \mathbf{X}^T(\mathbf{y}_M - \mathbf{Z}_M(\beta)) \end{bmatrix}$$

Estamos supondo aqui que os resíduos em cada uma das equações são homocedásticos e não autocorrelacionados. Em geral, ele é mais vantajoso do que o Método dos Mínimos Quadrados a Três Estágios porque tem ganhos de eficiência na presença de correlação e heterocedasticidade (a matriz do meio pode ser ajeitada com o uso do estimador de Newey-West para dar estimativas robustas). A matriz de variância-covariância das estimativas é dada pela matriz na página 758 do Greene.

Passemos agora à imposição de restrições não lineares e a realização de testes de hipóteses não lineares para as estimativas.

B.2 Teste de hipóteses e o Método Delta

Vimos na aula passada que boa parte dos sistemas de demanda são funções não lineares das variáveis independentes; além disso, as elasticidades-preço da demanda são funções não lineares dos parâmetros. Iremos agora detalhar como podemos usar as nossas estimativas, conjuntamente com as estimativas dos erros-padrão dos coeficientes, para obter intervalos de confiança para elasticidades-preço da demanda. Vamos imaginar que tenhamos um vetor de estimativas de parâmetros, denominado $\hat{\beta}$, e gostaríamos de obter um intervalo de confiança para uma função – possivelmente não linear – de β, que denominaremos $\gamma = g(\beta)$. O valor estimado de γ seria

$$\hat{\gamma} = g(\hat{\beta})$$

Fazendo uma expansão de Taylor em volta do valor "verdadeiro" de β, que denominaremos de β_0, temos:

$$\hat{\gamma} \simeq g(\beta_0) + \bigtriangledown_\beta g(\beta_0)(\hat{\beta} - \beta_0)$$

Uma vez que $g(\beta_0) = \gamma_0$, podemos reorganizar a equação acima:

$$\hat{\gamma} - \gamma_0 = \bigtriangledown_\beta g(\beta_0)(\hat{\beta} - \beta_0)$$

Multiplicando os dois lados por $N^{1/2}$, para assegurar uma velocidade de convergência assintótica adequada, temos que a variância de $\hat{\gamma} - \gamma_0$ é igual à seguinte forma quadrática:

$$(\bigtriangledown_\beta g(\beta_0))V(\beta)(\bigtriangledown_\beta g(\beta_0))^T$$

Esta forma quadrática nos dá a seguinte expressão para a função não linear dos parâmetros:

$$Var(\hat{\gamma}) = (\bigtriangledown_\beta g(\hat{\beta}))V(\beta)(\bigtriangledown_\beta g(\hat{\beta}))^T$$

B.2.1 Condições de Identificação

Como estamos falando de modelos em que existe um potencial problema de endogeneidade quando estamos falando de variáveis agregadas, é necessário que discutamos com maior profundidade o que a literatura arrola como fonte de instrumentos para assegurar identificação dos parâmetros em uma situação como esta.

O princípio é o mesmo: encontrar uma variável que seja correlacionada com a variável potencialmente problemática, para aí utilizá-la, juntamente com as outras variáveis explicativas, como instrumento para a variável problemática. Vamos detalhar alguns possíveis instrumentos:

- Preços dos Insumos: Geralmente são bons instrumentos, pois estão relacionados com o lado da oferta e não com os termos erro da regressão do lado da demanda; geralmente são chamados de "cost shifters". O problema é que o galho da endogeneidade envolve todos os preços dos produtos, o que indica que encontrar preços de insumos para todos os preços que estão nas equações de demanda fica sendo algo um pouco complicado.

- Preços dos mesmos produtos em outras áreas geográficas: este esquema de identificação foi adotado por Hausman et al. (1994) e Hausman (1997): a idéia deste esquema de identificação é a seguinte; supondo que os choques de custos afetem os produtos vendidos em todas as cidades, enquanto os choques de demanda são específicos á cada uma das cidades, podemos dizer que os preços dos produtos observados em outras cidades são instrumentos válidos para o preço do produto na cidade. O problema é que a eficácia deste procedimento é duvidosa. Bresnahan (1997) questiona a validade desta abordagem no caso dos cereais matinais, uma vez que existem campanhas de marketing construídas nacionalmente, o que faz que tenhamos choques de demanda gerais a todas as cidades.

- Preços dos Outros Produtos produzidos pela mesma empresa: Neste caso, a idéia de identificação reside no fato que os choques de custos são comuns às empresas.

- Preços de todos os outros produtos.

Finalmente, uma abordagem alternativa também discute se não é válido que não nos preocupemos com os problemas de endogeneidade, em especial se tivermos dados desagregados finamente o suficiente. Hausman (1997) afirma que, se os supermercados não ajustam os seus preços semanalmente para equilibrar oferta e demanda, e se as curvas de oferta são horizontais, os preços podem ser considerados como econometricamente predeterminados, e neste caso não existiria o feedback entre os shares e os preços. Mesmo assim, existem dúvidas sobre isso. Bresnahan (1997) afirma que pra isso ser verdade, os supermercados não poderiam ajustar os seus preços em resposta aos choques nacionais de demanda, uma hipótese pouco provável.

Apêndice C

Métodos de Integração

Apêndice D

Teste para Modelos não-aninhados

Uma outra crítica foi levantada por Nevo (1998) à abordagem de Breshanan (1982) e Gollop e Roberts (1979). Neste caso, a demanda pelo produto de uma empresa i é uma função não apenas do preço de mercado – que, dadas as premissas do modelo, seria o mesmo – mas também dos preços dos produtos que são substitutos imperfeitos dele. Neste caso, a demanda pelo produto da empresa i em um mercado com J empresas pode ser escrita da seguinte forma:

$$Q_i = D(p_1, \cdots, p_J, Y_i, \alpha), i = 1, \cdots, J$$

Em que α representa um vetor de parâmetros, $p_1, \cdots, p_J$ um vetor de preços dos produtos. Neste caso, os lucros da empresa podem ser escritos como

$$\Pi_i = \sum_{i \in I} (p_i - CMg_i(W, \beta)) \times Q_i - F$$

Neste caso, W e β representam variáveis exógenas do lado dos custos e os coeficientes da função custos. Além disso, F representa o total dos custos fixos. Supondo que as empresas compitam em preços à la Bertrand, temos que os preços de qualquer produto da empresa i deve atender às seguintes condições de primeira ordem:

$$Q_i(p) + \sum_{i \in I} (p_i - CMg_i) \times \frac{\partial Q_i}{\partial p_i} = 0$$

Para o caso de uma empresa que tem um conjunto – denotado I – de produtos, isto denota um sistema de I equações. Podemos resolver este sistema da seguinte forma. Vamos definir $S_{jr} = -\frac{\partial Q_r}{\partial p_j}, j, r = 1, \cdots, J$, que é o negativo do vetor de sensibilidades das demandas aos preços. Além disso, vamos definir o que seria uma matriz seletora de propriedade (ou seja, marca os produtos que são de uma mesma empresa), da seguinte forma:

$$\Theta_{jr} = \begin{cases} 1, & \exists f : \{r, j\} \in I \\ 0 & c.c. \end{cases}$$

Adicionalmente, podemos definir $\Omega_{jr} = \Theta_{jr} S_{jr}$. Desta forma, podemos representar em notação vetorial este sistema de condições de primeira ordem da seguinte forma:

$$Q(p) - \Omega(p - CMg) = 0$$

O que implica a seguinte relação de oferta:

$$p = CMg + \Omega^{-1} Q(p)$$

Nevo (1998) afirma que a modelagem de parâmetros de conduta como a que vimos equivaleria a substituir a matriz Θ por uma matriz de parâmetros a serem estimados, como no caso de Gollop e Roberts. O problema que o autor levanta é que, no caso de produtos diferenciados, é extremamente difícil identificar os parâmetros constantes nesta matriz Θ. Para isto, ele faz um exemplo, supondo duas empresas. A demanda de uma delas, denominada j é dada por:

$$Q_j = \alpha_{j0} + \alpha_{j1} p_1 + \alpha_{j2} p_2 + Y' \alpha_{j3} + \epsilon_j, j = 1, 2$$

Além disso, a função custo marginal é dada por:

$$CMg = \beta_{j0} + W' \beta_{j1}, j = 1, 2$$

Fazendo uma análise similar à de Breshanan (1982), podemos cozinhar as duas equações de forma que:

$$\begin{aligned} p_j &= \beta_{j0} + W' \beta_{j1} - A \left(\lambda_{ii} \frac{\partial Q_i}{\partial p_i} Q_i - \lambda_{ij} \frac{\partial Q_j}{\partial p_i} Q_i \right) + \eta \\ &= \beta_{j0} + W' \beta_{j1} - A(\lambda_{ii} \alpha_{ii} Q_j - \lambda_{ij} \alpha_{ji} Q_i) + \eta, j = 1, 2, i = 3 - j \end{aligned}$$

Em que:

$$A = \left(\lambda_{11} \lambda_{22} \frac{\partial Q_2}{\partial p_2} \frac{\partial Q_1}{\partial p_1} - \lambda_{12} \lambda_{21} \frac{\partial Q_2}{\partial p_1} \frac{\partial Q_1}{\partial p_2} \right)^{-1} = (\lambda_{11} \lambda_{22} \alpha_{22} \alpha_{11} - \lambda_{12} \lambda_{21} \alpha_{21} \alpha_{12})^{-1}$$

O problema é que, para identificar os λ's e os α's teríamos que ter a dimensão de variáveis exógenas nas equações de demanda igual ao número de produtos. Quando temos dois produtos, encontrar dois instrumentos é fácil. No entanto, quando temos um monte de produtos a coisa fica mais complicada. Como Nevo (1998, p. 394) afirma, "na prática, no entando, é difícil, se não impossível, encontrar tal enorme número de variáveis exógenas que influenciam a demanda mas são não correlacionadas com o choque na regra de oferta".

Para resolver este problema, os autores sugerem que, para a identificação da conduta do setor, sejam estimados diferentes modelos - correspondendo a diferentes estruturas para a variável Θ- e depois fazendo um teste estatístico formal. Evidentemente, os testes envolvidos são diferentes dos tradicionais, por uma simples razão: eles envolvem as chamadas "hipóteses não-aninhadas", ou seja, um dos possíveis resultados não pode ser obtido a partir de outra especificação, simplesmente colocando-se alguns parâmetros iguais a zero. Um dos testes mais comumente utilizados para a investigação de diferentes condutas em um contexto de Máxima Verossimilhança é o teste de Vuong (1989).

D.0.2 O Teste de Vuong (1989)

Vamos detalhar um pouco mais o princípio do Teste de Vuong (1989). A idéia básica subjacente é que um dos dois modelos, mesmo que não-aninhados, pode ser mais próximo de um limite que seria dado pela distribuição conjunta das observações e dos parâmetros dado o modelo "verdadeiro". Em especial, este conceito pode ser operacionalizado pelo Critério de Informação de Kullback-Leibler KLIC, que é uma medida da discrepância entre dois modelos. Supondo que a densidade condicional da variável dependente em relação às variáveis independentes segundo a distribuição conjunta "verdadeira" seja dada por $h_0(Y_i|X_i)$[1], e a densidade condicional da variável dependente na distribuição que escolhemos é dada por $f(Y_i|X_i, \beta^*)$, podemos expressar o KLIC da seguinte forma:

$$KLIC = E_0[\ln h_0(Y_i|X_i)] - E_0[\ln f(Y_i|X_i, \beta^*)]$$

Sendo que esta esperança é calculada tomando-se em consideração o modelo verdadeiro, e é denotada por E_0. Definimos adicionalmente β^* como os valores dos coeficientes da distribuição quando f não é o modelo verdadeiro. Neste caso, o modelo "melhor" seria aquele que possui o valor menor para este KLIC. Como não sabemos exatamente como é a função h_0, temos que a comparação entre os diferentes modelos envolverá a comparação das log-verossimilhanças entre os diferentes modelos. Esta é a essência do teste de Vuong (1989).

Para facilitar as idéias, vamos considerar dois modelos competidores, $M_f = \{F_\beta, \beta \in B\}$ e $M_g = \{F_\gamma, \gamma \in \Gamma\}$, em que F e G são funções distribuição que caracterizam os D.G.P. dos dois modelos, o que implicam f.d.p., f e g, respectivamente. Como estes modelos são não-aninhados, não podemos construir nenhuma densidade condicional $(Y_i|X_i)$, que seja elemento de M_f e M_g. A hipótese nula deste teste é que eles estão igualmente distantes do modelo "verdadeiro", o que

[1]Evidentemente, esta distribuição é desconhecida. Senão, a usaríamos e tudo estaria resolvido.

implica:

$$H_0 : E_0 \left[\ln \frac{f(Y_i|X_i,\beta^*)}{g(Y_i|Z_i,\gamma^*)} \right] = 0$$

Isto implica duas hipóteses alternativas:

$$H_f : E_0 \left[\ln \frac{f(Y_i|X_i,\beta^*)}{g(Y_i|Z_i,\gamma^*)} \right] > 0$$

Que implicaria que o modelo f é preferido ao modelo g, e a hipótese contrária:

$$H_g : E_0 \left[\ln \frac{f(Y_i|X_i,\beta^*)}{g(Y_i|Z_i,\gamma^*)} \right] < 0$$

Note que, como estamos falando de E_0 ainda, esta esperança não pode ser calculada. O truque do Vuong é que, sob condições de regularidade, podemos realizar a seguinte aproximação:

$$\begin{aligned} \frac{1}{N} LR_N(\hat{\beta},\hat{\gamma}) &\to^{a.s} E_0 \left[\ln \frac{f(Y_i|X_i,\beta^*)}{g(Y_i|Z_i,\gamma^*)} \right] \\ LR_N(\hat{\beta},\hat{\gamma}) &= \ln f(Y_i|X_i,\hat{\beta}) - \ln g(Y_i|X_i,\hat{\gamma}) \end{aligned}$$

Ou seja, é a nossa conhecida razão de verossimilhança dividida pelo número de observações. Mas ainda não temos o teste completo. Sob a hipótese nula, temos o seguinte resultado de convergência assintótica:

$$\frac{LR_N(\hat{\beta},\hat{\gamma})}{\sqrt{N}\hat{\omega}_N} \to^d N(0,1)$$

Sendo que:

$$\hat{\omega}_N^2 = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\ln f(Y_i|X_i)}{\ln g(Y_i|X_i)} \right]^2 - \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\ln f(Y_i|X_i)}{\ln g(Y_i|X_i)} \right]^2$$

O que poderíamos dizer é que esta estatística é uma média da razão de verossimilhança, adequadamente normalizada. O mais chato neste teste é conseguir calcular a normalização, que depende das contribuições das observações individuais à função verossimilhança. Considerando as hipóteses alternativas mencionadas anteriormente, podemos dizer que o modelo f seria mais próximo do modelo verdadeiro se a estatística de teste fosse positiva e diferente de zero; por outro lado, poderíamos dizer que o modelo g seria mais próximo do modelo verdadeiro se a estatística de teste fosse negativa e diferente de zero. Neste sentido, Vuong (1989) mostra que, sob H_f, temos:

$$\frac{LR_N(\hat{\beta},\hat{\gamma})}{\sqrt{N}\hat{\omega}_N} \to^{a.s} +\infty$$

E, sob H_g, teríamos:

$$\frac{LR_N(\hat{\beta},\hat{\gamma})}{\sqrt{N}\hat{\omega}_N} \rightarrow^{a.s.} -\infty$$

Caso o número de coeficientes difira entre os modelos, um ajustamento deve ser implementado. Uma sugestão é o critério de Schwartz (1978), que implica uma estatística LR corrigida da seguinte forma:

$$\tilde{LR_N}(\beta,\gamma) = LR_N(\beta,\gamma) - \left[\frac{p}{2}\ln(N) - \frac{q}{2}\ln(N)\right]$$

Em que p e q são as dimensões dos vetores β e γ, respectivamente.

No entanto, até o momento nós podemos aplicar este teste somente para modelos não aninhados que foram estimados por Máxima Verossimilhança. Para o caso de modelos estimados por GMM, Rivers e Vuong (2002) apresentam algumas alternativas. Seja $Q_N^{(f)}(\hat{\beta})$ o valor da função critério para o modelo f aos parâmetros que a minimizam. Da mesma forma, $Q_N^{(g)}(\hat{\gamma})$ é o valor da função critério minimizada do modelo g. O teste de Vuong neste caso ficaria sendo:

$$V_N = \frac{\sqrt{N}(Q_N^{(f)}(\hat{\beta}) - Q_N^{(g)}(\hat{\gamma}))}{\hat{\sigma}_N}$$

Caso V_N seja negativo e significante – usando a distribuição normal padronizada – temos evidências que o modelo (f) é o mais adequado, caso tenhamos um V_N positivo e significante, temos evidências que o modelo g é o mais adequado. Lembrando que N é o tamanho da amostra, a única coisa que não temos é o $\hat{\sigma}_N$, que é um estimador da variância do numerador da expressão. Este estimador possui a seguinte forma:

$$\hat{\sigma}_N = R_*^T V R_*$$

Em que:

$$V = \frac{1}{N} Var \sum_{i=1}^{N} \begin{pmatrix} m^{(f)}(X_i,\hat{\beta}) - E(m^{(f)}(X_i\hat{\beta})) \\ D_i^{(f)} \\ m^{(g)}(Z_i,\hat{\gamma}) - E(m^{(g)}(Z_i\hat{\gamma})) \\ D_i^{(g)} \end{pmatrix}$$

E o outro componente é dado por:

$$R_* = (R^{(f)} - R^{(g)})$$

Sendo que $R^{(f)}$ é dado por:

$$R^{(f)} = \begin{pmatrix} 2W^{(f)}E(m^{(f)}(X_i,\hat{\beta})) \\ -A_*^{(f)}\Delta^{(f)T}B^T\{E(m^{(f)}(X_i,\hat{\beta})) \otimes E(m^{(f)}(X_i,\hat{\beta}))\} \end{pmatrix}$$

Em que D, A e Δ são matrizes que são definidas implicitamente, dependendo de como seja a escolha de matriz de ponderação na função critério do GMM.Para detalhes destas matrizes, ver Rivers e Vuong (2002).

Para fechar, vamos discutir alguns limites desta abordagem. O primeiro deles, de ordem mais técnica, diz respeito ao fato que o teste de Vuong tem um poder reduzido – A convergência para a distribuição normal precisa de amostras de tamanho razoável para ocorrer. A segunda observação é que, na medida que a classificação de modelos é relativa, isto não necessariamente implica que um deles é o melhor, apenas que é melhor do que o outro.

Passemos à última das críticas sobre a estimação de parâmetros de conduta.