

**UNIVERSIDADE FEDERAL DO CEARÁ – UFC
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO,
ATUÁRIA E CONTABILIDADE – FEAAC
CURSO DE PÓS-GRADUAÇÃO EM ECONOMIA – CAEN**

FRANCIS CARLO PETTERINI

**TRÊS ENSAIOS SOBRE ESCOLHAS E
O USO DA TÉCNICA MIXED LOGIT**

FORTALEZA

2009

FRANCIS CARLO PETTERINI

TRÊS ENSAIOS SOBRE ESCOLHAS E
O USO DA TÉCNICA MIXED LOGIT

Tese submetida à Coordenação do Curso de Pós-Graduação em Economia da Universidade Federal do Ceará – CAEN/UFC, como requisito parcial para obtenção do Título de Doutor em Economia.

Orientador: Prof. Márcio Veras Corrêa
Co-Orientador: Prof. Sérgio Aquino de Souza

FORTALEZA

2009

P596t Petterini, Francis Carlo
Três ensaios sobre escolhas e o uso da técnica
Mixed LOGIT / Francis Carlo Petterini. 2009.
94 f.

Orientador: Prof. Dr. Márcio Veras Correa
Co-Orientador: Prof. Dr. Sérgio Aquino de Souza

Tese (Doutorado) – Universidade Federal do Ceará,
Curso de Pós-Graduação em Economia, CAEN, Forta-
leza, Ceará, 2009.

1. Microeconomia. 2. Teoria da Escolha. I. Título.

CDD 338.5

FRANCIS CARLO PETTERINI LOURENÇO

TRÊS ENSAIOS SOBRE ESCOLHAS E
O USO DA TÉCNICA MIXED LOGIT

Tese submetida à Coordenação do Curso de Pós-Graduação em Economia da Universidade Federal do Ceará – CAEN/UFC, como requisito parcial para obtenção do Título de Doutor em Economia.

Aprovada em 27 de novembro de 2009.

BANCA EXAMINADORA

Prof. Dr. Márcio Veras Correa (Orientador)
Universidade Federal do Ceará

Prof. Dr. Sérgio Aquino de Souza (Co-Orientador)
Universidade Federal do Ceará

Prof. Dr. Marcos Costa Holanda
Universidade Federal do Ceará

Prof. Dr. José Raimundo de Araújo Carvalho Júnior
Universidade Federal do Ceará

Prof. Dr. Ricardo Brito Soares
Universidade Federal do Ceará

AGRADECIMENTOS

À CAPES e à FUNCAP, pelo apoio financeiro com a manutenção da bolsa de auxílio.

Aos professores do CAEN, em particular aos meus dois orientadores, pelos conhecimentos transmitidos.

À Vitor Hugo Miro Couto Silva, por disponibilizar suas bases de dados e estar sempre disposto a ajudar.

Aos amigos Roberto Tatiwa Ferreira e Christiano Modesto Penna, por nossas conversas.

E todos os demais que de alguma forma contribuíram para a realização deste trabalho.

RESUMO

Como as pessoas fazem suas escolhas? O entendimento desta questão é crucial, por exemplo, para uma montadora de veículos que pretende lançar um novo modelo de automóvel no mercado, ou para um partido político que precisa decidir qual candidato lançar. Pois comprar um carro A ou B ou votar num candidato A ou B costumam ser decisões de escolha individual, então, ao se entender como as escolhas são feitas se aumentam as chances de sucesso de venda ou de votação.

Esta tese trata fundamentalmente disso: escolhas. Mais especificamente, trata de escolhas à luz da Teoria Econômica. E para conduzir uma seqüência de três exercícios empíricos sobre o tema usa-se uma técnica de vanguarda denominada de Mixed LOGIT.

No primeiro exercício o conjunto de escolhas das pessoas contém opções para compra de automóveis, e se estudará como são feitas as escolhas nesse ambiente de consumo. Por consequência desse entendimento, será possível computar elasticidades-preço, preço-cruzado, *markups* e excedentes do consumidor e do produtor.

No segundo exercício o conjunto de escolhas representará opções de tipos de trabalhadores, e se estudará como as firmas fazem escolhas para ocupar postos de trabalho vagos. Por consequência desse entendimento, será possível compreender como as redes sociais de trabalhadores podem afetar o mercado de trabalho.

E no terceiro exercício o conjunto de escolhas representará os candidatos a Presidência da República nas eleições de 2006, e se estudará como os eleitores podem ser induzidos a votar em determinado candidato. Por consequência desse entendimento, será possível estimar o quanto um programa de transferência direta de renda como o Bolsa Família é capaz de influenciar o resultado de uma eleição.

Palavras-chave: Teoria da escolha; *Mixed* LOGIT; Mercado de automóveis; Modelos de *matching*; Eleições.

ABSTRACT

How are choices made? The understanding of this matter is crucial, for instance, for an automobile industry firm that intends to launch a new model on the market, or for a political party that needs to decide which candidate will run for the office. It happens that buying the A or B car or voting for the A or B candidate are usually individual choices. Therefore, when one understands how the choices are made, the chances of selling more cars or winning the election are increased greatly.

That is the primary subject of this study: choices. More specifically, it deals with choices at the light of the Economical Theory. And in order to lead a sequence of three empiric exercises about the theme a forefront technique named *Mixed* LOGIT is used.

In the first exercise the set of the people's choices contains options for the buying of automobiles, and how the choices are made in that consumption environment will be studied. Consequently, it will be possible to compute own-price-elasticity, cross-price-elasticity, *markups* and surplus of the consumer and the producer.

In the second exercise the set of choices will represent options of kinds of workers, and how the firms make the choices on how to fill the job vacancies will be studied. Consequently, it will be possible to understand how the social networks can affect the labor market.

In the third exercise the set of choices will represent the Presidential candidates running for the 2006 election, and how the voters can be induced to voting to a certain candidate will be studied. Consequently, it will be possible to estimate how a cash transfer programme, such as Programa Bolsa Família (Family Grant Programme) can influence the result of the election.

Keywords: Choice theory; *Mixed* LOGIT; Auto market; *Matching* models; Elections.

SUMÁRIO

INTRODUÇÃO.....	10
1. A TRIBUTAÇÃO NAS VENDAS DE AUTOMÓVEIS NO BRASIL: QUEM PAGA A MAIOR PARTE DA CONTA?.....	12
1.1 APRESENTAÇÃO	12
1.2 O CONTEXTO DO MERCADO BRASILEIRO.....	13
1.3 A ESTRATÉGIA EMPÍRICA.....	15
1.3.1 O lado da oferta.....	16
1.3.2 O lado da demanda e os modelos LOGIT e <i>Mixed</i> LOGIT	17
1.3.3 A opção externa	22
1.3.4 As medidas de bem-estar	23
1.4 A AMOSTRA	25
1.5 RESULTADOS ESTIMADOS	26
1.6 SIMULAÇÕES.....	35
1.7 CONCLUSÕES	39
2. MAIS AMIGOS, MAIS CHANCE DE EMPREGO? UM EXERCÍCIO EMPÍRICO USANDO OS DADOS DE PORTUGAL	40
2.1 APRESENTAÇÃO	40
2.2 O CONTEXTO DAS <i>NETWORKS</i> E DO MERCADO DE TRABALHO	41
2.3 A ESTRATÉGIA EMPÍRICA.....	46
2.3.1 A taxa de <i>pseudo-matching</i>	46
2.3.2 O sistema de numeração posicional	47
2.3.3 O papel do <i>Mixed</i> LOGIT.....	48
2.3.4 As simulações de mudanças da <i>network</i>	50
2.4 A AMOSTRA	52
2.5 RESULTADOS ESTIMADOS.....	56
2.6 SIMULAÇÕES.....	58
2.7 CONCLUSÕES	61
3. POR QUANTOS VOTOS UMA BOLSA FAMÍLIA PODERIA SER TROCADA?.....	62
3.1 APRESENTAÇÃO	62
3.2 O CONTEXTO DAS ELEIÇÕES PRESIDENCIAIS DE 2006 E DO DEBATE ACADÊMICO	63
3.3 A ESTRATÉGIA EMPÍRICA.....	70
3.3.1 A questão do <i>viés de agregação</i>	70
3.3.2 O papel do <i>Mixed</i> LOGIT.....	72
3.4 A AMOSTRA	75
3.5 RESULTADOS ESTIMADOS	78
3.6 CONCLUSÕES	81
CONSIDERAÇÕES FINAIS.....	82
REFERÊNCIAS BIBLIOGRÁFICAS.....	83
ANEXO I.....	87
ANEXO II.....	88
ANEXO III	91
ANEXO IV.....	93
ANEXO V	94

INTRODUÇÃO

Como as pessoas fazem suas escolhas? O entendimento dessa questão é crucial, por exemplo, para uma montadora de veículos que pretende lançar um novo modelo de automóvel, ou para um partido político que precisa decidir qual candidato lançar. Pois comprar um carro A ou B ou votar num candidato A ou B costumam ser decisões de escolha individual, então, ao se entender como as escolhas são feitas se aumentam as chances de sucesso de venda ou de votação.

Esta tese trata fundamentalmente disso: escolhas. Mais especificamente, trata de escolhas à luz da Teoria Econômica. E para conduzir uma seqüência de exercícios empíricos sobre o tema se usa uma técnica denominada de *Mixed* LOGIT.

Na teoria, o ponto de partida para se entender as escolhas é a discussão acerca das preferências individuais. Nesse sentido, em um primeiro momento costuma-se definir um conjunto bastante genérico de alternativas, X , e uma estrutura de preferências que relaciona cada par de elementos desse conjunto.

Em um segundo momento adiciona-se mais estrutura nas preferências, de forma a representá-las através de uma função definida no espaço X . Formalmente isso se reflete numa “função utilidade”, $U : X \rightarrow \mathfrak{R}$, onde para todo $x \in X$ se terá um determinado número $U(x)$.

Mais ainda, assumem-se algumas propriedades para a relação de preferência para tornar quem escolhe em um ser “racional”. Formalmente, isto implicará que, dado o conjunto de alternativas, o indivíduo particularmente escolherá um $x \in X$ que gere o maior valor de $U(x)$. E até aqui a estrutura de escolhas é bastante geral.

A organização anterior leva a Teoria do Consumidor, que particulariza X como um conjunto de alternativas de consumo. E em seguida, o arcabouço leva o indivíduo (racional) a fazer suas escolhas maximizando $U(x)$ frente a uma restrição orçamentária $p \cdot x \leq m$, onde p é um vetor de preços e m é uma quantidade de dinheiro que dispõe. Esse processo acaba por gerar um vetor de funções de demandas $x = d(p, m)$. E assim a demanda torna-se o resultado final das escolhas.

Se um pesquisador está interessado em estudar empiricamente as escolhas de mercado, uma estratégia de trabalho é apelar para o paradigma do “agente representativo” – onde todos os indivíduos têm igual $U(x)$ – e modelar $x_i = d(p, m_i) + \varepsilon_i$, onde i indexa uma observação (um indivíduo) e ε_i é um termo que representa discrepâncias nos dados observados. Mas isso leva o pesquisador a um problema que pode inviabilizar o estudo empírico: o grande número parâmetros a se estimar. Pois se X possui N elementos, no mínimo $N \cdot (N + 1)$ parâmetros precisarão ser estimados, o que demandará uma base de dados extremamente rica.

Uma alternativa ao paradigma do “agente representativo”, popularizada pelos trabalhos de Daniel McFadden, passa a aceitar alguma heterogeneidade entre os indivíduos reescrevendo a função utilidade de um indivíduo i por $U_i(x) = V(x) + \varepsilon_i$, onde $V(x)$ é um termo comum a todos e ε_i é uma idiosincrasia.

Essa estrutura passa a ser empiricamente tratável ao se assumir que ε_i é independente e identicamente distribuído entre os agentes com estrutura de

probabilidade de Valor Extremo Tipo I. Pois quando se associada isso a condição de maximização de utilidade se obtêm que a probabilidade de que o indivíduo i venha a optar pela escolha x é dada por: $\Pr(x) = \exp(V(x)) / \sum_{y \in X} \exp(V(y))$.

Dado esse resultado, usando as idéias de Kevin Lancaster, mais do que se modelar a opção x em si, modelam-se as características da opção x que estão presentes (em maior ou menor grau) em todas as outras opções $y \in X$. E isso acaba por reduzir drasticamente o número de parâmetros a se estimar para desenhar empiricamente uma curva de demanda (o resultado final das escolhas).

Então, o procedimento padrão é igualar $\Pr(x)$ a parcela de mercado da opção x (uma forma de se escrever a demanda), destringir x em um vetor de características e usar um pouco de álgebra para encontrar uma estrutura que possa ser operacionalizada com técnicas econométricas tradicionais.

Mas a associação das idéias de McFadden e Lancaster vai além da possibilidade de se operacionalizar especificamente estudos empíricos de demanda de mercado. Pois o raciocínio vale para qualquer tipo de conjunto de alternativas X , desde que exista uma adequada correspondência para $\Pr(x)$. Por exemplo, X pode ser o conjunto de opções de uma eleição, de forma simplificada imagine $X = \{\text{candidato A, candidato B, votar branco ou nulo}\}$ e $\Pr(x)$ é o percentual de votos em cada alternativa. E modelando as características de cada candidato o pesquisador pode entender o que faz A ter maior votação do que B, ou vice-versa.

Seguindo os avanços feitos nessa técnica nos últimos anos, se o pesquisador tem algum domínio de programação computacional ele pode ir mais além: pode aceitar um maior grau de heterogeneidade entre os agentes. Isso se dá reescrevendo a função utilidade de um indivíduo i por $U_i(x) = V_i(x) + \varepsilon_i$, onde $V_i(x)$ e ε_i passam a ser idiossincráticos.

A técnica que provém destes desenvolvimentos é denominada de *Mixed LOGIT*, e com ela pode-se modelar de forma bastante geral qualquer estrutura de escolha. Esta tese faz uso dessa técnica para estudar três ambientes de escolha, cada um descrito em um dos ensaios que compõem o trabalho.

No primeiro ensaio o conjunto X contém opções para compra de automóveis, e se estudará como as pessoas fazem escolhas nesse ambiente de consumo. Por consequência desse entendimento, será possível computar elasticidades-preço, preço-cruzado, *markups* e excedentes do consumidor e do produtor.

No segundo ensaio o conjunto X representará opções de tipos de trabalhadores, e se estudará como as firmas fazem escolhas para ocupar postos de trabalho vagos. Por consequência desse entendimento, será possível compreender como as redes sociais de trabalhadores podem afetar o mercado de trabalho.

E no terceiro e último ensaio o conjunto X representará os candidatos a Presidência da República nas eleições de 2006, e se estudará como os eleitores podem ser induzidos a votar em determinado candidato. Por consequência desse entendimento, será possível estimar o quanto um programa de transferência direta de renda como o Bolsa Família é capaz de influenciar o resultado de uma eleição.

1. A tributação nas vendas de automóveis no Brasil: Quem paga a maior parte da conta?

1.1 Apresentação

O desenvolvimento da indústria automobilística foi determinante na fase de industrialização brasileira durante as décadas de 1950 e 1960. Anos depois, o setor foi emblemático no processo de abertura comercial e modernização da indústria na década de 1990. Em 2008, a participação do setor no PIB industrial foi de 23,3%, o que equivaleu a 5,5% do PIB nacional, empregando 1,5 milhão de trabalhadores (direta e indiretamente). Além disso, o Brasil já é quinto maior mercado consumidor do mundo (2,8 mil unidades vendidas em 2008), o sexto maior produtor (3,2 mil unidades) e possui a décima maior frota de automóveis do mundo (25,6 mil unidades)¹.

Fatos e números dessa magnitude também se refletem na tributação do setor, quando a soma da arrecadação de IPI, ICMS, PIS e COFINS da venda de veículos chegou a R\$ 39,4 bilhões em 2008 (quase o PIB do Ceará²). Um montante dessa ordem leva a duas perguntas clássicas na teoria da tributação: 1) Qual o ônus para sociedade da imposição desses impostos?; e, 2) Quem suporta a maior parte do ônus tributário? Os consumidores ou os produtores? Ou, em termos coloquiais esta pergunta pode ser reformulada da seguinte forma: quem paga a maior parte da conta?

Sabe-se que a existência de impostos altera o equilíbrio entre demanda e oferta de forma distorcida. Com a tributação, o preço recebido pelo produtor passa a ser diferente do preço pago pelo consumidor, gerando um resultado onde ambos perdem com a instituição do imposto. Os consumidores são prejudicados porque pagam um preço maior do que o que pagariam na ausência do imposto. Os vendedores também são punidos, porque poderiam receber um preço maior e ter maior demanda. O governo, por sua vez, é o único agente a ganhar alguma coisa: arrecadação. Este resultado reduz o bem-estar social, visto que a arrecadação do governo sempre é menor do que a combinação das perdas de consumidores e vendedores.

Tanto o tamanho da perda de bem-estar social como a distribuição do ônus tributário entre consumidores e vendedores dependem das condições de demanda e oferta (em particular, das elasticidades). Portanto, o primeiro passo para responder as duas questões postas acima consiste em determinar empiricamente os parâmetros que definem a demanda e a oferta. Já o segundo passo consiste em determinar as métricas a serem utilizadas para calcular as magnitudes econômicas desejadas, ou seja, determinar empiricamente a perda de bem-estar e a distribuição das perdas entre consumidores e produtores causadas pela tributação.

A partir dos dados do mercado, um painel não-balanceado com 66 modelos de automóveis (das oito principais montadoras instaladas no Brasil) nos anos de 2005 a 2008, aplica-se uma técnica de modelagem e estimação de demanda e oferta via *Mixed LOGIT*. Assim, obtêm-se estimativas para os parâmetros da demanda e os custos marginais.

As estimativas dos parâmetros da demanda e dos custos marginais darão base para a simulação de um cenário de isenção completa de tributos sobre as vendas de

¹ Os dados são apresentados em ANFAVEA (2006a e 2009).

² Estimado em R\$ 48 bilhões para 2008.

automóveis. Mais precisamente, darão base para o cálculo de preços decorrentes da desoneração tributária, que possibilitarão calcular a variação do lucro operacional por fabricante (pré e pós-isenção) e a variação compensatória dos consumidores. Assim, pode-se gerar uma medida de ganho de bem-estar social³ e de como este ganho é distribuído entre produtores e consumidores, ou seja, uma estimativa da partição do ônus tributário.

Além desta introdução, a seção 1.2 apresenta uma breve caracterização do mercado brasileiro de automóveis. A seção 1.3 discute a estratégia empírica. As seções 1.4 e 1.5 apresentam, respectivamente, a amostra e os resultados estimados para elasticidades-preço, preço-cruzado da demanda, *markups* e custos marginais. A seção 1.6 mostra as simulações da desoneração tributária. Depois se apresentam as considerações finais deste ensaio.

1.2 O contexto do mercado brasileiro

O mercado brasileiro de autoveículos⁴ (automóveis, veículos comerciais leves, caminhões e ônibus) é bastante relevante no contexto internacional. A tabela 1, abaixo, mostra que o país é sexto maior produtor, com 3.216 mil unidades em 2008, e o quinto maior mercado consumidor do mundo, com 2.820 mil unidades vendidas em 2008 (destas, 2.193 referem-se apenas aos automóveis, 78%). Mais ainda, o Brasil possui a décima maior frota no mundo, 25.596 automóveis registrados em 2007.

Tabela 1 – Frota de automóveis e produção e vendas de autoveículos novos – <i>ranking</i> dos países – valores em 1.000 unidades.						
	Frota de automóveis (2007)		Produção de autoveículos (2008)		Vendas de autoveículos (2008)	
1°	EUA	247.265	Japão	11.564	EUA	13.493
2°	Japão	75.715	China	9.345	China	9.380
3°	China	45.584	EUA	8.705	Japão	5.082
4°	Alemanha	44.021	Alemanha	6.041	Alemanha	3.425
5°	Itália	40.368	Coréia do Sul	3.807	Brasil	2.820
6°	França	37.033	Brasil	3.216	França	2.574
7°	Reino Unido	35.354	França	2.568	Reino Unido	2.485
8°	Espanha	27.174	Espanha	2.542	Itália	2.430
9°	México	25.686	Índia	2.315	Índia	1.980
10°	Brasil	25.596	México	2.191	Canadá	1.674

Fonte: ANFAVEA (2009).

No segmento dos automóveis, o mercado brasileiro é caracterizado por um oligopólio, onde oito empresas dominam a fabricação e as vendas. São elas: Chevrolet

³ Note que “perdas” causadas pela tributação é o mesmo que “ganhos” causados pela desoneração da tributação. Em algumas partes do texto, dependendo do contexto, essas duas formas de exposição podem se alternar.

⁴ Trata-se de autoveículos, mais do que apenas automóveis, para viabilizar comparações internacionais.

(General Motors), Ford, Volkswagen, Fiat, Toyota, Honda, Renault e Peugeot-Citroen.

Nos últimos cinco anos a Fiat vem mantendo a liderança nas vendas, em média com 26% do mercado, seguida de perto pela Volkswagen (VW) e pela GM, com médias de 24% e 22%, respectivamente. No mesmo período, também vem se observando uma gradual perda de mercado pela Ford a ascensão das marcas Honda, Renault, Peugeot, Citroen e Toyota. A participação das marcas no mercado brasileiro de automóveis é exposta na tabela 2, adiante, para o ano de 2008.

Tabela 2 - Participação das marcas no mercado brasileiro de automóveis - 2008 - %									
1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Fiat	VW	GM	Ford	Honda	Renault	Peugeot	Citroen	Toyota	Outras
25,4	24,0	21,3	8,7	5,1	5,0	3,6	3,0	2,3	1,6

Fonte: FENABRAVE (2009).

A explicação para a liderança da Fiat, VW e GM está no fato de produzirem os modelos populares mais vendidos no Brasil: VW Gol (líder absoluto de vendas a anos, com 11 a 12% do total do mercado, dependendo do período), Fiat Palio (8 a 9%), Fiat Uno (6 a 7%), GM Celta (6 a 7%) e VW Fox (3 a 4%). Uma fusão entre Fiat e GM, por exemplo, levaria essa nova firma a ser a líder em vendas de carros populares no Brasil, que representa entre 50 e 55% do total do mercado.

A configuração atual do setor foi estabelecida em grande parte pelas transformações ocorridas na década de 1990, marcada pela liberalização das importações que, acompanhada por uma redução nas alíquotas dos impostos, promoveu a entrada de modelos importados no mercado brasileiro (detalhes em Fiúza, 2002). O Plano Real também exerceu fortes impactos sobre o mercado de automóveis, não só ao promover a estabilidade de preços, mas também por gerar maior disponibilidade de crédito, impulsionando o consumo de bens duráveis.

No âmbito setorial, diversos acordos foram firmados objetivando desenvolver o parque industrial automobilístico nacional e aprimorar as relações de comércio exterior (principalmente acordos bilaterais e no Mercosul). O Acordo Automotivo de 1993 promoveu o fomento ao chamado “carro popular” e gerou um forte impulso nas vendas internas. Com um bônus tributário e a expansão do crédito para potenciais consumidores das classes de renda mais baixa nos anos seguintes, a demanda por carros da categoria 1.0 foi a que mais cresceu.

Já o Regime Automotivo de 1995 reduziu as tarifas de importação e promoveu benefícios fiscais para que as empresas instaladas no Brasil mantivessem suas plantas e novas empresas começassem a produzir em território nacional. Desta forma, na segunda metade da década de 1990, diversas montadoras se instalaram no país e as já existentes construíram novas e modernas plantas. O mercado, então, tornou-se mais diversificado e dinâmico.

A segunda metade dessa década também foi caracterizada por crises financeiras internacionais que afetaram a balança de pagamentos brasileira e as contas fiscais do governo, ameaçando em diversos momentos os incentivos tributários oferecidos. A crise asiática, por exemplo, refletiu fortemente no lado fiscal e o governo viu-se obrigado a aumentar a receita tributária, refletindo em aumentos nas alíquotas sobre os automóveis. As vendas passaram a cair, e só foram revertidas após novos acordos assinados em 1998 e 1999 promovendo uma reforma na estrutura tributária do setor.

A tabela 3, abaixo, expõe a razão entre vendas (licenciamentos) de veículos usados e novos nos últimos anos. Observe que a razão vem caindo, de 4,3 em 2005 para 2,8 em 2008. Conseqüência, principalmente, da paulatina facilidade em se financiar a compra de modelos novos, induzido a renovação da frota (com idade média de 12,7 anos, frente a 8,4 anos nos EUA, em 2007).

Tabela 3 - Razão entre vendas de automóveis usados e novos			
2005	2006	2007	2008
4,3	3,7	3,1	2,8

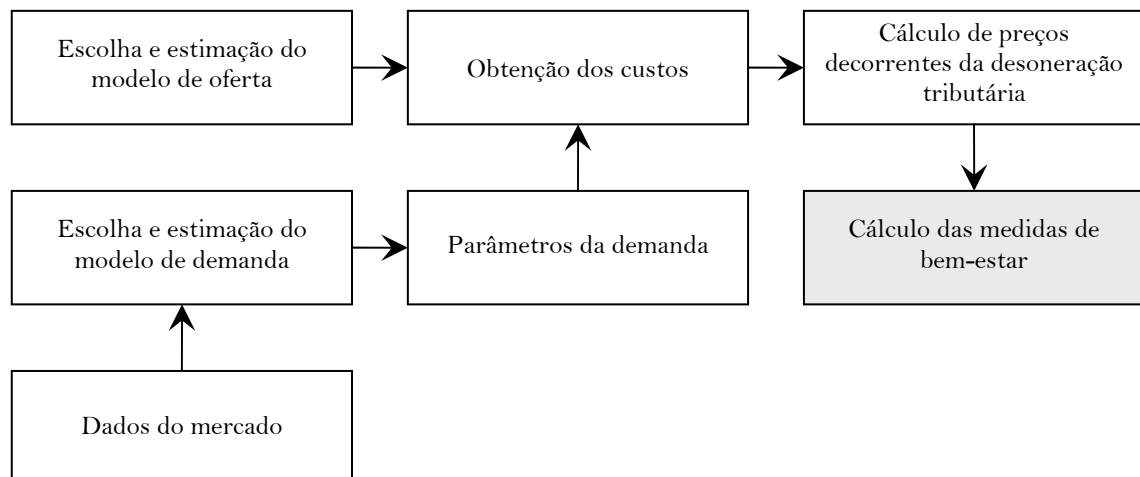
Fonte: FENABRAVE (2009).

Nos últimos anos o setor também tem se destacado por promover inovações como o dispositivo Flex (bi-combustível). Com a redução do preço do álcool e com a definição do governo de que os automóveis bi-combustível pagariam alíquota de IPI mais baixa (com os mesmos incentivos dos veículos a álcool), a produção em série de veículos do gênero passou a ser viável no Brasil. Assim, o primeiro modelo bi-combustível a chegar às revendas foi o VW Gol Total Flex 1.6, lançado em abril de 2003. A partir de então, a participação de veículos bi-combustível nas vendas de veículos novos cresceu de forma acelerada.

Em suma, os resultados apresentados nessa seção indicam que, apesar da dimensão atual já ser considerável, o mercado brasileiro ainda tem um bom potencial de crescimento.

1.3 A estratégia empírica

O diagrama abaixo descreve a estratégia empírica que será usada neste ensaio, e pormenorizada adiante nas sub-seções. A partir dos dados do mercado se aplicará a técnica de estimação de demanda e oferta proposta por Berry, Levinsohn e Pakes (1995), doravante BLP (que é uma aplicação do modelo *Mixed LOGIT*). Assim, obtêm-se estimativas para os parâmetros da demanda e os custos marginais.



As estimativas dos parâmetros da demanda e dos custos marginais darão base para a simulação de um cenário de isenção completa de tributos sobre as vendas de automóveis. Mais precisamente, darão base para o cálculo de preços decorrentes da desoneração tributária, que possibilitarão calcular a variação do lucro operacional por fabricante (pré e pós-isenção) e a variação compensatória dos consumidores. Assim, pode-se gerar uma medida de ganho de bem-estar social e de como este ganho é distribuído entre produtores e consumidores, ou seja, uma estimativa da partição do ônus tributário.

1.3.1 O lado da oferta

Seguindo a técnica BLP e as exposições Nevo (2000b) e Ferraz, Fiúza e Motta (2001) no lado da oferta supõe-se que os automóveis são produzidos por empresas multi-produtos que vendem itens diferenciados em um mercado oligopolista, como o descrito na seção anterior, e competem em preços (jogo Bertrand). Formalmente, cada empresa $f = 1, \dots, F$ produz um subconjunto J_f do conjunto composto por $j = 1, \dots, J$ modelos diferentes de carros do mercado. E então, cada empresa busca maximizar a seguinte função de lucro:

$$(1) \quad \pi_f = \sum_{j \in J_f} (p_j \times (1 - \tau_j) - c_j) \times s_j(p) \times M$$

Onde π_f representa o lucro da firma, p_j representa o preço, c_j o custo marginal, τ_j a tributação *ad valorem* sobre preço ao consumidor e s_j a parcela de mercado do modelo $j \in J_f$. Este último em função do vetor p , contendo os preços de todos os J modelos. Por fim, M é o tamanho do mercado.

As condições de primeira ordem do problema de maximização de lucros levam ao seguinte sistema (de onde se deduzem as curvas de reação):

$$(2) \quad \sum_{q \in J_f} \left\{ (p_q \times (1 - \tau_q) - c_q) \times \frac{\partial s_q(p)}{\partial p_j} \right\} + s_j(p) \times (1 - \tau_j) = 0 \quad ; \quad j = 1, \dots, J$$

Para computar o equilíbrio Bertrand-Nash usa-se uma matriz auxiliar Δ , de tamanho $J \times J$, onde:

$$(3) \quad \Delta_{jq} = \begin{cases} -\frac{\partial s_j}{\partial p_q} & , \quad \text{para } j \text{ e } q \text{ produzidos pela mesma firma} \\ 0 & , \quad \text{c.c.} \end{cases}$$

Então, do sistema (2) em forma matricial obtêm-se o seguinte vetor de preços de equilíbrio⁵:

⁵ Uma alternativa de decomposição é: $p = c + t(p) + \Delta(p^\tau)^{-1} s^\tau(p^\tau) = c + t(p) + l(p)$. Onde: t é um vetor onde em cada linha observa-se $t_j = \tau_j \times p_j$. A equação decompõe o vetor de preços ao consumidor na soma de outros três vetores, respectivamente: custo marginal (c); tributação marginal (t); e, lucro operacional (l).

$$(4) \quad p^\tau = c + \Delta(p^\tau)^{-1} s^\tau(p^\tau)$$

Onde o sobrescrito τ indica que os preços e as parcelas de mercado estão multiplicados por $(1 - \tau_j)$ em cada linha j .

Note que conhecendo preço e a carga tributária sobre o preço de cada veículo, bem como sua parcela de mercado, em (4) os custos marginais podem ser facilmente estimados a partir de uma estimativa da matriz (3) (fazendo $\hat{c} = p^\tau - \Delta(p^\tau)^{-1} s^\tau(p^\tau)$). E tendo estimativas dos custos marginais, e considerando-os constantes, podem-se simular mudanças de preços conseqüentes de mudanças de alíquotas de tributação ($p^{pós}$, em particular, para o cenário: $\tau_j = 0 ; \forall j$) resolvendo o sistema:

$$p^{pós} = \hat{c} + \Delta(p^{pós})^{-1} s(p^{pós}).$$

A próxima sub-seção tratará da construção da estimativa de (3).

1.3.2 O lado da demanda e os modelos LOGIT e *Mixed* LOGIT

Para se estimar (3), antes é preciso ter uma estimativa de demanda. Existem duas categorias de modelos de demanda: por produtos homogêneos e por produtos diferenciados. A primeira categoria encontra metodologias consolidadas e que, em geral, se apóiam em métodos empíricos relativamente simples. A segunda categoria, no entanto, possui uma série de desafios metodológicos que só foram (parcialmente) superados recentemente.

Os modelos de demanda por produtos homogêneos tratam de mercados onde o consumidor percebe todos os bens como idênticos, em todas as dimensões. São caracterizados, principalmente, por ter um preço único e por existir apenas uma equação para todo o mercado. Assim, pode-se utilizar uma econometria tradicional, de fácil execução. O problema mais comum de se encontrar nessa abordagem é a endogeneidade dos preços, quando se precisa aplicar o método das variáveis instrumentais (VI).

Já no que tange aos modelos de demanda por produtos diferenciados, que é o caso deste ensaio, a análise é mais complexa e existem duas classes de abordagem. A primeira é formada por modelos baseados em um “consumidor representativo” que atribui uma utilidade direta ao consumo dos bens ofertados no mercado.

Nesse caso, o pesquisador se depara com número de equações igual ao número de produtos, gerando um sistema com muitos parâmetros. De fato, em cada equação, além dos deslocadores de demanda, deve-se especificar o efeito próprio (a sensibilidade da demanda do bem j em relação a seu próprio preço) e o efeito cruzado (a sensibilidade da demanda do bem j em relação ao preço do bem rival r). Em termos teóricos o número elevado de parâmetros não impõe severas limitações, o que não é verdadeiro para análise empírica⁶.

⁶ Exemplos notórios que se inserem nesta classe de modelos são o LES (Linear expenditure system) e o AIDS (Almost Ideal Demand System), detalhados em Deaton e Muellbauer (1980). Ver Asano, Barbosa e Fiúza (2004) para uma aplicação desta classe de modelos no contexto brasileiro. Outro comentário a se fazer é que na estrutura de “consumidor representativo” não há como avaliar o efeito de introdução de novos produtos, como foi o caso da invenção da Minivan discutido em Petrin (2002). Ou, noutro exemplo, da inserção dos motores Flex.

A segunda classe de modelos usa a proposta de Lancaster (1966), que consiste em assumir que os consumidores atribuem utilidade às características dos bens, e não aos bens em si. Nesta perspectiva, o consumidor escolhe o bem que lhe confere a melhor combinação de atributos.

A partir da proposta de Lancaster, surgem os modelos de escolha discreta com utilidade aleatória (Random Utility Models, RUM⁷). A principal vantagem dessa abordagem reside na redução do número de parâmetros a serem estimados. Os modelos de escolha discreta rompem com a relação exponencialmente crescente entre número de produtos e número de parâmetros, permitindo aplicação em mercados caracterizados pela presença de muitas variedades.

O modelo LOGIT (multinomial) é o mais simples da classe RUM. Formalmente, contextualiza-se o modelo para um consumidor i atribuir ao produto j (entre J produtos) a seguinte utilidade:

$$(5) \quad U_{ij} = V_j + \varepsilon_{ij} \quad ; \quad V_j = -\alpha p_j + x_j \beta + \xi_j$$

Onde V_j é o componente da utilidade atribuído às características do produto (e é um termo comum a todos os consumidores) e ε_{ij} representa as idiosincrasias do consumidor em relação ao produto (é o único termo que diferencia a utilidade entre os consumidores). Mais ainda, V_j é decomposto de forma onde α é um escalar (positivo), p_j é o preço⁸, x_j representa um vetor (linha) de características dos produtos de dimensão K , β é um vetor (coluna) de parâmetros e ξ é um índice que agrupa outras características não incluídas no vetor x_j .

Mais ainda, a especificação da demanda precisa da definição da opção externa ($j = 0$), que representa várias atitudes: esperar para comprar, desistir de comprar, ganhar de presente etc. Sua definição é fundamental para evitar demandas agregadas perfeitamente inelásticas, algo pouco plausível na maioria dos casos. Uma normalização típica e conveniente⁹ de modelos de escolha discreta é assumir $V_0 = 0$.

Seguindo a teoria econômica, o consumidor escolhe o produto que lhe confere maior utilidade. Ao assumir essa racionalidade e que ε_{ij} é i.i.d. com distribuição probabilidade de Valor Extremo do Tipo I é possível obter uma forma analítica para a probabilidade do consumidor i escolher determinado produto j , cuja fórmula é dada por¹⁰:

$$(6) \quad \Pr_{ij} = \frac{\exp(V_j)}{1 + \sum_{r=1}^J \exp(V_r)}$$

Observe que o lado direito da equação acima não é indexado para o consumidor i . Portanto, a probabilidade não condicional de consumo é idêntica à probabilidade condicional ($\Pr_{ij} = \Pr_j$). Assim, a metodologia iguala essa probabilidade à fatia de mercado do produto j ($\Pr_{ij} = s_j$). Logo, mesmo sem supor a existência de um

⁷ Ver McFadden (2001) para a seqüência história dessa abordagem.

⁸ Que pela hipótese do jogo Bertrand é função de todos os outros preços – ver equação (4).

⁹ Isso não causa problemas, por conta das características ordinais da utilidade.

¹⁰ O resultado é apresentado no ANEXO I, e detalhes podem ser vistos em McFadden (1981).

consumidor representativo, chega-se a uma estrutura que depende apenas de dados observados em nível de mercado (algo muito apropriado para o uso empírico).

Por (5), cada V_j é função dos preços. E ao igualar o lado esquerdo de (6) as fatias de mercado s_j , as seguintes elasticidades podem ser computadas:

$$(7) \quad \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = -\alpha p_j (1 - s_j) \quad (\text{elasticidade preço}) \quad ; \quad \frac{\partial s_{j \neq r}}{\partial p_r} \frac{p_r}{s_{j \neq r}} = \alpha p_r s_r \quad (\text{cruzada})$$

Observe que tendo uma estimativa $\hat{\alpha}$ em mãos, substituindo-a em (7) tem-se uma estimativa de (3) e, por conseqüência, dos custos marginais de produção de cada modelo j , permitindo a execução dos exercícios de simulação.

Mais ainda, com uma simples manipulação da forma (6) e o conhecimento *a priori* de s_0 , chega-se a seguinte forma log-linear para se computar $\hat{\alpha}$:

$$(8) \quad \ln s_j - \ln s_0 = -\alpha p_j + x_j \beta + \xi_j$$

A equação (8) é facilmente implementável com uma econometria tradicional. No entanto, a exemplo da equação para produtos homogêneos, o preço é uma variável endógena devido a sua correlação com o erro (ξ , que representa o prestígio da marca, a influência da propaganda etc.). Mais uma vez, uma solução seria usar o método VI.

Mas mesmo de posse de bons instrumentos e de estimadores com propriedades econométricas desejáveis pelo pesquisador (coeficientes significantes e com sinal correto) o modelo LOGIT pode não ser adequado para análise econômica, pois apresenta alguns problemas conceituais.

De fato, observe a partir de (7) que o aumento da fatia de mercado do produto j decorrente do aumento percentual do preço p_r depende apenas de r . Isto significa que um aumento percentual de p_r afetará de forma idêntica todos os outros produtos no mercado (isso é chamado de *competição não localizada*¹¹). Trata-se de uma propriedade pouco plausível em mercados com produtos diferenciados.

Com efeito, uma das motivações principais para estudar este tipo de mercado é justamente, para cada par de produtos, distinguir entre mais próximos e mais distantes em relação ao grau de diferenciação. Tal restrição é uma manifestação da propriedade da Independência de Alternativas Irrelevantes (IAI) presente nestes modelos¹², amplamente discutida em modelos de demanda desagregados (em nível do consumidor).

Outro problema, apontado por Huse e Salvo (2005), é que a forma funcional do LOGIT também impõe que as elasticidades dependem diretamente do preço dos produtos. Desta forma, valores elevados de p_j levam a elasticidades-preço também elevadas (em valor absoluto). O que pode ser pouco plausível, uma vez que produtos mais caros tendem a estar em mercados com demandas mais inelásticas.

A solução mais simples para esta restrição do modelo LOGIT é oferecida pelo Modelo LOGIT Agrupado (ou *Nested LOGIT*). Neste caso, o pesquisador define *a priori* os agrupamentos (ou segmentos do mercado) e assume que produtos

¹¹ Para exemplificar, considere o Honda Civic e o Fiat Siena (que possuíam fatias de mercado semelhantes em 2008). É plausível assumir que variações no preço do Toyota Corolla tenham um impacto maior sobre a demanda do Civic do que do Siena, por serem produtos que visam atender o mesmo nicho de mercado. No entanto, as elasticidades-cruzadas estimadas entre Civic-Corolla e Siena-Corolla serão idênticas dependendo somente de variáveis relativas ao Corolla e independente do qual próximo estão esses produtos no espaço de características.

¹² Detalhes podem ser vistos em McFadden, 1981.

pertencentes ao mesmo grupo possuem grau de substituição mais elevado (ver Ferraz, Fiúza e Motta, 2001). Todavia, o *Nested* LOGIT também apresenta alguns problemas: 1) a definição *ad hoc* dos agrupamentos; e, 2) a permanência da propriedade IAI dentro de cada agrupamento.

Outra solução, bem mais sofisticada e que será usada nos três ensaios desta tese, passa a usar o modelo *Mixed* LOGIT, onde a utilidade marginal pelas características do produto, incluindo preços, varia de consumidor para consumidor. Formalmente, neste modelo o consumidor i atribui ao produto j a seguinte utilidade:

$$(9) \quad U_{ij} = V_{ij} + \varepsilon_{ij} \quad ; \quad V_{ij} = -\alpha_i p_j + x_j \beta_i + \xi_j \quad ; \quad V_{i0} = 0$$

Onde os coeficientes α_i e β_i são aleatórios (na perspectiva do pesquisador). E nesse ensaio, seguindo as exposições encontradas em Berry (1994), Berry, Levinsohn e Pakes (1995) e Berry e Pakes (2007) assume-se¹³:

$$(10) \quad V_{ij} = \delta_j + \mu_{ij} = \underbrace{\sum_{k=1}^K \bar{\beta}_k x_{jk}}_{\delta_j} + \xi_j + \underbrace{\sum_{k=1}^K \sigma_k v_{ik} x_{jk} - \overbrace{\exp(\alpha u_i)}^{\alpha_i} p_j}_{\mu_{ij}}$$

Onde V_{ij} é decomposto em dois termos: δ_j , que depende apenas das características do produto j , e μ_{ij} , que é uma interação entre idiosincrasias do consumidor i e características do produto j . A decomposição se completa a partir das suposições de que $\beta_{ik} \sim N(\bar{\beta}_k, \sigma_k)$ e $\alpha_i \sim \log N(0, \alpha)$, esta última é uma conveniência útil na construção da rotina computacional que será discutida adiante e não afeta a estrutura teórica do modelo.

Novamente, ao assumir que ε_{ij} é i.i.d. com distribuição probabilidade de Valor Extremo do Tipo I é possível obter uma forma analítica para a probabilidade do consumidor i escolher o produto j , dada por¹⁴:

$$(11) \quad \Pr_{ij} = \frac{\exp(V_{ij})}{1 + \sum_{r=1}^n \exp(V_{ir})}$$

O resultado (11) difere-se do resultado (6) em um aspecto fundamental: o lado direito da equação acima é indexado para o consumidor i . Portanto, a probabilidade não condicional de consumo não é idêntica à probabilidade condicional. Na realidade, para uma dada função de probabilidade cumulativa conjunta P , \Pr_j será dada pela esperança matemática:

$$(12) \quad \Pr_j = \int_{v,u} \frac{\exp(V_{ij})}{1 + \sum_{r=1}^n \exp(V_{ir})} dP(v,u)$$

¹³ Ver Nevo (2000a) e (2000b) para mais detalhes e referências.

¹⁴ A dedução de (6) e de (11) são idênticas – ver ANEXO I.

Ao fazer $\text{Pr}_j = s_j$ a equação acima representará um sistema de equações de demanda por produtos diferenciados, pois depende dos preços (e das características) de todos os produtos do mercado. No entanto o alto grau de não linearidade em ξ impede o uso imediato de técnicas econométricas tradicionais. O que demanda a construção de uma complexa rotina computacional, resumida nos quatro passos apresentados na seqüência¹⁵.

PASSO 1. Assumem-se valores iniciais para α e $\sigma' s$; geram-se os números u e v 's aleatoriamente de uma $N(0,1)$ para C consumidores simulados; e resolve-se um estimador para a forma (12). Tendo isso em mãos, Berry, Levinsohn e Pakes (1995) provam que a seguinte contração é válida:

$$(13) \quad \delta_{J \times 1}^{t+1} \leftarrow \delta_{J \times 1}^t + \ln s_{J \times 1} - \ln s_{J \times 1}^t$$

Onde δ^t é um vetor coluna de tamanho J gerado na iteração t e s é um vetor coluna de tamanho J da parcela de mercado observada nos dados (e com sobrescrito t é o resultado simulado dado pelo estimador de (12) na referida iteração). Esse procedimento é denominado de *looping interno*.

Quando o *looping interno* para, por um determinado critério de convergência¹⁶, o δ^t resultante é utilizado na iteração computacional do próximo procedimento, denominado de *looping externo*.

PASSO 2. Tendo um δ^t em mãos, observa-se por (10) que ele é linear nos parâmetros e em ξ , logo se pode programar um método econométrico tradicional¹⁷ para encontrar uma estimativa do vetor $\bar{\beta}$. Por conseqüência, encontra-se uma estimativa de ξ (ou seja, isola-se ξ em função dos parâmetros do modelo dados na iteração do *looping externo*).

PASSO 3. Computa-se o conjunto de instrumentos propostos por Pakes (1994) e os associa a estimativa de ξ formando condições de momento para a iteração do *looping externo*.

PASSO 4. Computa-se uma função objetivo GMM¹⁸ para a iteração do *looping externo* buscando minimizá-la em relação a α e $\sigma' s$. Verifica-se se houve convergência dessa função usando um algoritmo ausente de derivadas (o mais popular é técnica simplex Nelder-Mead, 1965). Então, ou a rotina é concluída porque houve convergência ou a rotina gera um novo conjunto de α e $\sigma' s$ para recomeçar o passo 1.

Uma vez que houve convergência, tem-se em mãos estimativas de α , $\bar{\beta}' s$ e $\sigma' s$ para (10), que são usadas para calcular uma (estimativa da) matriz de derivadas onde na diagonal principal e fora da diagonal tem-se, respectivamente:

¹⁵ A rotina computacional construída para os quatro passos é apresentada no ANEXO II.

¹⁶ Usou-se $\|\delta_{J \times 1}^{t+1} - \delta_{J \times 1}^t\| < 10^{-5}$.

¹⁷ Usou-se o estimador (17.50) para painel não balanceado da p. 579 de Wooldridge (2002).

¹⁸ Usou-se a função proposta por Berry, Levinsohn e Pakes (1995).

$$(14) \left\{ \begin{array}{l} \frac{\partial s_j}{\partial p_j} = \int_{v,u} \left(\left[\frac{\exp(V_{ij})}{1 + \sum_{r=1}^n \exp(V_{ir})} \right] \times \left[1 - \frac{\exp(V_{ij})}{1 + \sum_{r=1}^n \exp(V_{ir})} \right] \times \frac{\partial \mu_{ij}}{\partial p_j} \right) dP(v,u) \quad (NA \text{ DIAGONAL}) \\ \frac{\partial s_j}{\partial p_q} = \int_{v,u} \left(\left[\frac{\exp(V_{ij})}{1 + \sum_{r=1}^n \exp(V_{ir})} \right] \times \left[\frac{\exp(V_{iq})}{1 + \sum_{r=1}^n \exp(V_{ir})} \right] \times \frac{\partial \mu_{ij}}{\partial p_q} \right) dP(v,u) \quad (FORA \text{ DA DIAGONAL}) \end{array} \right.$$

Então (14) é usado na construção da matriz Δ da fórmula (3). Esses resultados implicarão que a matriz de elasticidades não possuirá valores cruzados idênticos, como os oriundos do modelo LOGIT simples, ou forçosamente diferentes, como acontece no *Nested* LOGIT. Espera-se, assim, encontrar valores mais próximos dos verdadeiros valores.

Em resumo, podemos listar como vantagens no uso do *Mixed* LOGIT: 1) número reduzido de parâmetros a se trabalhar, ou seja, a exemplo dos outros modelos de escolha discreta o *Mixed* LOGIT pode lidar com mercados caracterizados pela presença muitas variedades; 2) o modelo não impõe *a priori* um padrão de substituição entre os produtos; 3) o *Mixed* LOGIT gera medidas mais plausíveis de poder de mercado; e 4) o modelo é bastante flexível em relação a formas funcionais.

Já as desvantagens podem ser listadas por: 1) impossibilidade do uso de técnicas tradicionais; 2) o modelo não está disponível em pacotes econométricos, logo exige do pesquisador o conhecimento de técnicas de programação; 4) é um modelo de implantação complexa e possui uma carga computacional elevada, podendo demandar dias para se obter convergências; e 5) ainda é uma técnica em evolução¹⁹ (ver Berry e Pakes, 2007).

1.3.3 A opção externa

Um aspecto importante nas duas sub-seções anteriores é o tamanho do mercado e, por consequência, a parcela de mercado da opção externa (s_0). Quando o tamanho do mercado é facilmente identificado, digamos M , de posse das quantidades de venda (q_j 's) se calcula de forma direta as fatias de mercado ($s_j = q_j / M$) e a parcela da opção externa ($s_0 = 1 - \sum_{j \neq 0} s_j$). Todavia, nem sempre M é um valor evidente. Para abordar essa questão, usou-se uma adaptação da fórmula da elasticidade agregada derivada por DeSouza (2008), exposta abaixo:

$$(15) \quad \eta_I(\alpha, p, \delta) = \frac{E_v \left[\alpha_i \bar{P}_i(\alpha, p, \delta, v_i) s_{i0}(\alpha, p, \delta, v_i) \right]}{1 - s_0}$$

¹⁹ Por exemplo, ainda não há uma maneira eficiente de se calcular dos desvios-padrão das estimativas de α e $\sigma^1 s$, e conseqüentemente suas estatísticas t .

Onde η_i é a elasticidade agregada do mercado, \bar{P}_i é o preço médio das opções internas ponderado pelas probabilidades de compra de um consumidor i simulado e s_{i0} é a probabilidade de que um consumidor i escolha não demandar as opções internas.

Tendo uma informação externa para η_i , é possível adaptar (15) para se computar s_0 dentro do passo 1 da rotina computacional²⁰. A vantagem desse procedimento é que o pesquisador pode ter mais confiança no valor da elasticidade agregada do que no valor do tamanho do mercado. Assim, tendo um estimador \hat{s}_0 em mãos, computa-se o seguinte estimador do tamanho do mercado:

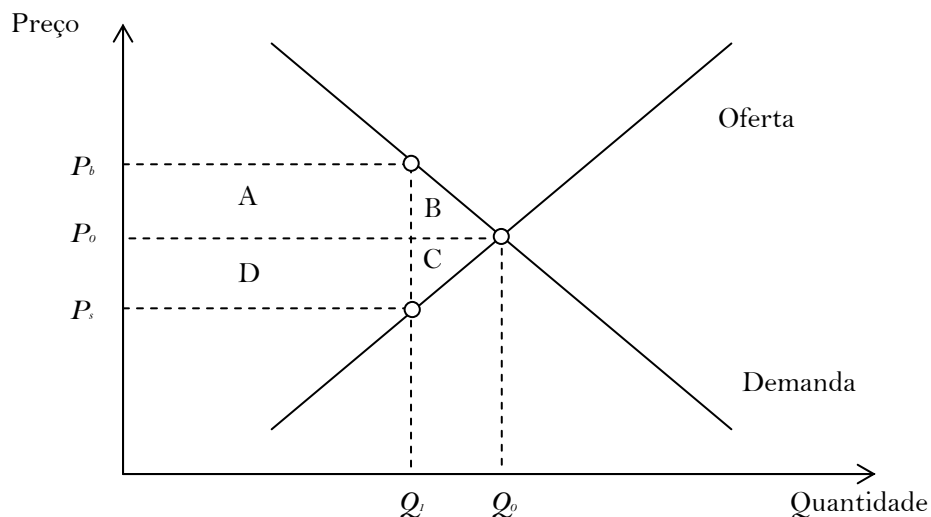
$$(16) \hat{M} = \sum_{j=1}^J \frac{q_j}{1 - \hat{s}_0}$$

DeNegri (1998) estima para o mercado brasileiro de automóveis um η_i em torno de 0,6 e 0,7, em valor absoluto. Neste trabalho será usado o valor de 0,6.

1.3.4 As medidas de bem-estar

Até aqui se discutiu como obter as estimativas dos parâmetros da demanda e dos custos marginais, para dar base para a simulação de um cenário de isenção completa de tributos sobre as vendas de automóveis ($\tau_j = 0 ; \forall j$). Esta sub-seção tratará de como mensurar o ganho de bem-estar decorrente da isenção tributária.

Para tanto, o diagrama abaixo apresenta a análise clássica presente nos livros de teoria microeconômica. Mas é importante salientar que o diagrama serve apenas para ilustrar a intuição dos resultados, pois num mercado com produtos diferenciados é impossível obter uma visualização gráfica.



O diagrama apresenta as curvas de oferta e demanda agregadas do mercado. Onde: P_0 e Q_0 são o preço e a quantidade vendida de automóveis na ausência de tributação,

²⁰ Ver a descrição da rotina computacional no ANEXO II.

respectivamente; P_b é o preço que o consumidor paga com tributação; P_s é o preço que o produtor recebe, dada uma tributação; e, Q_j é a quantidade vendida de automóveis com tributação. Assim, a tributação é: $t = P_b - P_s$.

Observe que a tributação leva a uma perda de excedente do consumidor representada pela soma das áreas A e B, a uma perda de excedente do produtor representada pela soma das áreas B e C, a uma receita tributária representada pela soma das áreas A e D, e ao “peso morto” representado pela soma das áreas B e C (pois é uma área que não é agregada nem pelo consumidor, nem pelo produtor e nem pelo governo).

Seguindo Nevo (2001), para estimar o ganho de bem-estar do consumidor gerado pela desoneração tributária do mercado (a soma das áreas A e B), usa-se a fórmula da variação compensatória de modelos de escolha discreta derivada por Small e Rosen (1981)²¹. Para um determinado consumidor i o resultado é dado por:

$$(17) VC_i = \frac{1}{\alpha_i} \times \ln \left(\frac{1 + \sum_{j=1}^J \exp(\delta_j + \mu_{ij}^{pós})}{1 + \sum_{j=1}^J \exp(\delta_j + \mu_{ij}^{pré})} \right)$$

Onde os sobrescritos “pós” e “pré” indicam as situações com e sem isenção de tributação, respectivamente. Assim, tendo em mãos um estimador para o tamanho do mercado, estima-se ganho de bem-estar total dos consumidores pelo produto de \hat{M} com a variação compensatória média:

$$(18) \hat{A} + \hat{B} = \hat{M} \times \int_{v,u} VC_i(u,v) dP(v,u)$$

Pela fórmula (1), descrita na seção que tratou do lado da oferta, é simples construir um estimador para a soma das áreas D e C, dado por:

$$(19) \hat{D} + \hat{C} = \sum_{f=1}^F (\pi_f^{pós} - \pi_f^{pré})$$

Tendo estimadores para os ganhos de bem-estar do consumidor e do produtor, estima-se o ganho total de bem-estar (\hat{W}) por:

$$(20) \hat{W} = (\hat{A} + \hat{B}) + (\hat{C} + \hat{D})$$

Por fim, para computar a estimativa do peso morto causado pela tributação faz-se:

$$(21) \text{Peso Morto Estimado} = \hat{W} - \sum_{j=1}^J (\tau_j \times p_j \times q_j)$$

Onde $\tau_j \times p_j$ é a tributação sobre o preço e q_j é a quantidade vendida de cada modelo (antes da isenção tributária).

²¹ É importante lembrar que a variação compensatória difere da medida de variação no excedente do consumidor. Detalhes, por exemplo, em Mas-Colell, Whinston e Green (1995), p. 80.

1.4 A amostra

A base de dados foi elaborada com informações da Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA) e com dados complementares obtidos em revistas especializadas (basicamente a *Quatro Rodas*). No total trabalhou-se com um painel não balanceado²² com 66 modelos e quatro anos.

A construção da base focou o mercado de veículos leves. Ou seja, foram desconsideradas as categorias caminhonetes, picapes, utilitários, *vans* e *jeeps*, bem como alguns modelos de luxo e os esportivos. No total, abrangeu-se mais de 80% do mercado de automóveis em todos os anos analisados.

Foram coletadas informações apenas para carros classificados como nacionais e importados de países do MERCOSUL e do México, conforme classificados pela revista *Quatro Rodas*. Dado que alguns modelos possuem sub-modelos (com diferenças de preços e características, como, por exemplo, potência), optou-se por agregar os sub-modelos em um único modelo representativo²³.

Com o objetivo de evitar problemas referentes a redução do IPI no final de 2008 e considerando que as montadoras costumam iniciar a comercialização dos novos modelos em outubro, optou-se por agregar as informações de outubro a setembro. Ou seja, o primeiro período da amostra é formado por 12 meses agregando as informações de outubro de 2004 a setembro de 2005, doravante será tratado apenas por 2005 (o mesmo ocorre para os outros anos).

A tabela 4, adiante, expõe algumas estatísticas descritivas. Nela se observa que o ano com maior número de modelos presente na amostra foi o de 2007, com 63, e o menor foi o de 2005, com 52.

Tabela 4 - Algumas estatísticas descritivas da amostra					
Variável		2005	2006	2007	2008
Número de modelos		52	55	63	59
	Preço - R\$ 1.000	42,62	40,84	40,20	38,32
	Populares	44,5%	58,2%	56,7%	40,4%
(médias ponderadas pelas fatias de mercado dos modelos)	Flex	73,9%	88,3%	90,4%	92,7%
	Potência - HP	89,79	87,32	85,44	87,11
	100×(HP/Kg)	9,28	9,43	9,24	9,10
	Hatch	70,3%	69,8%	66,7%	67,6%
	Minivan	6,9%	5,6%	4,8%	4,2%
	Sedan	18,1%	20,4%	24,1%	24,5%
	Perua	4,8%	4,2%	4,4%	3,7%

Nota-se também que o preço médio dos modelos, ponderando pelas fatias de mercado e deflacionando pelo IPCA para valores de dezembro de 2008, reduziu-se

²² Não balanceado porque no decorrer do período alguns modelos deixaram de ser fabricados, e outros passaram a ser.

²³ Por exemplo, o Peugeot 206 possui os sub-modelos: *Sensation*, *Presence*, *Allure* e *Moonlight*. Quando possível, agregou-se preço e características por uma média ponderada pelas vendas dos sub-modelos. Ou então, de forma *ad hoc*, usando ponderações condizentes com as matérias das revistas especializadas.

gradualmente ao longo dos anos observados, de R\$ 42,62 mil em 2005 para R\$ 38,32 mil em 2008. Em particular, observou-se que os modelos mais caros tiveram reduções de preço mais acentuadas.

Esse comportamento descendente também se observou na participação dos carros populares (com motor 1.0) de 2006 a 2008, embora tenha ocorrido um aumento entre 2005 e 2006. Como será observado adiante, na seção dos resultados, isto se reflete na maior elasticidade preço da demanda estimada para estes modelos.

Por outro lado, os modelos com motores Flex (ou bi-combustível) avançam ano após ano, sendo que em 2008 já representavam 92,7% do mercado (na amostra). Muitos modelos que possuíam motor apenas a gasolina nos anos iniciais passaram a ser Flex no decorrer do período analisado.

Já a potência média dos motores vem se mantendo relativamente estável em torno dos 87 HPs. O mesmo comportamento ocorre com a razão entre potência e peso, em torno de 9,2 HPs por 100 Kg, uma vez que o peso dos modelos também quase não muda entre os anos.

Outra observação é que a participação dos modelos Hatch vem se mantendo relativamente estável, enquanto a demanda por modelos Sedan vem aumentando em detrimento a modelos Minivan e Perua. O que pode ser reflexo de uma preferência mais forte por carros Sedan, na média.

1.5 Resultados estimados

Nesta seção, primeiro se apresentarão os resultados estimados para a equação (8) (o modelo LOGIT) e suas conseqüentes elasticidades preço e preço-cruzado da demanda. Depois, serão apresentados os resultados estimados para a equação (10) (o modelo *Mixed* LOGIT) e suas conseqüentes elasticidades preço e preço-cruzado da demanda. A apresentação nessa ordem busca ilustrar a importância de se usar a uma técnica mais sofisticada para se obter resultados mais plausíveis. Depois, finalizando a seção, será apresentada a decomposição (4) dos preços dos modelos da amostra, entre custo marginal, tributação marginal e lucro operacional por unidade.

A tabela 5 apresenta os resultados dos parâmetros estimados das equações de demanda LOGIT²⁴, com e sem instrumentalização. Os instrumentos usados foram os propostos por Berry, Levinsohn e Pakes (1995)²⁵. Como variáveis explicativas (no vetor x_j de (5)) usaram-se: uma constante unitária; uma *dummy* para automóvel popular (1 se for o caso); uma *dummy* para automóvel com motor Flex (1 se for o caso); a razão entre potência e peso (em HPs por 100 Kg, uma medida de potência relativa); e um Trend (1 para 2005, ..., 4 para 2008).

²⁴ Assim como na especificação BLP, usou-se o estimador (17.50) em dois estágios de Wooldridge (2002) para os parâmetros das variáveis Flex, Potência/Peso, Trend e Preço. Para a constante e a *dummy* Popular usou-se um terceiro estágio, via OLS, tomando a constante de efeito fixo do passo anterior por variável dependente. Isso foi necessário por conta do referido estimador do painel ser do tipo *within*, e a *dummy* Popular não muda de valor no tempo. Para s_0 usou-se o valor estimado no modelo BLP que será apresentado adiante.

²⁵ Existem duas técnicas mais comuns de construção de variáveis instrumentais para esse caso: VI's construídas a partir das características dos produtos, propostas por Berry, Levinsohn e Pakes (1995), em decorrência do trabalho de Pakes (1994); e VI's construídas a partir da observação de preços do mesmo produto em mercados geograficamente distintos (ver Hausman *et. al*, 1994).

Tabela 5 – Parâmetros estimados das equações LOGIT – Especificação (8)							
Especificação	LOGIT <u>com</u> instrumentos			LOGIT <u>sem</u> instrumentos			
Variável	Estimativa	Desvio Padrão	Estatística t	Estimativa	Desvio Padrão	Estatística t	
Constante	0,640	0,182	3,523	0,606	0,180	3,356	
$\bar{\beta}$	Popular	-0,437	0,494	-0,886	-0,400	0,490	-0,817
	Flex	0,323	0,129	2,500	0,375	0,131	2,866
	100×(HP/Kg)	-0,176	0,034	-5,107	-0,203	0,033	-6,075
Trend	-0,261	0,037	-6,956	-0,245	0,038	-6,440	
α	Preço	-0,068	0,007	-10,005	-0,064	0,007	-9,576

Primeiro observe que a ausência de instrumentos tende a subestimar (em valor absoluto) o parâmetro dos preços, que se mostrou estatisticamente significativo e com o sinal esperado nas duas regressões.

O parâmetro estimado para a *dummy* Popular não se mostrou estatisticamente significativo, ao contrário do para a *dummy* Flex. Este último também apresentou o sinal esperado, ou seja, automóveis com motor bi-combustível levam a maiores níveis de utilidade (na média).

Quanto, ao parâmetro estimado para a razão entre potência e peso, apesar de ter apresentado significância estatística, apresentou o sinal inverso ao que se esperava. Intuitivamente, é de se esperar que um modelo com maior potência relativa leve a maiores níveis de utilidade (na média).

O parâmetro estimado para a variável Trend se mostrou estatisticamente significativo e sinal negativo. Note que este parâmetro pode representar uma série de mudanças na utilidade ao longo do tempo, reflexo de mudanças de renda, gosto, intensidade da propaganda etc. Portanto, sua interpretação não é direta e nem trivial, sendo aqui considerada apenas como um fator de ajuste entre os anos.

A tabela 6, adiante, mostra os valores das elasticidades LOGIT estimadas para 2008. Para tanto, usou-se a estimativa $\hat{\alpha} = 0,068$.

Dois pontos devem ser observados na tabela 6 (e depois contrastados com os da tabela 9, que adiante apresentará os resultados estimados via *Mixed* LOGIT). O primeiro é a já citada presença da propriedade da IAI. Por exemplo, tome dois modelos como o Honda Civic e o Fiat Siena, com fatias de mercado semelhantes. Mesmo se esperando que um aumento de preço no Citroen C4 Pallas gere um impacto maior sobre a demanda do Civic, a estimação por meio do modelo LOGIT vai oferecer padrões de substituição semelhantes entre o Civic e o Siena (elasticidade-preço cruzada de 0,01 em ambos). O que é pouco plausível, uma vez que o C4 e o Civic são do mesmo nicho de mercado, e este nicho não é o do Siena.

O segundo ponto a se observar é que a forma funcional LOGIT também impõe que as elasticidades dependem diretamente do preço dos produtos. Desta forma, preços elevados levam a elasticidades também elevadas (o que pode ser pouco razoável). Note que a elasticidade-preço da demanda estimada de um modelo popular como Uno Mille, por exemplo, foi de 1,62 (em valor absoluto). Intuitivamente, era de se esperar um número maior que o de um modelo de luxo como o C4 Pallas, por exemplo, com elasticidade-preço estimada em 3,48.

Tabela 6 – Uma amostra das elasticidades-preço e preço-cruzado pela especificação LOGIT – 2008

Ka 1.0	1,71	0,00	Clio h.	0,01	0,01	Celta	Fox 1.0	Siena	Classic	Palio	Gol 1.0	Parati	Palio W.	206 SW	Space	Idea	Meriva	Golf	Stilo	Fit	Corolla	C4	Civic	Zafira
Clio hatch 1.0	0,00	1,86	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Uno Mille	0,00	0,00	1,62	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Prisma	0,00	0,00	0,01	2,25	0,01	0,01	0,01	0,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Celta 1.0	0,00	0,00	0,01	0,01	1,90	0,01	0,01	0,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Fox 1.0	0,00	0,00	0,01	0,01	0,01	2,26	0,01	0,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Siena 1.0	0,00	0,00	0,01	0,01	0,01	0,01	2,15	0,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Classic	0,00	0,00	0,01	0,01	0,01	0,01	0,01	2,01	0,01	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Palio 1.0	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	1,94	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Gol 1.0	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	1,98	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Parati	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	2,90	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Palio Weekend	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	3,14	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
206 SW	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	3,38	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Spacefox	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	3,36	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Idea	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	3,42	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Meriva	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	3,25	0,00	0,00	0,00	0,00	0,01	0,01	0,00
Golf	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	3,49	0,00	0,00	0,00	0,01	0,01	0,00
Stilo	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	3,62	0,00	0,00	0,01	0,01	0,00
Fit	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	3,88	0,00	0,01	0,01	0,00
Corolla	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	3,85	0,01	0,01	0,00
C4 Pallas	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	3,48	0,01	0,00
Civic	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	4,53	0,00
Zafira	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	4,69

Nota: Os elementos da diagonal são as elasticidades-preço em valor absoluto. Fora da diagonal lê-se a tabela como: aumento do preço do modelo da coluna para o aumento de demanda do modelo da linha.

Mesmo ao se recorrer a solução parcial de agrupar os modelos em nichos e estimar um LOGIT hierárquico, o problema da IAI ainda estaria presente dentro dos nichos. Assim, a fórmula (3) continuaria a carregar valores pouco plausíveis, comprometendo o sucesso do exercício proposto de se simular o impacto da desoneração tributária sobre o mercado. Por isso é necessário recorrer a uma estrutura mais sofisticada como o *Mixed* LOGIT.

A tabela 7, adiante, apresenta os resultados para a especificação (10), para C igual a 1.000 e a 100 consumidores²⁶. Note primeiro que o parâmetro $\bar{\beta}$ estimado para a constante mostrou o mesmo sinal do estimado nos LOGIT. Todos se mostraram estatisticamente significantes, embora os valores do *Mixed* LOGIT tenham se mostrado cerca de duas ou três vezes maiores. Já a estimativa do parâmetro σ se mostra significativa apenas a níveis de confiança mais modestos²⁷.

Tabela 7 – Resultados dos parâmetros estimados das equações da especificação (10) do modelo <i>Mixed</i> LOGIT							
Especificação	1.000 Consumidores			100 Consumidores			
Variável	Estimativa	Desvio Padrão	Estatística t	Estimativa	Desvio Padrão	Estatística t	
$\bar{\beta}$	Constante	1,432	0,276	5,196	2,199	0,349	6,298
	Popular	0,633	0,749	0,846	-0,636	0,948	-0,671
	Flex	1,442	0,267	5,396	1,931	0,366	5,269
	100×(HP/Kg)	0,729	0,029	25,398	1,727	0,039	43,878
	Trend	-3,598	0,078	-46,065	-0,187	0,107	-1,745
σ	Constante	1,987	1,305	1,523	1,849	1,321	1,400
	Popular	1,693	1,874	0,903	1,736	0,373	4,658
	Flex	1,541	0,566	2,722	1,622	0,234	6,936
	100×(HP/Kg)	0,613	0,041	14,975	0,699	4,526	0,154
	Trend	2,204	0,547	4,031	2,217	0,550	4,030
α	Preço	2,079	0,457	4,544	2,174	0,283	7,675

Os valores estimados do $\bar{\beta}$ para a *dummy* Popular invertem os sinais entre as especificações *Mixed* LOGIT com $C = 100$ e 1.000, embora em ambas não tenham se mostrado estatisticamente significantes (nem mesmo a do σ para $C = 1.000$).

A estimativa do parâmetro da *dummy* Flex apresentou os sinais esperados e significância estatística, em todas as especificações. Note que, considerando as propriedades da distribuição normal, $\beta_i \in (-0,099 ; 2,983)$ em 68,3% dos 1.000 consumidores simulados. Ou seja, apesar da maioria dos consumidores simulados auferirem números positivos para característica de motor bi-combustível, alguns ainda dão pesos maiores a motores tradicionais.

²⁶ A semente aleatória usada para gerar u e v 's foi 7654321. No primeiro momento, os *starts* do algoritmo Nelder-Mead para α e σ 's foram a unidade e para δ 's um vetor de zeros. Depois, como *starts*, usou-se valores aleatórios de distribuições uniformes entre 0 e 2 para assegurar que o algoritmo levaria aos mesmos resultados de convergência da função objetivo.

²⁷ Para se computar a matriz de variância-covariância GMM, que gera as estatísticas t dos σ estimados, usou-se um gradiente numérico, uma vez que não existe uma forma funcional analítica para a derivada das condições de momento em relação aos parâmetros.

O parâmetro estimado para a razão entre potência e peso, que se mostrou com o sinal inverso ao esperado nas especificações LOGIT, mostrou o sinal esperado nas especificações *Mixed* LOGIT. Mais uma vez considerando as propriedades da distribuição normal, $\beta_i \in (0,116 ; 1,342)$ para a variável em 68,3% dos consumidores simulados.

O parâmetro estimado para o Trend, assim como nas especificações LOGIT, apresentou sinal negativo nas especificações *Mixed* LOGIT. Também se observa nessa variável o maior desvio padrão estimado entre os consumidores, $\sigma_i = 2,204$.

Quanto ao α estimado, mostrou-se estatisticamente significativa na especificação *Mixed* LOGIT tanto para $C = 100$ e como para $C = 1.000$. Mais ainda, para $C = 1.000$ uma estimativa de α é $\exp(2,079 \times u_i)$. A título de ilustração, observe que se $u_i = 1$ o parâmetro dos preços será de 7,997 e se $u_i = -1$ o parâmetro será de 0,125.

A tabela 8 apresenta a quantidade vendida, a parcela de mercado estimada da opção externa e o tamanho do mercado (em mil unidades) para os anos trabalhados, usando a fórmula (16) e as estimativas *Mixed* LOGIT com $C = 1.000$.

Tabela 8 – Quantidade de automóveis vendida e opção externa e tamanho do mercado estimados – mil unidades				
Parâmetro	2005	2006	2007	2008
Quantidade Vendida	1.175	1.435	1.803	2.216
\hat{s}_0	72,86%	79,21%	84,60%	88,62%
\hat{M}	4.331	6.902	11.704	19.477

Observe que o tamanho estimado do mercado para 2008 é de aproximadamente 19,5 milhões de unidades. Este valor será usado adiante nos exercícios de simulação.

Salienta-se que a estratégia usada aqui para estimar \hat{M} é diferente da usada por Fiúza (2002), que computou a opção externa a partir de uma estimação do tamanho do mercado em relação à renda das famílias que poderiam comprar um carro novo no Brasil: $\hat{M} \approx 12,5$ milhões de famílias, para 1996.

Outra referência para esta estimação é que Berry, Levinsohn e Pakes (1995) apresentam uma estimativa do tamanho do mercado norte-americano de 66 milhões de unidades para 1990. Miller (2000) aponta que as vendas de veículos leves nos EUA em 1990 foram de 9 milhões de unidades. Combinando estes dois números chega-se a estimativa de $s_0 = 86\%$.

A tabela 9 apresenta uma amostra das elasticidades-preço e preço-cruzado pela especificação *Mixed* LOGIT no ano de 2008 e as estimativas para $C = 1.000$. Note que agora a propriedade da IAI não é observada. Agora, um aumento de 1% no preço do Citroen C4 Pallas leva a uma aumento de demanda de 0,19% no Honda Civic e 0,01% no Fiat Siena.

Note também que a tabela 9 está organizada por ordem decrescente de valor absoluto da elasticidade-preço da demanda, da esquerda para a direita e de cima para baixo. Na amostra, o modelo com a maior elasticidade-preço estimada (em valor absoluto) foi o Ford Ka 1.0, com 7,53. E o modelo com a menor foi a Minivan Chevrolet Zafira, com 2,08. Logo, os resultados *Mixed* LOGIT não fazem preços elevados implicarem em elasticidades também elevadas.

Tabela 9 – Uma amostra das elasticidades-preço e preço-cruzado pela especificação *Mixed LOGIT* – ano de 2008

	Ka 1.0	Clio h.	Uno	Prisma	Celta	Fox 1.0	Siena	Classic	Pallo	Gol 1.0	Parati	Pallo W.	206 SW	Space	Idea	Meriva	Golf	Stilo	Fit	Corolla	C4	Civic	Zafira
Ka 1.0	7,53	0,35	0,94	0,09	0,27	0,08	0,08	0,16	0,18	0,18	0,03	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00
Clio hatch 1.0	0,08	6,32	0,11	0,03	0,05	0,02	0,01	0,03	0,03	0,04	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Uno Mille	2,56	1,24	5,87	0,26	0,91	0,20	0,21	0,47	0,54	0,52	0,06	0,04	0,03	0,03	0,03	0,03	0,02	0,02	0,02	0,01	0,01	0,01	0,01
Prisma	0,14	0,18	0,15	5,27	0,14	0,20	0,13	0,08	0,07	0,09	0,14	0,15	0,10	0,09	0,09	0,08	0,04	0,05	0,03	0,05	0,03	0,03	0,02
Celta 1.0	0,90	0,74	1,10	0,31	5,08	0,23	0,21	0,45	0,45	0,48	0,12	0,09	0,07	0,07	0,07	0,06	0,05	0,05	0,05	0,03	0,03	0,03	0,02
Fox 1.0	0,16	0,16	0,16	0,28	0,15	4,65	0,32	0,13	0,13	0,13	0,13	0,09	0,08	0,08	0,08	0,07	0,08	0,06	0,09	0,03	0,02	0,02	0,03
Siena 1.0	0,14	0,10	0,13	0,14	0,11	0,25	4,62	0,12	0,14	0,12	0,07	0,04	0,04	0,04	0,04	0,04	0,06	0,03	0,08	0,01	0,01	0,01	0,02
Classic	0,20	0,16	0,21	0,06	0,16	0,07	0,08	4,48	0,17	0,16	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,03	0,01	0,01	0,01	0,01
Pallo 1.0	0,81	0,58	0,87	0,20	0,59	0,27	0,36	0,62	4,11	0,60	0,10	0,06	0,06	0,07	0,06	0,06	0,07	0,05	0,10	0,02	0,02	0,02	0,03
Gol 1.0	1,03	0,87	1,10	0,34	0,84	0,36	0,40	0,76	0,80	4,01	0,17	0,12	0,10	0,11	0,10	0,10	0,10	0,08	0,12	0,04	0,04	0,03	0,05
Parati	0,02	0,02	0,01	0,06	0,02	0,04	0,03	0,02	0,02	0,02	3,67	0,05	0,04	0,04	0,04	0,03	0,02	0,03	0,02	0,02	0,02	0,02	0,01
Pallo Weekend	0,02	0,03	0,01	0,09	0,02	0,04	0,02	0,02	0,01	0,02	0,07	3,52	0,07	0,06	0,06	0,05	0,03	0,04	0,02	0,06	0,04	0,04	0,02
206 SW	0,00	0,01	0,00	0,02	0,01	0,01	0,01	0,00	0,00	0,00	0,01	0,02	3,20	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Spacefox	0,01	0,02	0,01	0,06	0,02	0,04	0,03	0,02	0,02	0,02	0,06	0,06	0,05	3,09	0,05	0,05	0,04	0,04	0,03	0,04	0,04	0,04	0,03
Idea	0,02	0,03	0,01	0,08	0,03	0,04	0,03	0,02	0,02	0,02	0,07	0,08	0,07	0,06	3,08	0,06	0,04	0,05	0,03	0,06	0,05	0,05	0,03
Meriva	0,01	0,02	0,01	0,05	0,02	0,04	0,02	0,02	0,01	0,02	0,05	0,06	0,05	0,05	0,05	2,95	0,04	0,04	0,03	0,04	0,04	0,03	0,03
Golf	0,01	0,01	0,01	0,02	0,01	0,03	0,03	0,01	0,01	0,01	0,03	0,02	0,03	0,03	0,03	0,03	2,64	0,03	0,03	0,02	0,02	0,02	0,02
Stilo	0,01	0,01	0,00	0,03	0,01	0,02	0,02	0,01	0,01	0,01	0,03	0,04	0,03	0,03	0,03	0,03	0,03	2,63	0,02	0,03	0,03	0,03	0,02
Fit	0,02	0,02	0,01	0,04	0,03	0,07	0,08	0,04	0,04	0,03	0,06	0,04	0,05	0,05	0,05	0,05	0,07	0,05	2,61	0,03	0,03	0,02	0,04
Corolla	0,01	0,03	0,01	0,07	0,02	0,03	0,01	0,01	0,01	0,02	0,08	0,13	0,10	0,09	0,10	0,08	0,05	0,08	0,03	2,58	0,10	0,12	0,05
C4 Pallas	0,00	0,01	0,00	0,02	0,01	0,01	0,01	0,01	0,00	0,01	0,03	0,04	0,04	0,03	0,04	0,03	0,02	0,03	0,01	0,05	2,27	0,05	0,03
Civic	0,01	0,04	0,01	0,09	0,03	0,04	0,02	0,02	0,01	0,02	0,12	0,20	0,16	0,14	0,16	0,14	0,08	0,13	0,05	0,24	0,19	2,26	0,10
Zafira	0,00	0,00	0,00	0,01	0,00	0,01	0,01	0,01	0,00	0,01	0,01	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	2,08

Nota: Os elementos da diagonal são as elasticidades-preço em valor absoluto. Fora da diagonal lê-se a tabela como: aumento do preço do modelo da coluna leva ao aumento de demanda do modelo da linha.

Observe que o bloco superior esquerdo da tabela 9 concentra os modelos populares e o bloco inferior direito os modelos de luxo. Dos modelos populares, como era de se esperar, o VW Gol 1.0 e o Fiat Pálio 1.0 apresentaram as menores elasticidades-preço, 4,01 e 4,11, respectivamente, em valor absoluto.

Existe ainda uma categoria intermediária formada pelos modelos perua VW Parati, Fiat Palio Weekend, Pegout 206 SW e VW Spacefox com elasticidades entre 3,67 e 3,09. Uma outra categoria pode ser formada pelas Minivans que não são de luxo, a Fiat Idea e a Chevrolet Meriva, com elasticidades de 3,08 e 2,95, respectivamente. E ainda um bloco com elasticidades menores de 2,64 formado pelos modelos VW Golf, Fiat Stilo e Honda Fit e, por fim, um bloco formado pelos modelos de luxo Toyota Corolla, C4 Pallas e Honda Civic.

No que tange as elasticidades cruzadas, os valores que mais se destacam são os ligados ao Ford Ka 1.0, Renault Clio Hatch 1.0 e Uno Mille, pois são os que mais transmitem demanda por aumentos próprios no preço. Por exemplo, um aumento de 1% no preço do Ka leva a um aumento de 2,56% na demanda do Uno, de 1,03% na do Gol 1.0 e de 0,81% na do Pálio 1.0, *ceteris paribus*. Na outra ponta da tabela, pode-se observar que o Honda Civic é o que menos transmite demanda e o que mais recebe entre os modelos de luxo.

A tabela 10, adiante, apresenta as estimativas dos *markups*, do custo marginal, imposto marginal e do lucro variável por modelo e fabricante da amostra, para o ano de 2008 (trata-se da decomposição descrita pela fórmula (4), usando os resultados estimados para o *Mixed LOGIT* com $C = 1.000$).

O *markup* médio estimado (ponderado pelas fatias de mercado) foi de 21,08%. Modelos populares, como o Ka, por exemplo, com 10,2%, apresentam valores menores que os modelos mais sofisticados como o Renault Scenic, por exemplo, com 32,3%.

O custo marginal, como era de se esperar, varia conforme o preço do modelo. Modelos mais caros apresentam maiores custos marginais estimados. O Honda Civic e o Corolla Fielder foram os modelos que apresentaram os maiores valores estimados, R\$ 29,42 mil e R\$ 29,14 mil, respectivamente. Por outro lado, o Uno Mille apresentou o menor valor: R\$ 13,51 mil.

Para se computar os valores da tributação marginal tomou-se como base os levantamentos da ANFAVEA (2006 e 2009) que computa τ_j (formado por IPI, ICMS, PIS e COFINS) como 27,1% para veículos de 1.000 cc, 30,4% para veículos com mais de 1.000 cc e menos de 2.000 cc a gasolina, 29,2% para veículos com mais de 1.000 cc e menos de 2.000 cc bi-combustível, 36,4% para veículos com 2.000 cc ou mais a gasolina e 33,1% para veículos com 2.000 cc ou mais bi-combustível.

Assim, apenas o VW Gol 1.0 teria gerado R\$ 2,0 bilhões de receita tributária para o governo em 2008, foi o maior arrecadador seguindo pelo Honda Civic e o Fiat Pálio 1.0, ambos com R\$ 1,5 bilhão. No total, a massa de tributação estimada para 2008

$(\sum_{j=1}^J (\tau_j \times p_j \times q_j))$ foi da ordem de R\$ 24,8 bilhões²⁸.

O modelo com o maior lucro variável estimado foi o Honda Civic, com R\$ 1,61 bilhão. Uma consequência direta de ser um veículo de luxo que é o mais caro e o mais vendido de sua categoria. Em seguida surgem o Gol 1.0 (R\$ 1,49 bilhão) e o Pálio 1.0 (R\$ 1,22 bilhão). No total, a massa de lucro variável estimada para a indústria (somando os 59 modelos da amostra) foi de R\$ 19,0 bilhões.

²⁸ Como era de se esperar, inferior aos 39,4 bilhões totais descritos na introdução, uma vez que a amostra não considerou todos os veículos produzidos em 2008.

Tabela 10 - Estimativas de elasticidades-preço, *markups*, custo marginal, imposto marginal e lucro variável por modelo e fabricante da amostra para o ano de 2008

Modelo	Elasti- cidade preço	<i>Mark</i> <i>-up</i> (%)	Decomposição (4) R\$ mil por unidade				Quanti- dade Vendida (1.000 unid.)	Tributa- ção Total (R\$ milhões)	Lucro Variável Total (R\$ milhões)
			Preço	Custo Marginal	Tributo Marginal	Lucro Variável			
Doblo	-2,61	29,8	56,63	22,76	16,99	16,89	9	153	152
Idea	-3,08	25,5	50,35	22,40	15,11	12,84	32	484	411
Palio > 1.0	-5,91	16,3	30,74	16,52	9,22	5,00	33	304	165
Palio 1.0	-4,11	21,7	28,79	14,77	7,77	6,24	195	1.515	1.217
Palio Weekend	-3,52	22,6	46,26	21,93	13,88	10,46	25	347	262
Punto	-3,01	26,9	45,39	19,58	13,62	12,20	45	613	549
Siena > 1.0	-3,34	24,7	42,55	19,26	12,76	10,52	41	523	431
Siena 1.0	-4,62	20,2	31,75	16,76	8,57	6,41	68	583	436
Stilo	-2,63	29,4	57,16	23,23	17,15	16,78	18	309	302
Uno Mille	-5,87	16,7	24,02	13,51	6,49	4,02	145	941	583
Fiat							610	5.771	4.507
Fiesta Sedan > 1.0	-4,50	16,7	37,37	19,90	11,21	6,25	21	235	131
Fiesta Sedan 1.0	-4,58	17,7	32,91	18,20	8,88	5,82	23	204	134
Fiesta > 1.0	-4,07	18,4	40,47	20,87	12,14	7,45	11	134	82
Fiesta 1.0	-3,88	20,0	36,16	19,14	9,76	7,25	56	547	406
Focus Hatch	-2,99	24,8	47,37	21,42	14,21	11,74	19	270	223
Focus Sedan	-2,66	27,6	52,37	22,22	15,71	14,44	20	314	289
Ka > 1.0	-6,12	12,6	31,80	18,24	9,54	4,02	4	38	16
Ka 1.0	-7,53	10,2	25,19	15,83	6,80	2,56	50	340	128
Ford							205	2.082	1.409
Astra Hatch	-3,21	23,6	50,92	20,58	18,33	12,01	29	532	348
Astra Sedan	-3,14	24,2	52,76	20,98	18,99	12,78	9	171	115
Celta 1.0	-5,08	16,4	28,11	15,90	7,59	4,61	150	1.139	692
Classic	-4,48	19,0	29,61	15,99	7,99	5,62	52	415	292
Corsa Hatch > 1.0	-4,47	19,6	37,35	18,84	11,21	7,31	27	303	197
Corsa Hatch 1.0	-5,43	16,5	31,31	17,69	8,45	5,17	27	228	140
Corsa Sedan > 1.0	-4,82	18,6	35,70	18,36	10,71	6,63	33	353	219
Meriva	-2,95	26,7	51,39	22,24	15,42	13,73	25	386	343
Prisma	-5,27	17,0	33,23	17,61	9,97	5,65	58	578	328
Vectra Hatch	-2,39	29,4	65,09	22,51	23,43	19,15	14	328	268
Vectra Sedan	-3,75	20,6	66,62	28,89	23,98	13,74	15	360	206
Zafira	-2,08	32,8	71,13	22,17	25,61	23,35	12	307	280
GM							449	5.100	3.428

(continua na próxima página)

(continuação)

Modelo	Elasti- cidade preço	Mark- up (%)	Preço	Custo Margina l	Tributo Margina l	Lucro Variável	Quantidade Vendida	Tributa- ção Total	Lucro Variável Total
Civic	-2,26	31,2	75,87	29,42	22,76	23,69	68	1.548	1.611
Fit	-2,61	27,5	51,23	21,79	15,37	14,07	42	646	591
Honda							111	2.193	2.202
206	-3,65	19,5	39,98	20,19	11,99	7,79	47	564	366
307	-2,58	27,6	56,64	24,02	16,99	15,62	22	374	344
206 SW	-3,20	22,6	49,65	23,56	14,90	11,20	6	89	67
C3	-2,89	24,5	47,48	21,59	14,24	11,64	38	541	442
C4 Pallas	-2,27	28,6	69,02	24,42	24,85	19,76	18	447	356
Picasso	-2,29	31,4	63,29	24,43	18,99	19,87	11	209	219
Peugeot- Citroen							142	2.224	1.794
Clio Sedan > 1.0	-3,87	18,7	45,49	23,33	13,65	8,51	2	27	17
Clio Sedan 1.0	-3,88	19,6	39,76	21,23	10,73	7,80	2	21	16
Clio 1.0	-6,32	12,1	27,34	16,66	7,38	3,30	11	81	36
Kangoo	-3,07	23,5	47,86	22,23	14,36	11,27	1	14	11
Logan > 1.0	-4,40	16,6	36,68	19,59	11,00	6,09	16	176	97
Logan 1.0	-4,69	16,0	29,64	16,89	8,00	4,74	22	176	104
Megane	-2,44	29,1	61,70	25,23	18,51	17,97	9	167	162
Sandero > 1.0	-4,10	17,7	38,09	19,94	11,43	6,73	27	309	182
Sandero 1.0	-5,30	14,5	31,69	18,55	8,56	4,58	13	111	60
Scenic	-2,20	32,3	60,11	22,67	18,03	19,41	4	72	78
Renault							108	1.155	763
Corolla	-2,58	27,2	66,74	28,53	20,02	18,18	40	801	727
Corolla Fielder	-2,46	29,8	72,45	29,14	21,74	21,57	4	87	86
Toyota							44	888	813
Fox > 1.0	-4,34	20,5	37,73	18,68	11,32	7,73	53	600	410
Fox 1.0	-4,65	19,5	33,45	17,89	9,03	6,52	81	731	528
Golf	-2,64	29,6	53,34	21,54	16,00	15,80	19	304	300
Gol > 1.0	-6,29	15,5	30,35	16,54	9,11	4,71	48	437	226
Gol 1.0	-4,01	20,1	29,46	15,60	7,96	5,91	252	2.006	1.489
Parati	-3,67	23,3	42,67	19,91	12,80	9,96	19	243	189
Polo Hatch	-3,45	24,3	45,44	20,76	13,63	11,05	22	300	243
Polo Sedan	-2,95	27,1	51,49	22,09	15,45	13,95	29	448	405
Spacefox	-3,09	26,2	49,44	21,66	14,83	12,95	27	400	350
VW							550	5.470	4.140
Total							2.216	24.883	19.056

1.6 Simulações

A tabela 11 apresenta os resultados estimados para quantidades vendidas, preços, lucro operacional e *markups* no cenário de desoneração tributária completa em 2008, ou seja, $\tau_j = 0$ para todo modelo j (uma análise complementar a essa é exposta em ANEXO III, quando se compara os preços recebidos pelo produtor com e sem tributação).

Para se obter esses resultados, tomou-se como fixo os custos marginais expostos na tabela 10 para resolver o sistema (4) e obterem-se os preços “pós isenção” ($p^{pós}$). Em seguida, computou-se o novo vetor de lucro operacional dos modelos: $p^{pós} - \hat{c}$. E por fim, computou-se o vetor das novas parcelas de mercado: $s^{pós} = \hat{\Delta} \times (p^{pós} - \hat{c})$. Para se computar as quantidades vendidas manteve o tamanho do mercado fixo em 19,5 milhões de unidades.

Com a isenção completa, estima-se um aumento de vendas da ordem de 388 mil unidades, um preço médio (ponderado pelas fatias de mercado) que cai de R\$ 39,06 mil para R\$ 28,78 mil, um ganho de lucro operacional da indústria ($\hat{D} + \hat{C}$) na ordem de R\$ 6,95 bilhões e *markups* que aumentam 11,23% na média.

O aumento de 388 mil unidades demandas é bastante intuitivo, uma vez que se está usando o valor $\eta_I = 0,6$ (que é a elasticidade agregada do mercado estimada por DeNegri, 1998). Como o preço médio caiu em 25,32%, uma “conta de bolso” é que a oferta aumentaria em 15,79%, que sobre a base de 2.216 mil unidades vendidas levariam a 350 mil unidades.

Note na tabela 11 que quatro modelos de veículos tiveram reduções nas vendas na simulação, são eles: Uno Mille, vendendo oito mil unidades a menos; Ford Ka 1.0, menos cinco mil unidades; e Corsa Hatch 1.0 e Fox 1.0, ambos com mil unidades a menos. Um efeito direto das elasticidades cruzadas da tabela 9.

As versões não populares do VW Gol teriam os maiores aumentos de demanda (28 mil unidades), seguido pela versão popular deste modelo (23 mil unidades) e pelos modelos GM Prisma e Astra Hatch (ambos com 20 mil unidades).

O maior ganho de lucro operacional (estimado) seria do Honda Civic (R\$ 752,14 milhões), seguido do Toyota Corolla (R\$ 328,80 milhões) e do Gol 1.0 (R\$ 300,19 milhões).

Uma inspeção na tabela 11 também mostra que a redução de preços faz com que a demanda por automóveis populares cresça menos que as dos não populares. A tabela 12, mais adiante, busca sintetizar esse resultado. Observe que a demanda por automóveis populares aumenta em 38 mil unidades (ou 4%), ao passo que a dos não populares aumenta em 350 mil unidades (ou 26%).

Tabela 11 – Simulações de uma desoneração total sobre as quantidades, preços, lucros operacionais e *markups* para o ano de 2008

Modelo	Quantidade (mil unidades)		Preço (R\$ milhões)		Lucro Operacional (R\$ milhões)		Markup (%)	
	Pós	Pós-Pré	Pós	Pós-Pré	Pós	Pós-Pré	Pós	Pós-Pré
Doblo	11	2	42,10	-14,53	214,00	62,38	45,95	16,13
Idea	37	5	37,53	-12,82	561,88	154,47	40,30	14,80
Palio > 1.0	43	10	21,81	-8,93	226,37	61,72	24,28	8,01
Palio 1.0	204	9	21,84	-6,94	1.441,0	224,47	32,37	10,69
Palio Weekend	28	3	34,49	-11,77	351,85	87,22	36,42	13,82
Punto	58	13	32,89	-12,50	768,98	224,66	40,48	13,61
Siena > 1.0	53	13	30,50	-12,05	600,33	173,53	36,84	12,11
Siena 1.0	72	4	23,99	-7,75	523,24	86,99	30,13	9,94
Stilo	22	4	42,64	-14,51	424,79	123,45	45,52	16,16
Uno Mille	137	-8	17,86	-6,16	595,57	13,15	24,35	7,60
Fiat	665	55			5.708,1	1.212,1		
Fiesta Sed. > 1.0	26	6	26,57	-10,80	176,11	46,03	25,09	8,35
Fiesta Sedan 1.0	29	5	24,10	-8,81	168,17	33,99	24,48	6,79
Fiesta > 1.0	14	2	29,11	-11,36	111,38	29,00	28,30	9,88
Fiesta 1.0	63	6	27,16	-8,99	502,53	93,03	29,53	9,47
Focus Hatch	24	6	34,13	-13,24	310,98	92,21	37,23	12,46
Focus Sedan	27	6	38,01	-14,36	419,27	126,24	41,55	13,97
Ka > 1.0	9	5	20,93	-10,87	25,31	9,01	12,83	0,19
Ka 1.0	46	-5	19,10	-6,09	148,93	19,89	17,13	6,98
Ford	237	32			1.862,7	449,40		
Astra Hatch	49	20	33,46	-17,46	629,96	283,77	38,51	14,92
Astra Sedan	15	6	34,64	-18,12	202,85	91,46	39,41	15,19
Celta 1.0	158	8	21,08	-7,02	818,71	127,02	24,56	8,15
Classic	56	4	22,31	-7,30	352,94	61,83	28,32	9,32
Corsa Hat. > 1.0	30	3	27,38	-9,97	258,23	60,85	31,19	11,62
Corsa Hatch 1.0	26	-1	23,76	-7,55	158,55	18,26	25,56	9,05
Corsa Sed. > 1.0	37	5	25,87	-9,83	278,91	63,40	29,02	10,45
Meriva	28	2	39,48	-11,91	478,01	130,46	43,66	16,94
Prisma	78	20	23,25	-9,97	438,18	112,53	24,26	7,26
Vectra Hatch	27	14	40,11	-24,98	481,10	216,91	43,87	14,46
Vectra Sedan	19	4	49,12	-17,49	385,45	182,15	41,18	20,55
Zafira	27	16	39,99	-31,14	489,51	220,50	44,56	11,73
GM	551	101			4.972,4	1.569,2		

(continua na próxima página)

(continuação)

Modelo	Quantidade		Preço		Lucro Operacional		Markup	
	Pós	Pós-Pré	Pós	Pós-Pré	Pós	Pós-Pré	Pós	Pós-Pré
Civic	83	15	58,00	-17,87	2.371,2	752,14	49,27	18,05
Fit	56	14	37,03	-14,20	858,38	261,37	41,16	13,69
Honda	139	29			3.229,6	1.013,5		
206	65	18	28,16	-11,82	536,41	158,90	29,39	9,12
307	27	5	42,13	-14,51	499,61	147,88	44,34	15,74
206 SW	7	1	37,10	-12,55	97,09	27,09	37,69	14,25
C3	49	11	34,47	-13,01	654,88	195,14	38,63	13,19
C4 Pallas	24	6	49,70	-19,32	605,90	250,99	51,84	22,52
Picasso	14	3	46,35	-16,94	307,74	92,83	48,43	16,19
Peugeot-Citroen	185	44			2.701,6	872,84		
Clio Sedan > 1.0	2	0	34,13	-11,36	23,31	5,14	31,63	12,93
Clio Sedan 1.0	2	0	31,04	-8,71	19,23	3,26	31,62	12,01
Clio 1.0	13	2	20,01	-7,33	44,34	8,05	16,72	4,67
Kangoo	1	0	34,55	-13,30	12,51	3,69	35,65	12,11
Logan > 1.0	22	6	25,87	-10,81	139,93	40,34	24,28	7,69
Logan 1.0	26	4	21,81	-7,83	129,17	24,50	22,53	6,52
Megane	12	3	45,03	-16,67	240,25	72,87	43,98	14,86
Sandero > 1.0	37	10	26,88	-11,21	257,66	74,65	25,83	8,17
Sandero 1.0	14	1	23,70	-7,99	71,09	10,99	21,72	7,27
Scenic	6	2	40,84	-19,27	105,47	32,70	44,50	12,21
Renault	136	28			1.043,0	276		
Corolla	48	8	50,41	-16,33	1.053,4	328,80	43,39	16,15
Corolla Fielder	5	1	54,15	-18,30	123,53	39,12	46,18	16,41
Toyota	53	9			1.177,9	368,92		
Fox > 1.0	65	12	26,98	-10,75	542,05	132,56	30,75	10,26
Fox 1.0	80	-1	25,42	-8,02	603,65	73,60	29,62	10,12
Golf	24	5	39,08	-14,26	423,77	124,87	44,88	15,26
Gol > 1.0	76	28	20,98	-9,38	336,74	111,70	21,15	5,64
Gol 1.0	275	23	22,10	-7,36	1.789,0	300,19	29,42	9,37
Parati	24	5	30,86	-11,81	259,58	68,97	35,48	12,15
Polo Hatch	27	4	33,25	-12,19	334,42	88,35	37,57	13,26
Polo Sedan	35	6	38,12	-13,37	556,74	155,99	42,05	14,95
Spacefox	32	5	36,48	-12,96	475,51	131,06	40,63	14,44
VW	638	88			5.321,5	1.187,3		
Total	2.604	388			26.016	6.948		

Observe na tabela 12 que os preços dos automóveis populares caem, na média, R\$ 6,86 mil (ou 25%), e o dos não populares R\$ 12,16 mil (ou 27%). Em contraponto, o *markup* médio dos automóveis não populares aumenta mais (12,13% contra 9,01%).

Tabela 12 – Quantidade vendida e preços e <i>markups</i> médios, ponderados pelas fatias de mercado, para o ano de 2008 (Pré) e simulação de desoneração tributária (Pós) – análise entre modelos populares e não populares				
Variável		Total	Popular	Não Popular
Quantidade (mil unidades)	Pré	2.216	877	1.339
	Pós	2.604	915	1.689
	Pós-Pré	388 (↑18%)	38 (↑4%)	350 (↑26%)
Preço (R\$ mil)	Pré	38,32	27,92	45,13
	Pós	28,79	21,06	32,97
	Pós-Pré	-9,54 (↓25%)	-6,86 (↓25%)	-12,16 (↓27%)
<i>Markup</i> (%)	Pré	21,08	18,41	22,83
	Pós	32,31	27,42	34,96
	Pós-Pré	11,23	9,01	12,13

A tabela 13 apresenta a mesma análise separando o mercado em modelos Hatch, Sedan, Perua e Minivan. Note que o segmento com maior aumento relativo de demanda é o das Minivans, com 35%. Muito embora a demanda dos modelos Hatch seja a que tem o maior aumento em termos absolutos, 220 mil unidades.

Tabela 13 – Quantidade vendida e preços e <i>markups</i> médios, ponderados pelas fatias de mercado, para o ano de 2008 (Pré) e simulação de desoneração tributária (Pós) – análise entre modelos Hatch, Sedan, Perua e Minivan					
Variável		Hatch	Sedan	Perua	Minivan
Quantidade (mil unidades)	Pré	1,498	545	81	92
	Pós	1,718	666	96	124
	Pós-Pré	220 (↑15%)	121 (↑22%)	15 (↑19%)	32 (↑35%)
Preço (R\$ mil)	Pré	33,91	46,05	47,97	55,68
	Pós	25,49	34,22	35,46	40,02
	Pós-Pré	-8,42 (↓25%)	-11,83 (↓26%)	-12,51 (↓26%)	-15,66 (↓28%)
<i>Markup</i> (%)	Pré	19,93	22,52	24,36	28,19
	Pós	29,81	34,67	38,15	43,35
	Pós-Pré	9,88	12,15	10,79	15,16

A variação relativa de preços médios entre os modelos Hatch, Sedan, Perua e Minivan fica em torno de -26%. Todavia, em termos absolutos, a redução média vai de R\$ 15,66 mil para as Minivans a R\$ 8,42 mil para os modelos Hatch. Movimento semelhante, mas inverso, ocorre com os *markups*, que na média aumentam em 15,16% para as Minivans e 9,88% para os modelos Hatch.

Quanto as medidas de bem-estar, a VC_i média foi calculada em R\$ 1,28 mil. Seguindo a fórmula (16) com $\hat{M} = 19,5$ milhões de famílias chega-se a $\hat{A} + \hat{B}$ na ordem de R\$ 24,92 bilhões. Então, pela fórmula (20), o ganho total de bem-estar da desoneração tributária é de R\$ 31,87 bilhões. E o peso morto, dado pela fórmula (21), é de R\$ 7,04 bilhões. A tabela 14 sintetiza estes resultados.

Tabela 14 – Síntese dos resultados da simulação de desoneração tributária				
Ganho de Bem-Estar (\hat{W})	=	Parcela do Consumidor ($\hat{A} + \hat{B}$)	+	Parcela do Produtor ($\hat{D} + \hat{C}$)
R\$ 31,87 bilhões		R\$ 24,92 bilhões (78,2%)		R\$ 6,95 bilhões (21,8%)
Peso Morto Estimado = R\$ 7,04 bilhões ($\hat{B} + \hat{C}$)				

Logo, 78,2% da conta tributária é paga pelo consumidor e 21,8% é paga pelo produtor.

1.7 Conclusões

Como apresentado no início desse ensaio, em 2008 a soma da arrecadação de IPI, ICMS, PIS e COFINS da venda de veículos chegou a R\$ 39,4 bilhões. Sabe-se que a tributação altera o equilíbrio entre demanda e oferta de forma distorcida. Os consumidores são prejudicados porque pagam um preço maior do que o que pagariam na ausência do imposto. Os vendedores são punidos, porque poderiam receber um preço maior e ter maior demanda. Mas qual é o peso morto dessa tributação? Qual seria o ganho de bem-estar decorrente de uma desoneração tributária? E como esse ganho seria distribuído entre produtores e consumidores? Essas perguntas foram elucidadas nesse trabalho.

Para tanto, recorreu-se a técnica proposta por Berry, Levinsohn e Pakes (1995), usando um modelo *Mixed* LOGIT em um exercício empírico. E até chegar as respostas finais, encontraram-se outros resultados a se considerar: 1) o tamanho estimado do mercado brasileiro de automóveis é de 19,5 milhões de unidades para 2008; 2) as elasticidades-preço da demanda oscilam entre -7,53 e -4,01 para modelos populares como o Ford Ka 1.0 e o VW Gol 1.0, respectivamente, e -2,26 e -2,08 para modelos de luxo como o Honda Civic e o GM Zafira, respectivamente; 3) a estimativa do *markup* médio praticado no mercado em 2008 foi de 21,08%; e, 4) uma isenção tributária completa aumentaria as vendas em 2008 em 388 mil unidades. Desse montante, 38 mil seriam de modelos populares e 350 mil de modelos não populares. Mais ainda, os automóveis Sedan e Minivan seriam os que mais teriam aumento relativo de demanda.

A isenção tributária levaria preços e quantidades demandas a se ajustarem de tal forma que o lucro operacional das oito principais montadoras aumentaria em R\$ 6,95 bilhões em 2008. Por outro lado, computando a variação compensatória, uma estimativa do excedente ganho pelo consumidor é de R\$ 24,92 bilhões para 2008. Acabando com um peso morto em torno de R\$ 7,04 bilhões. Assim, o ganho total de bem-estar estimado seria da ordem de R\$ 31,87 bilhões. Conclui-se então que 78,2% da conta tributária é paga pelo consumidor e 21,8% paga pelo produtor.

2. Mais amigos, mais chance de emprego? Um exercício empírico usando os dados de Portugal

2.1 Apresentação

Um considerável interesse tem emergido recentemente na literatura econômica sobre as interações sociais e a forma como as estruturas sociais afetam o comportamento individual. Isso é de fundamental interesse, em particular, para a economia do trabalho. Pois este é um mercado muito diferente da maioria dos outros mercados, uma vez que reflete inúmeras idiosincrasias de cunho social.

Do lado dos trabalhadores, existem diferentes níveis de produtividades (associadas as diferentes escolhas de níveis de educação, habilidades natas etc.), diferentes intensidades e formas de busca por emprego (via agências governamentais, anúncios de jornal etc.), diferentes salários de reserva (associados a preferências individuais em relação à dicotomia trabalho-lazer) e diferentes hábitos culturais, religiosos etc. Do lado das firmas, existem diferentes intensidades do uso de capital e trabalho, diferentes técnicas de seleção de pessoal etc. E de ambos os lados esses aspectos são notadamente influenciados pelo ambiente social onde se encontram trabalhador e firma.

Uma proeminente forma de se tratar essas idiosincrasias no mercado de trabalho surgiu com a teoria de *matching* em meados da década de 1970. Tal forma de análise nasce da adaptação dos processos estatísticos de “urna-bola” feitos nos trabalhos seminiais de Robert Hall, Peter Diamond, Dale Mortensen e Christopher Pissarides²⁹. A idéia é de que trabalhadores desempregados podem ser vistos como bolas e empregos (vagas) como urnas. Em cada momento do tempo existem processos de comportamento social que “lançam” as bolas em direção as urnas. Ao final do processo, algumas bolas estarão em *matching* com algumas urnas, mas sempre existirão bolas perdidas e urnas vazias.

Na ótica da teoria de *matching*, a menor taxa de desemprego possível se dá quando não existem urnas vazias após os lançamentos. Todavia, a probabilidade disto acontecer é bastante reduzida devido a uma série de fricções que existem no mercado de trabalho. Uma vez que não existe um planejador central que lance cada bola em cada urna, existe um lançamento não coordenado dado pela estrutura social. E trabalhadores desempregados mais bem informados podem localizar mais (e melhores) postos de trabalho. Quando se diz “mais bem informado”, isso pode significar: mais educação, mais intensidade de busca, melhor rede social etc.

Este ensaio irá analisar o processo de *matching* (e sua conseqüente taxa de desemprego) em razão do tamanho da rede social (a *network*, um conjunto de trabalhadores que pode indicar um posto de trabalho vago a um desempregado). Em termos coloquiais, o objetivo é buscar a resposta da seguinte pergunta: mais amigos, mais chance de emprego?

A idéia é analisar os aspectos teóricos que contextualizam esta questão, e depois conduzir um exercício empírico usando um modelo *Mixed* LOGIT e uma base de dados de Portugal que contém perguntas desenhadas para se entender esse processo.

²⁹ Os aspectos históricos desta teoria são apresentados em Petrongolo e Pissarides (2001).

Especificamente, o *Mixed* LOGIT é usado para simular a escolha da firma de como preencher um posto de trabalho com determinado tipo de trabalhador³⁰.

Como será apresentado, observaram-se indícios de que ter poucos amigos é ruim, pois reflete pouca informação. Mas ter muitos amigos também é ruim, pois mais gente acaba compartilhando a mesma informação e externalidade positiva da *inside information* fica fragmentada.

Além desta introdução, a seção 2.2 apresenta contexto teórico que fundamenta o ensaio. A seção 2.3 discutirá a estratégia empírica a ser adotada. As seções 2.4 a 2.6 abordam a amostra, os resultados encontrados para o *Mixed* LOGIT proposto e as simulações conduzidas, respectivamente. Por fim, a seção 2.7 aponta as conclusões.

2.2 O contexto das *networks* e do mercado de trabalho

Entende-se por rede social (*network*, em inglês) um agrupamento de pessoas com interesses em comum. A rede é responsável pelo compartilhamento de informações. Neste ensaio, a *network* é um conjunto de trabalhadores que pode indicar um posto de trabalho vago a um amigo desempregado.

A literatura econômica que trata das *networks* no contexto acima descrito pode ser dividida em três fases³¹. A primeira fase começa com o trabalho de Rees (1966), documentando que uma pesquisa na cidade de Chicago apontou que quase 50% dos entrevistados haviam encontrado seus empregos pela influência de amigos ou familiares.

O trabalho de Rees foi o estopim de uma série de levantamentos estatísticos separando a busca por emprego em dois canais: o *formal* (ou convencional), que são as agências de emprego, os anúncios de jornal, a entrega direta de currículos etc.; e o *informal* (ou não convencional), que são as informações ou indicações recebidas de amigos e familiares.

Do trabalho de Rees até o fim da década de 1990 a literatura focou-se em estabelecer fatos estilizados sobre o assunto. Os trabalhos de Corcoran, Datcher e Duncan (1980), Datcher (1983), Holzer (1988) e Topa (2001) entre outros³² apontam que três desses fatos podem ser estabelecidos para os Estados Unidos:

- F1. Cerca de $\frac{1}{4}$ dos trabalhadores (desempregados e empregados em busca *on-the-job*) usam sua *network* como principal meio de busca de vagas;
- F2. O uso e a eficiência da *network* em aumentar as chances de *matching* podem variar substancialmente entre faixas etárias, grupos étnicos e posições geográficas;

³⁰ Em contraponto com o ensaio anterior: lá simulou-se a escolha do consumidor em relação a compra de um modelo de automóvel (ou não comprar nenhum automóvel).

³¹ Um *survey* foi providenciado por Ioannides e Loury (2004).

³² Entre esses, um trabalho singular é o de Bartlett e Miller (1985), que usaram os dados de uma pesquisa respondida por 132 mulheres executivas das maiores empresas dos Estados Unidos em 1982 e observaram que 42% delas responderam “sim” a seguinte pergunta: é quem você conhece, e não o que você sabe que realmente importa? Nesse contexto, aplicaram regressões tendo como variável dependente os salários destas executivas contra seus perfis de respostas, e observaram que maiores salários estavam associados a fatos como ser membro de clubes privados desde antes de encontrarem o emprego e sentirem que tinha uma boa *network*.

F3. Trabalhadores que usam sua *network* apresentam maiores chances de *matching*, mas isso não implica em salários diferenciados.

A segunda fase da literatura se dá durante a década de 1990 (se sobrepondo à primeira) com a publicação dos primeiros modelos teóricos que tratam da influência das redes sociais no mercado de trabalho. Esta fase é marcada fortemente pelos trabalhos de Montgomery (1991 e 1994), que particularmente focou no terceiro fato estilizado descrito acima. Esses trabalhos apontam que se o trabalhador tem uma *network* eficiente não implica necessariamente que ele é mais produtivo, e acaba por receber um salário maior. Muito pelo contrário, o trabalhador pode ter uma ótima *network*, mas ser pouco produtivo.

A literatura nessas duas primeiras fases indicava que o uso das *networks*, no mínimo, não era prejudicial ao mercado de trabalho. Mais especificamente, parecia haver um consenso de que quanto maiores fossem as redes sociais (mais amigos se tivesse) mais chances de *matching* teriam os trabalhadores (e isso tenderia a reduzir o desemprego).

Todavia, iniciando a terceira fase da literatura, os trabalhos de Calvó-Armengol (2004) e Calvó-Armengol e Zenou (2005), doravante denominados por C-A&Z, sugeriram que o uso de *networks* (ou, mais especificamente, o tamanho delas) não necessariamente implica em menores taxas de desemprego. Mais ainda, C-A&Z mostraram que a relação entre *network* e mercado de trabalho é bem mais complexa do que se imaginava inicialmente.

O modelo C-A&Z é desenvolvido sobre uma estrutura de *matching* ao estilo Pissarides (2000), mas em tempo discreto. Para que o modelo seja algebricamente tratável, os trabalhadores são considerados homogêneos (mesma educação, mesma habilidade, mesma idade etc.). E existem duas suposições de base:

- S1. Um trabalhador empregado que ouve falar de um posto de trabalho vago só detém esta informação durante o período vigente. Ou, de forma alternativa, se uma vaga não for preenchida no período vigente ela é destruída;
- S2. Um trabalhador recentemente empregado passa por um período probatório, quando sua produtividade é (normalizada em) zero. É uma fase na qual o ele apenas aprende a cultura da firma. O trabalhador passa a ser produtivo só após o período de provação, e mantém sua produtividade constante até se desligue do emprego. E quando ele achar um novo trabalho passará por um novo período probatório.

Juntas, estas duas suposições implicam que um trabalhador empregado que ouve falar de um posto desocupado não terá receio de transmitir esta informação a um amigo desempregado. Pois, como as vagas são válidas por apenas um período e existem diferenças de salário decorrentes das diferenças de produtividade³³, nenhum trabalhador empregado tem incentivo a esconder informação de um amigo.

Em cada ponto do tempo existem n trabalhadores, que podem estar empregados ou desempregados. Ao término do período $t-1$, as taxas de desemprego e emprego são u_{t-1} e $1-u_{t-1}$, respectivamente. Cada período t começa com essas taxas.

Em C-A&Z as firmas anunciam as vagas de forma que todos os trabalhadores (empregados e desempregados) têm igual acesso a informação. Então, cada trabalhador

³³ Como no período probatório a produtividade é nula o salário é zero, mas quando o trabalhador sai do período probatório ele passa a receber um salário maior que zero.

pode isoladamente se defrontar com um determinado posto de trabalho desocupado com igual probabilidade $v_i = V_i / n$ (onde V_i é o número de postos vagos no momento).

O “isoladamente” significa: sem usar a *network*. Mas se o trabalhador não estiver isolado os amigos podem ajudá-lo a arrumar um emprego. E para estruturar isso, supõe-se que a cada período cada trabalhador está potencialmente³⁴ em contato s dos n trabalhadores. É tamanho da *network*, e é exógeno e igual para todos.

Desempregados não têm incentivos a indicar vagas a outros desempregados (pois são concorrentes). Logo a única forma da *network* funcionar (alguém indicar uma vaga a um amigo) é de empregado para desempregado. Então, para se calcular a probabilidade individual de encontrar uma vaga via *network*, fixa-se um trabalhador desempregado i e considera-se algum outro trabalhador empregado j em contato direto com i . A probabilidade de j ter um determinado número a de amigos desempregados além de i é:

$$\binom{s-1}{a} \cdot (1 - u_{t-1})^{s-a-1} \cdot u_{t-1}^a$$

A probabilidade de j não ter amigos desempregados além de i , ou ter um, ou ter dois, ..., ou ter todos seus amigos desempregados além de i é dada:

$$\sum_{a=0}^{s-1} \binom{s-1}{a} \cdot (1 - u_{t-1})^{s-a-1} \cdot u_{t-1}^a$$

Se todo trabalhador desempregado pertencente à *network* de j tem igual chance de ser encontrado (se todos são igualmente amigos), a probabilidade de j encontrar um amigo desempregado além de i é:

$$\sum_{a=0}^{s-1} \binom{s-1}{a} \cdot \frac{(1 - u_{t-1})^{s-a-1} \cdot u_{t-1}^a}{a+1}$$

E com um pouco de álgebra a expressão acima é reduzida a:

$$\frac{1 - (1 - u_{t-1})^s}{u_{t-1} \cdot s}$$

O trabalhador j está empregado e atento a uma oferta de trabalho redundante a ele, porque é importante para um amigo. A informação de um posto de trabalho vago chegará a j com probabilidade $v_i \cdot (1 - u_{t-1})$ (probabilidade se defrontar com uma vaga e estar empregado). Então a probabilidade de j encontrar uma oferta de trabalho redundante e um amigo desempregado além de i para repassá-la é:

$$v_i \cdot (1 - u_{t-1}) \cdot \frac{1 - (1 - u_{t-1})^s}{u_{t-1} \cdot s}$$

E a probabilidade de que nenhum integrante da *network* transmita a informação da vaga para i é dada por:

³⁴ Ter s amigos não significa que o trabalhador irá encontrar todos eles durante o período.

$$\left(1 - v_t \cdot (1 - u_{t-1}) \cdot \frac{1 - (1 - u_{t-1})^s}{u_{t-1} \cdot s}\right)^s$$

Então, a probabilidade de que i em particular receba a informação de um posto de trabalho vago via *network* é dada por:

$$(1) \quad P_t(s, u_{t-1}, v_t) = 1 - \left(1 - v_t \cdot (1 - u_{t-1}) \cdot \frac{1 - (1 - u_{t-1})^s}{u_{t-1} \cdot s}\right)^s$$

O resultado (1) é a chave do modelo C-A&Z, particularmente porque para cada par (u, v) existe um único máximo global para $P_t(\cdot)$, denotado por \bar{s} . Intuitivamente, isso significa que o tamanho da *network* tem dois efeitos sobre $P_t(\cdot)$: um positivo, que é a disseminação de informação; e um negativo, que é a aglomeração da informação. E isso vai de encontro ao que parecia ser senso comum na literatura até o início dos anos 2000.

O efeito positivo da disseminação da informação é um resultado bem visível. Mas menos intuitivo é o resultado negativo. Para enxergá-lo, imagine que o trabalhador tem inicialmente poucos amigos e note que a medida que a *network* cresce, a margem de aumento da probabilidade de que um trabalhador em particular receba uma informação útil via *network* é cada vez menos importante (é menor). Isso é reflexo do fato que cada vez mais gente terá a mesma informação, reduzindo a relevância da *inside information* da *network*.

Quando as conexões sociais aumentam, cada trabalhador desempregado tem conexões mais diretas que lhe podem transmitir informações de trabalho. Mas cada de um dos amigos diretos também têm conexões mais diretas, de forma que a informação do trabalhador empregado é compartilhada agora por um grupo maior de trabalhadores desempregados. Esse é precisamente o efeito negativo da aglomeração da informação, pois a notícia de uma mesma vaga chega a muitos trabalhadores, de forma que a chance de cada um individualmente se empregar se reduz. Em termos coloquiais: a *inside information* vai deixando de ser *inside information*.

Dando continuidade ao raciocínio, a probabilidade de um trabalhador desempregado ser contratado no período t é a probabilidade de ser contratado da forma convencional (v_t) ou não achar uma vaga da forma convencional e achar via *network*: $v_t + (1 - v_t) \cdot P_t(s, u_{t-1}, v_t)$. Como existem u_{t-1} trabalhadores desempregados no início do período, a probabilidade de que ocorra um *matching* nesta economia no decorrer do período é dada por:

$$(2) \quad m_t(s, u_{t-1}, v_t) = u_{t-1} \cdot (v_t + (1 - v_t) \cdot P_t(s, u_{t-1}, v_t))$$

Proposição (em Calvó-Armengol e Zenou, 2005). *No mercado de trabalho com network, $m(\cdot)$ possui as seguintes propriedades: é crescente em u e v ; e, sendo \bar{s} o único máximo global de $P(\cdot)$, $m(\cdot)$ é crescente no intervalo $[0, \bar{s}]$ e decrescente no intervalo $[\bar{s}, n]$.*

A função *matching* crescente em u e v está em perfeita consonância com o processo “urna-bola” (descrito em Pissarides, 2000 – sem estrutura de *network*). A grande

novidade está na função *matching* ter um ponto de máximo em \bar{s} . E como consequência disso, C-A&Z também mostram que em estado estacionário a taxa de desemprego será minimizada se o tamanho da rede social for \bar{s} (e não quando $s \rightarrow n$, como parecia esperar a literatura até então).

A idéia de C-A&Z foi aplicada empiricamente em três trabalhos. Primeiro em Wahba e Zenou (2005), que usaram uma base de dados para o Egito, onde cerca de 20 mil trabalhadores foram entrevistados em 1998 (32,5% dos empregados afirmaram ter encontrado a vaga devido a indicações de amigos ou familiares).

Wahba e Zenou usaram um modelo LOGIT simples no estudo, de variável dependente binária. Definindo $z = 0$ se um trabalhador não encontrou a vaga devido a *network* (dado que estava empregado, E) e $z = 1$ no caso contrário. Então, para uma matriz de controles X e um vetor de parâmetros β : $\Pr(z = 0 | E) = 1/(1 + \exp(X \cdot \beta))$ e $\Pr(z = 1 | E) = \exp(X \cdot \beta)/(1 + \exp(X \cdot \beta))$. Nesse contexto, o modelo C-A&Z prevê que $\Pr(z = 1 | E)$, que é uma *proxy* de (1), tem um comportamento côncavo em relação ao tamanho da *network*.

A matriz de controles X de Wahba e Zenou possuía um vetor que apontava a densidade populacional do local onde estava empregado o trabalhador. Supostamente, essa densidade seria uma *proxy* para s . E os resultados estimados indicaram que $\Pr(z = 1 | E)$ seria côncava em relação a densidade populacional, corroborando com o modelo C-A&Z.

Já Ioannides e Soetevent (2006) usaram modelo C-A&Z para um exercício de simulação com dados dos Estados Unidos e da França, calibrando os diversos parâmetros do modelo (completo) e testando várias possibilidades para s acrescido de algo que chamaram de “viés de conexão”. O “viés de conexão” é um instrumento que construíram para modificar $P(\cdot)$, de forma a ser diferente entre os trabalhadores (ou seja, incluíram um determinado tipo de heterogeneidade). A idéia é ajustar o modelo ao fato de que desempregados de longo prazo terão menores probabilidades de incorrer em *matchings*.

Seus resultados, com essa modificação, apontam que maiores *networks* tendem a reduzir as taxas de desemprego da economia no estado estacionário. Conclusão: Ioannides e Soetevent contradizem a previsão do modelo C-A&Z.

Fontaine (2008) avança na estrutura de Ioannides e Soetevent desenhando um s aleatório e diferente para cada um dos n trabalhadores e acrescentando diferentes intensidades de busca de vagas. Supostamente alguns trabalhadores são mais esforçados que outros na busca de emprego, assim Fontaine introduz outra forma de heterogeneidade no modelo.

Em exercícios de simulação, usando os mesmos dados de Ioannides e Soetevent, Fontaine encontra resultados ambíguos para $P(\cdot)$ e $m(\cdot)$ em relação ao tamanho da *network*. Dependendo dos parâmetros calibrados no modelo, $P(\cdot)$ e $m(\cdot)$ podem ser crescentes no tamanho médio da *network* ou ter o comportamento previsto por C-A&Z.

Como pequenas modificações no modelo original acabam por gerar resultados ambíguos, os trabalhos mais recentes como Ballester, Calvó-Armengol e Zenou (2006) e Calvó-Armengol, Patacchini e Zenou (2009) buscam entender melhor a relação entre *network* e o mercado de trabalho saindo do escopo de modelagem de *matching* e passando usar Teoria dos Jogos. A idéia, em termos gerais, é que essa relação é bem mais complexa do que inicialmente se imaginava, e, portanto, demanda um instrumental matemático mais complexo para descrever a forma como as *networks* se constroem ao longo da vida dos indivíduos, tanto em quantidade quanto em qualidade.

2.3 A estratégia empírica

A estratégia empírica deste ensaio é aprimorar a idéia de Wahba e Zenou e depois promover simulações para se estudar as relações do modelo C-A&Z. Para tanto, nesta seção se fará uma seqüência de quatro considerações. O uso conjunto dessas quatro considerações permitirá estudar a topologia da função *matching* em um ambiente de *network*. A idéia que permeia esta estratégia é simular a escolha das firmas em preencher vagas com determinados tipos de desempregados.

2.3.1 A taxa de *pseudo-matching*

Seguindo C-A&Z, u_t e $1 - u_t$ são as taxas de desemprego e emprego do período t , respectivamente. A taxa de emprego, por sua vez, pode ser vista como a probabilidade de um determinado trabalhador estar empregado. Mas se o trabalhador está empregado no período t , das duas uma: i) ou ele incorreu em *matching* em $t - 1$, com probabilidade $m_{t-1}(s, u_{t-2}, v_{t-1})$; ii) ou ele já estava empregado em $t - 1$.

A probabilidade de que um determinado trabalhador empregado em t já estivesse empregado em $t - 1$ é a probabilidade de que seu posto não tenha sido destruído no período e que ele já estivesse empregado. Como é de praxe nos modelos de *matching*, assumamos uma taxa de destruição exógena dada por $0 < \lambda < 1$. Portanto, concluí-se que:

$$1 - u_t = m_{t-1}(s, u_{t-2}, v_{t-1}) + (1 - \lambda) \cdot (1 - u_{t-1})$$

Replicando o raciocínio anterior para os períodos passados obtém-se:

$$\begin{aligned} 1 - u_t &= m_{t-1}(\cdot) + (1 - \lambda) \cdot (1 - u_{t-1}) \\ 1 - u_t &= m_{t-1}(\cdot) + (1 - \lambda) \cdot (m_{t-2}(\cdot) + (1 - \lambda) \cdot (1 - u_{t-2})) \\ &\vdots \\ 1 - u_t &= \sum_{i=0}^{\infty} (1 - \lambda)^i \cdot m_{t-1-i}(\cdot) \end{aligned}$$

Ou seja, a probabilidade de encontrar um trabalhador empregado em t é a probabilidade de que se tenha encontrado alguém que incorreu em *matching* ontem, ou anti-ontem, ..., dando maiores pesos para datas mais próximas do presente.

Faça agora a seguinte suposição: $m_t = \bar{m} + \zeta_t$. Onde \bar{m} é uma constante que representa a probabilidade média de *matching* da economia ao longo do tempo e ζ_t é um distúrbio tal que $E(\zeta_t) = 0$. Ou seja, supõe-se que a taxa de *matchings* da economia é estacionária em torno \bar{m} ao longo do tempo. Isso nada mais é do que a base da derivação da *Curva de Beveridge*, que relaciona inversamente taxa de desemprego e taxa de postos de trabalhos vagos na economia (ver Blanchard *et al.*, 1989).

E ao se aplicar essa suposição no resultado anterior se obtém³⁵:

$$(3) \quad 1 - u_t = \bar{m} \cdot (1 + 1/\lambda) + \sum_{i=0}^{\infty} (1 - \lambda)^i \cdot \zeta_{t-1-i} \quad \Rightarrow \quad E(1 - u_t) = \bar{m} \cdot (1 + 1/\lambda)$$

Ou seja, a expressão acima aponta que a atual taxa de emprego da economia é a taxa média de *matching* média ampliada por um determinado fator mais uma soma ponderada de distúrbios (que perdem força ao longo do tempo). O que implica que, na média, a atual taxa de emprego da economia é a taxa média de *matching* ampliada por um fator inverso a taxa de destruição dos postos de trabalho.

Por exemplo, usando os valores de Blanchard e Portugal (2001), têm-se como fatos estilizados um λ de 0,039 para Portugal e de 0,073 para os Estados Unidos. Com esses valores deduz-se que $E(1 - u_t) \approx 26,6 \cdot \bar{m}$ e $E(1 - u_t) \approx 14,7 \cdot \bar{m}$ para estes países, respectivamente. Ou, em outros termos, a taxa média de *matching* de Portugal equivale a 3,8% de sua taxa de emprego, e a dos Estados Unidos 6,8%. Por simplicidade de notação, doravante se tratará $m^* = \bar{m} \cdot (1 + 1/\lambda)$ por taxa de *pseudo-matching*.

Como $1 - u$ e λ podem ser determinados para uma economia, por (3) poderia se construir um vetor para o lado esquerdo de (2), com o intuito de se usar uma técnica econométrica tradicional para se estudar sua topologia em relação a s . Todavia, o lado direito de (2) é demasiadamente intrincado para se usar uma técnica econométrica tradicional. Por isso é necessário recorrer aos recursos que serão discutidos adiante.

2.3.2 O sistema de numeração posicional

Os resultados (1) e (2) de C-A&Z foram deduzidos a partir da hipótese de trabalhadores homogêneos. Todavia, em um exercício empírico isto não é sustentável. Mas parece razoável imaginar que existe um determinado número de características capaz para exaurir os tipos plausíveis de trabalhadores que caracterizam um mercado de trabalho. Isto é, uma forma de se agregar trabalhadores heterogêneos por tipos relativamente homogêneos.

O que se propõe aqui é fragmentar o mercado de trabalho de forma a se usar um caminho do meio: nem um único trabalhador representativo, nem o universo de idiosincrasias de todos os trabalhadores. Essa alternativa visa melhorar a tratabilidade dos dados. E para tanto, se recorrerá a um sistema de numeração posicional³⁶.

Como ilustração da idéia, imagine a matriz X dos dados Wahba e Zenou. Nas linhas estão as informações de cada entrevistado e nas colunas está uma seqüência de K *dummies* que indicam características do trabalhador. Por exemplo, tome $K = 3$ da seguinte forma: a primeira coluna indica o gênero (1 para homem e 0 para mulher), a segunda a faixa etária (1 para menos de 30 anos e 0 caso contrário) e a terceira a forma com que o emprego foi encontrado (1 via *network* e 0 caso contrário).

³⁵ Note que $\sum_{i=1}^{\infty} (1 - \lambda)^i = 1/\lambda$.

³⁶ Discussões sobre o sistema podem ser encontradas em qualquer livro de matemática discreta aplicada à computação. Recomenda-se Scheinerman (2005).

Então existem $2^3 = 8$ possibilidades de “tipos” de trabalhadores observados através das linhas de X . Onde: uma posição binária $0\sim 1\sim 0$ significa *mulher com menos de 30 anos que não encontrou emprego via network*; e uma posição binária $1\sim 0\sim 1$ significa *homem de mais de 30 anos empregado via network*. E para transcrever as posições binárias em índices decimais usa-se o sistema posicional de base 2: isto é, $0\sim 1\sim 0$ é o tipo 2 ($2^0 \cdot 0 + 2^1 \cdot 1 + 2^2 \cdot 0 = 2$); e $1\sim 0\sim 1$ é o tipo 5 ($2^0 \cdot 1 + 2^1 \cdot 0 + 2^2 \cdot 1 = 5$).

Assim, se fragmentará o mercado de trabalho por K características, em até 2^K tipos, indexados por $j = \sum_{k=1}^K 2^{k-1} \cdot d_k$, onde d_k é uma *dummy* para a característica $k = 1, \dots, K$. E $j = 0$ é o “tipo base”, quando todas as *dummies* são zero.

2.3.3 O papel do *Mixed* LOGIT

Em uma economia do mundo real existe uma miríade de trabalhadores e empregos. Do lado dos trabalhadores, o sistema de numeração posicional permitirá que se trabalhe um número razoável de tipos de trabalhadores. Mas do lado dos empregos isso parece ser mais difícil de aplicar, até porque comumente existe mais informação disponível sobre trabalhadores do que sobre empregos.

É para resolver o lado dos empregos que se precisará do *Mixed* LOGIT, nos seguintes termos: em uma economia, em um determinado momento, existem T postos de trabalho (ocupados ou não); T não precisa ser conhecido, tampouco cada posto precisa ser identificado; o pesquisador não sabe por qual perfil j cada um dos $\tau = 1, \dots, T$ postos está sendo ocupado (ou se não está sendo ocupado), mas sabe que as firmas escolhem o melhor para si.

Suponha então que cada firma tem uma função escolha $E_{j,\tau}$, que associa um número ao perfil j em relação a ocupação do trabalho τ . A função escolha é bastante geral e pode representar produtividade, lucratividade ou qualquer outro critério de escolha. Defina a função escolha por:

$$E_{j,\tau} = c_\tau + \underbrace{\sum_{k=1}^K \beta_{k,\tau} \cdot d_{k,j}}_{=\Omega_{j,\tau}} + \xi_j + \varepsilon_{j,\tau}$$

Onde: c_τ representa a ponderação do “tipo base” para o emprego τ ; $\beta_{k,\tau}$ é o peso que relaciona a característica k do trabalhador ao emprego τ ; $d_{k,j}$ é uma *dummy* que indica se o trabalhador do tipo j possui a característica k ; ξ_j é um índice que agrega outras características tipo j que não puderam ser captadas pelo sistema de numeração posicional³⁷; e $\varepsilon_{j,\tau}$ é um termo que indica a subjetividade da escolha do tipo j para o emprego τ , que não é observado pelo pesquisador.

³⁷ Pode ser interpretado como um índice de heterogeneidade “intra-tipo”. Quanto maior forem os seus valores absolutos, menor será a qualidade da seleção das K características escolhidas pelo pesquisador.

Seguindo a construção do *Mixed* LOGIT, defina $E_{u,\tau}$ o resultado da função escolha ao não se escolher ninguém³⁸. Mais ainda, normalize $\Omega_{u,\tau} = 0$. Então, se a firma escolhe ocupar o posto τ com um perfil j implica que $\Omega_{j,\tau} > 0$.

Como discutido no primeiro ensaio desta tese, ao atribuir uma distribuição de probabilidade de Valor Extremo do Tipo I para $\varepsilon_{j,\tau}$ conclui-se que a probabilidade de que um determinado trabalhador tipo j esteja ocupando um determinado emprego τ é:

$$\Pr_{j,\tau} = \frac{\exp\left[c_\tau + \sum_{k=1}^K \beta_{k,\tau} \cdot d_{k,j} + \xi_j\right]}{1 + \sum_{r=0}^{2^K} \exp\left[c_\tau + \sum_{k=1}^K \beta_{k,\tau} \cdot d_{k,r} + \xi_r\right]}$$

Considere agora que \bar{c} e $\bar{\beta}_k$ são, respectivamente, a constante média e o peso médio da característica k no universo dos T postos de trabalho existentes na economia. Mais ainda, considere que σ_c e σ_k são os respectivos desvios-padrão destes pesos neste universo. Então $z_{c,\tau} = (c_\tau - \bar{c})/\sigma_c$ e $z_{k,\tau} = (\beta_{k,\tau} - \bar{\beta}_k)/\sigma_k$ possuem distribuição normal padrão, e a probabilidade de que um determinado tipo j esteja empregado será dada pela seguinte esperança matemática:

$$(4) \quad \Pr_j = \int \frac{\exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,j} + \xi_j\right) + \left(\sigma_c \cdot z_{c,\tau} + \sum_{k=1}^K \sigma_k \cdot d_{k,j} \cdot z_{k,\tau}\right)\right]}{\sum_{r=0}^{2^K} \exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,r} + \xi_r\right) + \left(\sigma_c \cdot z_{c,\tau} + \sum_{k=1}^K \sigma_k \cdot d_{k,r} \cdot z_{k,\tau}\right)\right]} \cdot dF(\mathbf{z})$$

Onde \mathbf{z} é o vetor com $(1+K) \cdot T$ entradas e $F(\mathbf{z})$ é a sua distribuição cumulativa conjunta dos z 's. Duas ponderações são importantes nesse momento. A primeira é que, em termos empíricos, o lado esquerdo de (4) equivale a taxa de emprego do perfil j , $1 - u_j$. Ou melhor, nos termos deduzidos anteriormente, é igual a taxa de *pseudo-matching* ($\Pr_j = m_j^*$) e pode ser observada nos dados devido ao uso do sistema de numeração posicional³⁹.

A segunda é que a princípio o pesquisador não conhece T e tampouco \mathbf{z} , então o lado direito de (4) não pode ser calculado diretamente. Mas trata-se de uma esperança matemática, e então fica evidente o papel do *Mixed* LOGIT. Pois dada a estrutura organizada aqui, essa esperança pode ser calculada usando métodos numéricos aos moldes de McFadden e Train (2000), Train (2003) ou Berry e Pakes (2007). Assim, obtêm-se estimativas para \bar{c} , $\bar{\beta}_k$'s, ξ_j 's, σ_c e σ_k 's dado um determinado conjunto de \tilde{z} 's, em um determinado número Q de postos de trabalho simulados, que se ajustam as taxas de *pseudo-matchings* observadas na base de dados.

³⁸ Note que a frase “não escolher ninguém” pode significar muitas possibilidades. Pode não se ter encontrado um candidato, ou a vaga pode não ter sido anunciada ainda, ou a vaga está reservada para alguém etc.

³⁹ E a taxa de desemprego da economia é dada por $u = \sum_j u_j$.

2.3.4 As simulações de mudanças da *network*

O cerne deste ensaio é contrastar os resultados empíricos com os resultados previstos no modelo C-A&Z. Para tanto, a proposta é seguir uma seqüência de quatro passos descrita adiante.

PASSO 1. Imagine que a característica K seja ter (ou não ter) encontrado o emprego através da *network*. Então re-indexe os tipos de trabalhadores em até 2^{K-1} categorias. Isto é, junte todos os tipos semelhantes que encontraram ou não emprego através da *network*. Esse novo índice de tipo é dado por: $J = \sum_k^{K-1} 2^{k-1} \cdot d_k$. E assim a probabilidade de um trabalhador do tipo J estar empregado é dada por: $m_J^* = \int_z (m_{j'}^*(.,z) + m_{j''}^*(.,z)) \cdot dF(z)$. Onde $m_{j'}^*$ é a probabilidade dele estar empregado no posto τ e ser via *network* e $m_{j''}^*$ é a probabilidade dele estar empregado no posto τ e não ser via *network*.

PASSO 2. Para ocorrerem *matchings* é preciso existir trabalhadores desempregados (u) e postos de trabalho vagos (v). Por (3), a taxa de desemprego e a taxa média de *matchings* estão diretamente relacionadas. A taxa de desemprego é observada pelo pesquisador, e está a esquerda de (4). A direita de (4) imagine os Q postos de trabalho simulados. A idéia é destruir a relação de trabalho de alguns destes postos e transforma-los em vagas⁴⁰. Essa idéia está em perfeita consonância com a dedução da função *matching* de C-A&Z, pois note que até se chegar em (2) a taxa de vagas (v) não é endógena⁴¹.

Operacionalmente, o pesquisador possui uma lista de postos de trabalho simulados indexados por $q = 1, 2, \dots, Q$. Então, de repente, o pesquisador pode transformar V destes Q postos em vagas. Indexe estas vagas por $v = 1, 2, \dots, V < Q$. Cada vaga v tem sua idiosincrasia herdada de q e dada por um vetor \tilde{z}_q , que agora será \tilde{z}_v . Essas idiosincrasias fazem com que cada vaga seja mais adequada a um determinado perfil J do que outro (ou nenhum). Esse é o aspecto fundamental das simulações, pois dependendo das características dos \tilde{z} 's associados às vagas que surgem mais (ou menos) trabalhadores desempregados podem incorrer em *matching* com essas vagas.

A fricção do processo “urna-bola” é gerada aqui pelo fato que cada vaga simulada (urna) se encaixa melhor a um determinado tipo j (um tipo de bola). E quando os tipos são lançados em direção as urnas nem sempre haverá encaixe (*matching*). A questão então é: se, na hora de lançar as bolas, o pesquisador tirar todos os tipos que usam a *network* os encaixes aumentarão? É nessa direção que estão os passos seguintes.

⁴⁰ Lembre-se que alguns deles estão desocupados, mas isso não impede que estes postos possam ser mantidos abertos para uma futura seleção.

⁴¹ Ela passa a ser endógena depois da dedução de (2), quando o modelo ganha mais estrutura para se estudar um equilíbrio de longo prazo.

PASSO 3. A partir de (3), (4) e dos passos anteriores, conclui-se com alguma álgebra⁴² que, em termos da estrutura estimada do *Mixed* LOGIT, a taxa média de *matching* de um trabalhador do tipo J é dada por:

$$(5) \quad \bar{m}_J = \frac{\sum_{v=1}^V \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,J} + \hat{\sigma}_c \cdot \tilde{z}_{c,v} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,J} \cdot \tilde{z}_{k,v}\right] \cdot \left[\frac{\gamma' \cdot \exp(\hat{\beta}_K + \hat{\sigma}_K \cdot \tilde{z}_{K,v} + \hat{\xi}_{j'})}{V \cdot (1 + 1/\lambda)} + \gamma'' \cdot \exp(\hat{\xi}_{j''}) \right]}{1 + \sum_{R=0}^{q^{K-1}} \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,R} + \hat{\sigma}_c \cdot \tilde{z}_{c,v} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,R} \cdot \tilde{z}_{k,v}\right] \cdot \left[\frac{\gamma' \cdot \exp(\hat{\beta}_K + \hat{\sigma}_K \cdot \tilde{z}_{K,v} + \hat{\xi}_{r'})}{V \cdot (1 + 1/\lambda)} + \gamma'' \cdot \exp(\hat{\xi}_{r''}) \right]}$$

Onde os “chapéus” indicam parâmetros estimados via *Mixed* LOGIT, j' e r' são os índices dos tipos que encontraram emprego via *network* e j'' e r'' são os índices dos tipos que não encontraram emprego via *network*. Já os números γ' e γ'' podem assumir os valores 0 ou 1, e por construção serão idênticos a todos os tipos J (sua função será explicitada no passo seguinte).

PASSO 4. No modelo de C-A&Z, a cada período t a economia possui uma diferentes taxas de desemprego (u_t) e emprego ($1 - u_t$), diferentes taxas de vagas (v_t) etc., mas sempre um mesmo tamanho de *network* (s). Então, para dar mais fidedignidade ao exercício, se conduzirá um experimento de Monte Carlo da seguinte forma:

- Tome três cenários: i) no *CENÁRIO BASE* existe um determinado tamanho de *network* $s = s_0$ refletido por $\gamma' = \gamma'' = 1$ em (5); ii) no *CENÁRIO AUSENTE DE NETWORK* existe $s = 0$ refletido por $\gamma' = 0$ e $\gamma'' = 1$; e, iii) no *CENÁRIO PLENO DE NETWORK* existe $s = n$ refletido por $\gamma' = 1$ e $\gamma'' = 0$.
- Em cada cenário, considere a experimentação de t' períodos, indexados por $t = 1, 2, \dots, t'$. Em cada um destes períodos, aleatoriamente aparecerão V_t postos de trabalhos vagos no conjunto dos Q postos conhecidos.
- Em cada cenário e cada período experimentado será coletado as taxas de *matching* estimadas por (5) e suas conseqüentes taxas de desemprego $\hat{u} = 1 - \sum_J m_J^*(\cdot, \gamma', \gamma'')$.

Note que, por construção, o experimento não possui uma característica *ceteris paribus* no seguinte sentido: a cada experimentação mudam vagas, taxas de desemprego e taxas de *matching*. Esse é um ponto bastante positivo acerca da estratégia que se está propondo usar, pois a faz mais próxima do mundo real.

⁴² Ver o ANEXO IV.

Mais ainda, o experimento está em perfeita consonância com a dedução da forma (3), pois a taxa de *matching* de cada simulação $t=1,2,\dots,t'$ (e sua resultante taxa de emprego) sempre tem por base a taxa de *matching* média observada na amostra (\bar{m}_J) corrigida por $1+1/\lambda$ (como se fosse o período anterior).

Por fim, salienta-se que se o resultado do experimento estiver de acordo com o resultado previsto no modelo C-A&Z, a lista de taxas de *matching* gerará a forma de um U invertido quando colocada em um plano cartesiano contra s . Mais ainda, a taxa de desemprego resultante gerará um formato de U neste plano. E isso sendo observado é um indício a favor da idéia de C-A&Z, uma vez que, além de admitir certo grau de heterogeneidade entre os trabalhadores, a técnica proposta aqui não faz suposição nenhuma acerca da estrutura de probabilidade descrita por (1), tampouco sobre a forma funcional ou os rendimentos de escala da função *matching* como utilizado em trabalhos similares (ver Petrongolo e Pissarides, 2001).

2.4 A amostra

Para conduzir o exercício empírico foram usadas as informações do *The European Community Household Panel* (ECHP) para Portugal, dos anos 1994 a 2001. O ECHP é uma pesquisa (ao estilo PNAD) feita nos países da Comunidade Européia com o objetivo de coletar informações socioeconômicas padronizadas.

O questionário inclui blocos de perguntas sobre saúde, educação, condições de moradia, migração, características de emprego etc. Em particular, o questionário possui a seguinte pergunta para quem se declara empregado:

Para encontrar trabalho você...

- a) aplicou um currículo diretamente ao empregador;
- b) encontrou a vaga em anúncio de jornal ou assemelhado;
- c) encontrou a vaga através de uma agência de empregos;
- d) recebeu a indicação de amigos ou familiares; ou
- e) começou seu próprio negócio ou se juntou aos negócios da família.

A mesma pergunta é feita aos desempregados da seguinte forma: nas últimas quatro semanas, qual é o principal meio que você está usando para encontrar trabalho? Ou seja, o fato do ECHP possuir essas perguntas o faz a base adequada para o exercício empírico idealizado neste trabalho.

Em Portugal, no período analisado, foram aplicados 103.063 questionários. Destes, 55.302 entrevistados responderam que estavam empregados no momento e 4.138 desempregados (o restante é formado por inativos e *missings*). Como 48,5% dos questionários foram aplicados todos os anos no mesmo domicílio, na amostra existem pessoas que ao longo do período mudaram sua condição de empregado para desempregado e vice-versa. Mas considerando que pessoas iguais em anos diferentes são trabalhadores diferentes (porque de um ano para o outro estão mais velhos, mais experientes, podem ter se casado ou terminado a faculdade etc.), o número de trabalhadores ativos na amostra foi de 59.440. Isso reflete uma estimativa de PEA de 57,67% em relação ao total de entrevistados, e uma taxa de desemprego de 7,48%.

Estes números estão em consonância com vários levantamentos feitos pelo Instituto Nacional de Estatísticas de Portugal, independentes do ECHP, e analisados em Portugal (2003).

Sobre o tratamento da amostra, ao invés de se trabalhar com um painel (de oito anos e 2^K unidades *cross-section*) optou-se em agregar todos os anos como um único período por três razões:

- R1. A primeira razão é uma consequência do uso do sistema de numeração posicional. Para perceber isso, imagine $K = 10$. Então o pesquisador poderia encontrar nestes 59.440 questionários até 1.024 tipos de pessoas. Mas já nas primeiras tentativas de se operacionalizar a idéia percebeu-se que muitos tipos deixam de aparecer em muitos anos, por mais criativa que for a seleção das K características. O que acaba por gerar um grande número de *missings* para se montar um painel. Mas esse número de *missings* é radicalmente reduzido quando se considera os oito anos em um único período.
- R2. O número de *missings* é uma dificuldade para se trabalhar com dados em painel. Mas obviamente isso não gera uma impossibilidade, uma vez que existem estimadores específicos para esses casos que devem ser processados dentro da rotina computacional do *Mixed LOGIT* (como foi feito no ensaio anterior). Todavia, estes estimadores necessitam que as variáveis explicativas mudem de valor entres os períodos das unidades *cross-section*⁴³. Então a segunda razão de se optar em usar todos os anos como um único período é de cunho econométrico. Pois para que o sistema proposto na seção anterior seja aplicável, todas as variáveis explicativas inseridas na computação do *Mixed LOGIT* são *dummies* que caracterizam os tipos, logo não podem mudar de valor.
- R3. A terceira razão é de origem numérica/computacional. A rotina usada para se estimar os parâmetros do *Mixed LOGIT* envolve um *contraction mapping*⁴⁴ que computaria, nesse exemplo, um vetor de $1.024 \times 8 = 8.182$ entradas (muitas delas *missing*). As taxas de *pseudo-matching* para 1.024 tipos acabam por envolver alguns números extremamente pequenos, da ordem de 1 em $55.302 \cdot (1 + 1/0,039) = 1.473.302$, o que demandaria critérios para aceitação de convergência da contração ainda menores (pelo que se observou poderiam chegar facilmente a ordem $10.E - 20$). Isso, além de aumentar os erros de simulação, pode facilmente demandar dias para se obter convergências em computadores convencionais.

Sobre a determinação das K características a se usar, note que existe um *trade-off* aqui: se K for pequeno demais, o exercício fica irrealista; e se K for grande demais, o exercício fica inviável. A maioria das características de fragmentação do mercado de trabalho está largamente documentada na literatura: sexo, idade, nível de educação etc. Então a determinação de K envolve essencialmente a determinação *thresholds* para essas variáveis “clássicas”.

Após uma série de “tentativas e erros” de seleção usando como critério a minimização de *missings*, selecionou-se $K = 18$ na estrutura na Tabela 1, adiante. Note que por essa estrutura o “tipo base” é alguém com menos de 20 anos, com ensino fundamental, casado, homem e que encontrou emprego sem usar a sua *network*.

⁴³ Ver, por exemplo, a discussão da p. 578 de Wollridge (2002).

⁴⁴ O resultado (13) do primeiro ensaio.

Tabela 1 – Definição das características utilizadas

<i>k</i>	<i>thresholds para as dummies</i>
1	1 se a idade é maior ou igual 20 e menor que 25 anos
2	1 se a idade é maior ou igual 25 e menor que 30 anos
3	1 se a idade é maior ou igual 30 e menor que 35 anos
4	1 se a idade é maior ou igual 35 e menor que 40 anos
5	1 se a idade é maior ou igual 40 e menor que 45 anos
6	1 se a idade é maior ou igual 45 e menor que 50 anos
7	1 se a idade é maior ou igual 50 e menor que 55 anos
8	1 se a idade é maior ou igual 55 e menor que 60 anos
9	1 se a idade é maior ou igual 60 e menor que 65 anos
10	1 se a idade é maior ou igual 65 e menor que 70 anos
11	1 se a idade é maior que 70 anos
12	1 se ensino médio é o maior nível de educação
13	1 se ensino superior é o maior nível de educação
14	1 se nunca casou
15	1 se é viúvo
16	1 se é separado ou divorciado
17	1 se é mulher
18	1 se encontrou emprego com indicação de amigos ou familiares

A Tabela 2, adiante, apresenta as estatísticas descritivas da amostra usando as caracterizações expostas acima. No período de 1994 a 2001, 57,03% dos entrevistados que estavam empregados eram homens. A taxa de desemprego na amostra reduziu-se no período: de 9,04% em 1994 a 5,08% em 2001. Em relação ao estado civil, 67,74% dos entrevistados empregados estavam casados e 25,65% responderam que nunca casaram (e, portanto, estão sendo tratados por solteiros). A faixa etária com maior número de trabalhadores é a entre 25 e 30 anos. E a grande maioria, 81,00%, possuía escolaridade no nível do ensino fundamental.

As estatísticas descritivas expostas na Tabela 2 são todas condizentes com as documentadas por Portugal (2003), com exceção da análise sobre *network* (pois não foi objeto daquele estudo). O percentual de trabalhadores que declarou ter encontrado emprego com indicação de amigos ou familiares foi de 28,7% no período. Muito próximo do que Wahba e Zenou documentaram na amostra do Egito: 32,5%. Mais ainda, note que o número aumentou paulatinamente no período: de 26,19% em 1994 a 32,21% em 2001.

Tabela 2 – Estatísticas descritivas da amostra

Características / Anos		1994	1995	1996	1997	1998	1999	2000	2001	Total
Empregados		5.883	6.241	6.223	12.416	6.124	6.176	5.929	6.310	55.302
Desempregados		585	520	501	998	477	401	318	338	4.138
Total de Trabalhadores		6.468	67.61	67.24	13.414	6.601	6.577	6.247	6.648	59.440
Taxa de Desemprego (%)		9,04	7,69	7,45	7,44	7,23	6,10	5,09	5,08	7,48
Gênero (%)		58,61	57,94	57,21	56,84	57,02	56,82	56,15	55,91	57,03
Mulheres		41,39	42,06	42,79	43,16	42,98	43,18	43,85	44,09	42,97
Casados		68,80	68,01	68,19	67,27	67,60	67,58	67,58	67,37	67,74
Separados		3,17	3,31	3,54	3,91	4,06	4,26	4,05	4,18	3,82
Viúvos		2,69	2,81	2,75	2,73	2,74	2,86	2,93	2,87	2,79
Solteiros		25,34	25,87	25,52	26,09	25,60	25,30	25,44	25,57	25,65
Menos de 20		4,61	4,19	3,82	3,62	3,50	3,27	3,17	2,80	3,62
Entre 20 e 25		10,36	11,14	12,33	12,60	12,50	12,18	11,17	10,80	11,75
Entre 25 e 30		9,91	10,41	11,30	12,48	13,60	14,75	15,93	16,22	13,01
Entre 30 a 35		12,77	12,05	11,97	11,64	11,09	10,49	11,24	11,55	11,60
Entre 35 a 40		12,00	11,79	11,62	11,27	11,68	11,86	11,64	11,39	11,61
Entre 40 a 45		11,83	12,02	11,69	11,60	11,42	10,87	10,66	10,80	11,39
Entre 45 a 50		10,56	10,34	10,60	10,69	10,15	10,25	10,16	10,42	10,43
Entre 50 a 55		9,43	9,29	8,77	8,57	8,48	8,53	8,31	8,84	8,76
Entre 55 a 60		7,93	7,68	7,42	7,37	6,98	6,86	6,88	6,41	7,21
Entre 60 a 65		6,39	5,99	5,92	5,40	5,51	5,75	5,27	5,46	5,67
Entre 65 a 70		2,63	3,00	2,63	2,91	2,92	2,93	3,09	2,78	2,87
Mais de 70		1,59	2,10	1,92	1,85	2,15	2,27	2,48	2,53	2,08
Superior		5,47	5,21	5,37	5,47	9,06	9,14	9,84	10,85	7,32
Média		9,43	9,98	10,92	11,26	11,74	12,79	13,35	14,35	11,68
Fundamental		85,10	84,81	83,72	83,27	79,20	78,08	76,80	74,80	81,00
Convencional		73,81	73,44	73,13	72,06	71,60	69,55	68,30	67,79	71,30
Via de emprego (%)		26,19	26,56	26,87	27,94	28,40	30,45	31,70	32,21	28,70
Network										

Características dos que se declararam empregados

Uma análise combinatória das *dummies* apresentadas acima mostra o que maior número de tipos que poderia ter sido encontrado era 576. Encontrou-se 406.

Abaixo, na Tabela 3 lista-se a descrição dos 10 tipos com maior taxa de emprego na amostra. Como era de se esperar, a maior frequência é de homens, casados com mais de 35 anos.

Tabela 3 – Descrição dos 10 tipos com maior taxa de emprego na amostra	
1 - u (%)	Descrição do tipo
3,47	45 a 50 anos, ens. fundamental, casado, homem e empregou-se sem usar <i>network</i>
3,25	40 a 45 anos, ens. fundamental, casado, homem e empregou-se sem usar <i>network</i>
3,07	50 a 55 anos, ens. fundamental, casado, homem e empregou-se sem usar <i>network</i>
2,88	55 a 60 anos, ens. fundamental, casado, homem e empregou-se sem usar <i>network</i>
2,57	35 a 40 anos, ens. fundamental, casado, homem e empregou-se sem usar <i>network</i>
2,48	60 a 65 anos, ens. fundamental, casado, homem e empregou-se sem usar <i>network</i>
2,33	45 a 50 anos, ens. fundamental, casado, mulher e empregou-se sem usar <i>network</i>
2,21	40 a 45 anos, ens. fundamental, casado, mulher e empregou-se sem usar <i>network</i>
2,17	20 a 25 anos, ens. fundamental, solteiro , homem e empregou-se sem <i>network</i>
2,10	20 a 25 anos, ens. fundamental, solteiro , homem e empregou-se pela <i>network</i>

Dois últimos pontos sobre a amostra. Primeiro, o uso da *network* se mostrou mais freqüente entre os jovens: 45,46% dos trabalhadores com idade inferior a 25 anos afirmaram ter se empregado com a ajuda de amigos e familiares (isso corrobora com o que é documentado para os Estados Unidos em Ioannides e Loury, 2004). O segundo ponto é que com a caracterização utilizada se registrou 37 tipos que apareceram apenas uma vez, 17 apenas duas e 13 apenas três vezes nos oito anos abordados.

2.5 Resultados Estimados

Adiante a tabela 4 apresenta os valores dos parâmetros estimados para equação (4), bem como seus respectivos desvios-padrão e estatísticas *t*. A metodologia utilizada seguiu o mesmo processo descrito no primeiro ensaio desta tese, utilizando:

1. $Q = 1.000$;
2. Semente aleatória 7654321 para formar o vetor \tilde{z} ;
3. Estimador do OLS para \bar{c} e $\bar{\beta}_k$'s na regressão $\delta_j = \bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,j} + \xi_j$, onde δ_j é o termo resultante do *contraction mapping* do *Mixed LOGIT* (veja detalhes no primeiro ensaio desta tese); e,
4. Não houve instrumentalização, pois não é necessário para o caso deste ensaio.

Tabela 4 – Valores dos parâmetros estimados para equação (4) para $Q = 1.000$

Parâmetro	$\hat{\beta}$	Desvio Padrão	Estatística t	$\hat{\sigma}$	Desvio Padrão	Estatística t
Constante	-2,775	0,316	-8,795	1,052	0,459	2,293
Idade 20 a 25	1,596	0,356	4,490	0,280	0,221	1,266
Idade 25 a 30	1,963	0,346	5,680	0,506	0,674	0,751
Idade 30 a 35	1,994	0,342	5,828	0,044	0,035	1,250
Idade 35 a 40	1,995	0,345	5,792	0,319	0,194	1,643
Idade 40 a 45	1,326	0,342	3,883	0,792	0,837	0,946
Idade 45 a 50	1,394	0,342	4,079	0,580	0,584	0,994
Idade 50 a 55	1,181	0,341	3,466	0,105	0,038	2,749
Idade 55 a 60	0,938	0,352	2,668	0,115	0,078	1,467
Idade 60 a 65	0,479	0,358	1,339	0,925	0,629	1,471
Idade 65 a 70	-0,090	0,370	-0,245	0,356	0,599	0,593
Mais de 70	-0,130	0,383	-0,339	0,531	0,799	0,664
Edu. Média	-2,300	0,146	-15,713	1,073	0,955	1,123
Edu. Superior	-2,456	0,153	-15,999	0,905	0,404	2,240
Solteiro	-1,662	0,156	-10,657	1,481	2,877	0,515
Viúvo	-3,176	0,196	-16,218	0,766	0,523	1,465
Separado	-2,547	0,164	-15,524	0,261	0,307	0,852
Mulher	0,144	0,122	1,182	1,229	0,730	1,684
<i>Network</i>	-1,284	0,123	-10,410	1,252	0,951	1,316

Começando a análise da Tabela 4 pelos parâmetros que ponderam as faixas etárias, note que os $\hat{\beta}$'s crescem da faixa de 20 a 25 anos até a faixa de 35 a 40 anos, e depois decrescem. Esse é o mesmo comportamento que se observa na Tabela 2 com o percentual de trabalhadores empregados por faixa etária. E não poderia ser diferente, uma vez que o modelo está sendo ajustado para as taxas de emprego por tipo de trabalhador. As estatísticas t , por sua vez, só indicam baixos níveis de significância para faixas acima dos 60 anos.

Quanto aos $\hat{\sigma}$'s, justamente chama a atenção os maiores valores (relativos) para as faixas acima dos 60 anos. Por exemplo, os números apontam que para uma determinada vaga com característica $\tilde{\alpha} = -1$, um trabalhador da faixa de 60 a 65 anos

terá peso de $0,479 - 0,925 = -0,446$, ou seja, quem escolhe candidatos para essa vaga não deseja alguém nesse perfil. Todavia, se $\tilde{z} = +1$ a pontuação será de $0,479 + 0,925 = 1,404$, o que indica que esta firma simulada se interessa por alguém deste perfil (lembre-se que pelas características da distribuição normal padrão em 68,3% dos casos $\tilde{z} \in [-1,1]$). Notadamente as estatísticas t para os $\hat{\sigma}'s$ são baixas, mas é importante lembrar (do primeiro ensaio desta tese) que os desvios-padrão são calculados por gradientes numéricos, o que leva essas estatísticas a serem menos confiáveis.

Quanto aos $\hat{\beta}'s$ relativos aos níveis de educação média e superior, note que os valores se mostraram negativos. Mas isso também só está refletindo o menor número de trabalhadores com estas características. Já seus $\hat{\sigma}'s$ são relativamente altos, mas, por exemplo, para o peso da educação superior se tornar positivo seria necessário um \tilde{z} de no mínimo 2,714, o que ocorrerá em apenas 0,33% das vezes.

Quanto as características do estado civil, os $\hat{\beta}'s$ negativos para viúvos e separados refletem o fato de que eles (juntos) representem apenas 6,61% dos trabalhadores empregados. Já a situação dos solteiros é tal que, para terem avaliações positivas, a vaga simulada precisa ter no mínimo um $\tilde{z} = 1,122$, o que acontecerá em 13,1% das vezes.

Em relação ao gênero, o respectivo $\hat{\beta}$ não se mostrou estatisticamente significativo. E quando a estimativa $\hat{\beta} = 0,144$ é composta com o respectivo desvio-padrão estimado em $\hat{\sigma} = 1,229$, conclui-se que em 45,4% das vezes mulheres terão avaliações positivas para uma determinada vaga. O que é muito próximo dos seus 42,9% de participação no mercado, como era de se esperar.

Em relação ao uso da *network*, o respectivo $\hat{\beta}$ se mostrou estatisticamente significativo. E apesar de apresentar um valor negativo, teve o segundo maior desvio-padrão estimado (só menor que a *dummy* para Solteiro). Isso implica que existe uma grande heterogeneidade entre as firmas quanto a forma de pontuar esta característica. Mas, ainda assim, para que a pontuação seja positiva será necessário um \tilde{z} de no mínimo 2,714, o que ocorrerá em 15,3% das vezes.

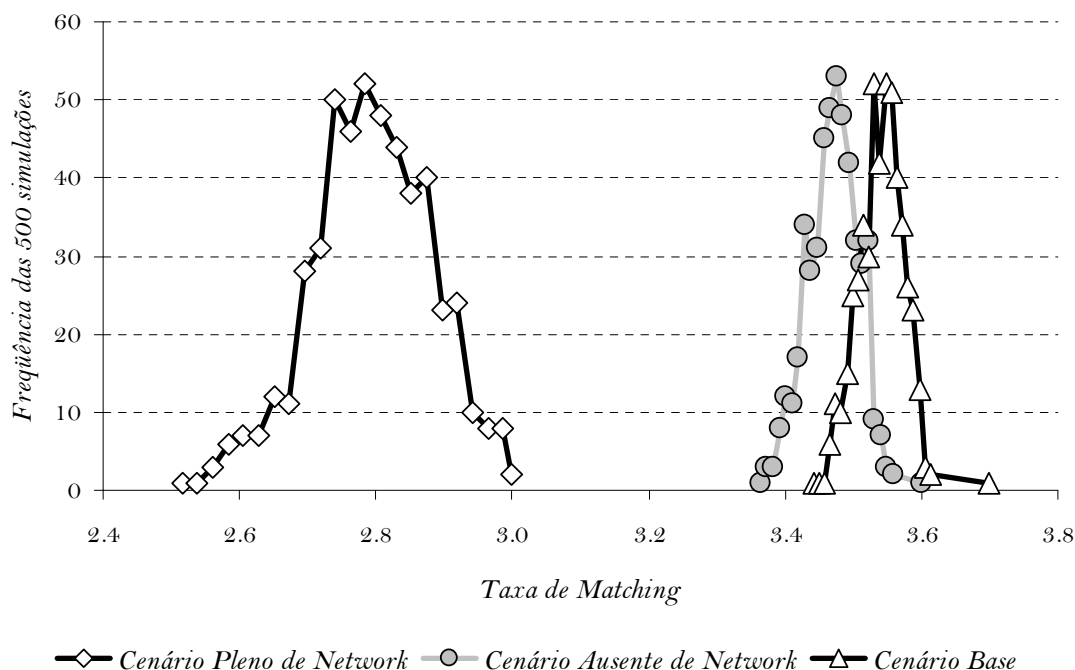
E por fim, como discutido anteriormente, o termo ξ_j representa um índice que agrega outras características do tipo j que não puderam ser captadas pelo sistema de numeração posicional. Por sua vez, o R^2 resultante da regressão $\delta_j = \bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,j} + \xi_j$ é um indicador da eficiência do uso do sistema de numeração posicional para as K variáveis selecionadas e seus *thresholds* determinados. Tal R^2 foi 0,6665, e considerou-se um bom número pois foi o maior dentre muitas seleções de K e *thresholds* experimentadas.

2.6 Simulações

O Gráfico 1, adiante, apresenta as frequências das taxas de *matching* simuladas por cenário para $t' = 500$. Em cada cenário trabalhou-se com V entre 100 e 150, onde tanto o número de vagas como quais dos Q postos de trabalho se tornariam vagas em

cada simulação foram selecionados de forma aleatória. As taxas de *matching* foram computadas através de (5) fazendo $\bar{m} = \sum_J m_J$. E, seguindo Blanchard e Portugal (2001), usou-se $\lambda = 0,039$.

Gráfico 1 – Frequências das taxas de *matching* simuladas por cenário (em %)



No cenário *base* a média da taxas de *matching* foi de 3,54%, ao passo que nos cenários *ausentes* e *plenos* de *network* as médias foram de 3,47% e 2,78%⁴⁵, respectivamente. Ou seja, o exercício apresenta indício a favor da proposição de Calvó-Armengol e Zenou (2005).

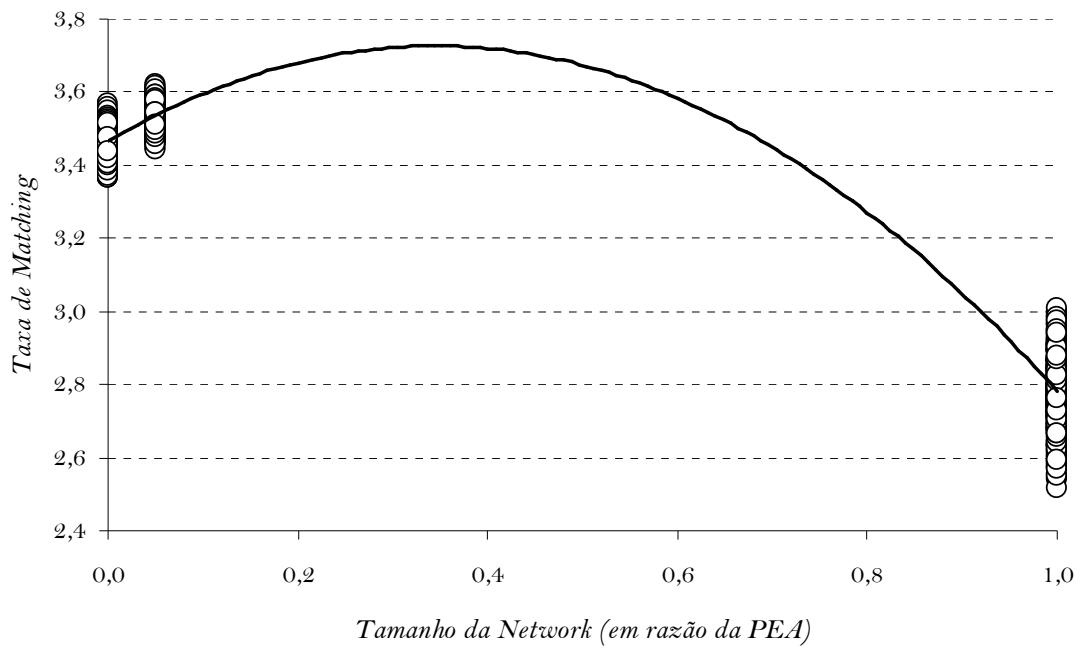
Note também que desvios em relação as médias das taxas de *matching* no Gráfico 1 equivalem na fórmula (3) a uma composição dos ζ_t . Ou seja, em cada simulação $t = 1, 2, \dots, t'$ esses desvios representam as fricções do processo “urna-bola” quando os tipos são lançados em direção as urnas em cada experimento t .

Já o Gráfico 2, adiante, apresenta a dispersão das taxas de *matching* simuladas por tamanho da *network* (como razão da PEA). Nos extremos estão as dispersões nos cenários *ausentes* e *plenos* de *network* ($s/n = 0$ e 1 , respectivamente). Ao centro está a dispersão do cenário *base* considerando, de forma ilustrativa, $s/n = 0,05$. A linha que conecta os aglomerados de pontos é o simples resultado da regressão de uma forma quadrática, e é meramente ilustrativa.

Note que $s/n = 0,05$, no universo dos aproximadamente 5 milhões de trabalhadores portugueses, equivale a dizer que na média cada trabalhador tem uma *network* próxima de 250 mil amigos. Provavelmente este número é exagerado, mas para valores mais baixos a ilustração gráfica não transmitiria a idéia de que os resultados da simulação corroboram com os resultados preditos por C-A&Z. Pois note também que, conforme discutido anteriormente, se observa no plano cartesiano um U invertido.

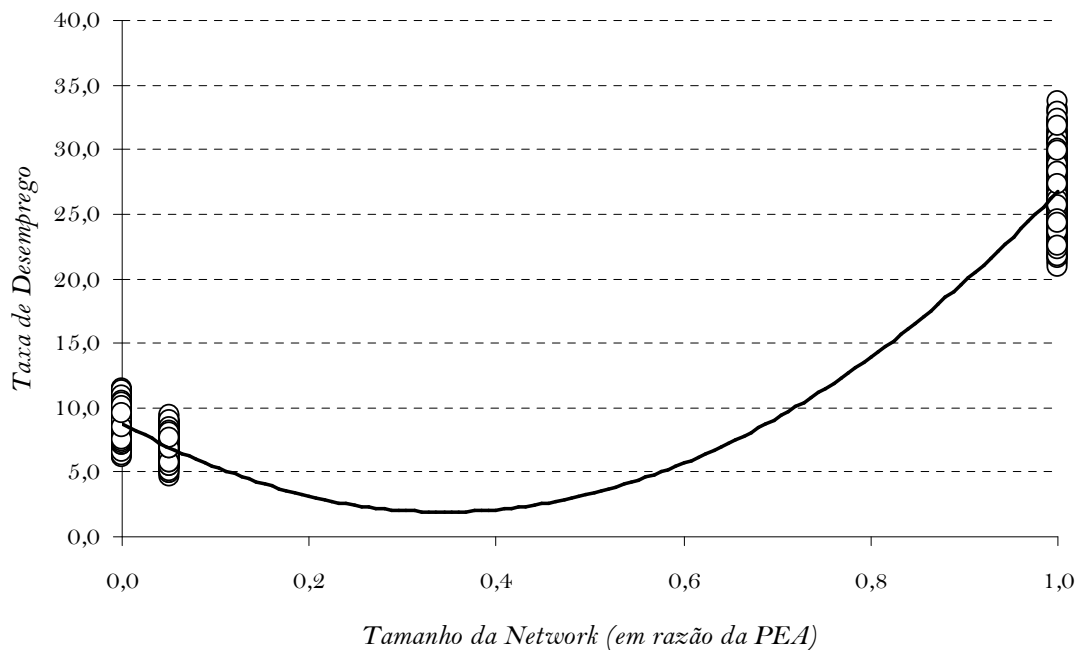
⁴⁵ Por (3), as taxas são em relação a PEA.

Gráfico 2 – Dispersão das taxas de *matching* simuladas por tamanho da *network*



Já o Gráfico 3 apresenta a dispersão das taxas de desemprego simuladas por tamanho da *network* usando $\hat{u} = 1 - \sum_J m_J^*(., \gamma', \gamma'')$. A linha que conecta os conjuntos de pontos também é resultante da regressão de uma forma quadrática e é meramente ilustrativa – note que tem a forma de um U.

Gráfico 3 – Dispersão das taxas de desemprego simuladas por tamanho da *network*



Apesar dos resultados expostos acima corroborarem com o que prediz C-A&Z, note que eles indicam também que os efeitos negativos da *network* no mercado de trabalho só se manifestam quando as redes sociais são extremamente grandes. Nos números utilizados acima, para tais efeitos se revelarem as *networks* deveriam ter tamanhos acima de 250 mil integrantes, o que talvez seja impossível de acontecer.

2.7 Conclusões

Nos últimos anos a literatura econômica vem crescendo seu interesse sobre as interações sociais e as formas como as estruturas sociais afetam o comportamento individual, em particular no mercado de trabalho.

Na ótica da teoria de *matching*, os trabalhadores desempregados mais bem informados podem localizar mais (e melhores) postos de trabalho vagos. Se o trabalhador desempregado está inserido em uma rede social (uma *network*) ele pode fazer uso da *inside information* da rede para localizar as vagas. Mas o modelo de C-A&Z mostra que a *inside information* perde força se a *network* for muito grande, o que contradiz o entendimento de boa parte da literatura na área até então.

Este ensaio propôs uma maneira de tratar empiricamente a questão fazendo uso do *Mixed LOGIT* para simular as escolhas das firmas por tipos de trabalhadores, particularmente atentando para o uso da *network*, e a aplicou em uma base de dados de Portugal que contém perguntas desenhadas para se entender esse processo. A proposta dá ao ensaio um aspecto inovador, cuja lógica é descrita na seguinte seqüência:

1. O pesquisador, na base de dados, observa as características dos empregados, mas não observa as características dos empregadores;
2. Então o pesquisador classifica os trabalhadores por “tipos”, e observa a parcela destes “tipos” no mercado de trabalho;
3. Tendo estas parcelas de mercado nas mãos (e conhecendo os “tipos”) o pesquisador pode criar artificialmente postos de trabalho ocupados de forma a mimetizar os dados – é neste ponto onde se usa o *Mixed LOGIT*;
4. Para existir *matching* é preciso desempregados e postos de trabalho vagos. O perfil dos trabalhadores desempregados é observado na base de dados e o pesquisador conhece os postos de trabalho da economia – ele os criou. Mostra-se então que pesquisador pode também criar vagas artificiais;
5. Então, pela observação dos trabalhadores desempregados e pela simulação de postos de trabalho vagos as estruturas de *matching* podem ser estudadas.

Assim, de fato, observaram-se indícios de que uma sociedade onde na média têm-se poucos amigos tende a ter menores taxas de *matching* e conseqüentes menores taxas de emprego. Mas uma sociedade que na média possui grandes redes de amigos também é ruim para o mercado de trabalho, pois mais gente acaba compartilhando das mesmas informações sobre empregos vagos, e a externalidade positiva da *inside information* fica fragmentada. E isso reduz a chance do trabalhador desempregado encontrar um posto de trabalho vago.

Todavia, para que os efeitos negativos da *network* se manifestem, os resultados indicam que elas deveriam ser extremamente grandes. O que parece ser improvável de acontecer no mundo real.

3. Por quantos votos uma Bolsa Família poderia ser trocada?

3.1 Apresentação

Nas eleições presidenciais de 2006 ocorreu uma clara mudança no perfil de votação do candidato reeleito, Luiz Inácio Lula da Silva (Lula), em relação às eleições de 2002. Em 2002, Lula obteve uma maior aprovação nas regiões mais urbanizadas e desenvolvidas do Brasil, mas em 2006 sua base eleitoral migrou para as regiões menos desenvolvidas do país. Ou seja: a base se inverteu!

Para explicar esse fenômeno surgiram várias hipóteses⁴⁶ no pós-eleições de 2006, longe de serem mutuamente excludentes. Entre elas, pareceram predominar:

- H.1) o aumento dos programas sociais de transferência de renda (em especial o Bolsa Família) atraiu o eleitorado de baixa renda;
- H.2) a seqüência de denúncias e escândalos de corrupção durante o primeiro mandato (em particular o “mensalão”) afastou o eleitorado mais instruído;
- H.3) o resultado seria uma consequência da simples avaliação positiva das políticas econômica e social do primeiro mandato, que geraram mais impactos positivos nas regiões menos desenvolvidas; e,
- H.4) o resultado foi uma consequência do fato de que o candidato da base do governo historicamente tem votações maiores em regiões mais dependentes do Estado.

Hunter e Power (2007) publicaram um dos primeiros estudos econômicos para testar as hipóteses listadas acima, e creditaram enfaticamente a mudança no perfil de votação ao Bolsa Família. A partir deste trabalho, se iniciou um debate acadêmico onde os resultados se mostraram bastante díspares: variando desde conclusões de que o programa Bolsa Família não foi determinante para a reeleição de Lula (Shikida *et al*, 2009), até conclusões de que o programa foi o fator decisivo da vitória e da mudança da base eleitoral (Marques *et al*, 2009).

Canêdo-Pinheiro (2009) observou que a discussão iniciada com o trabalho de Hunter e Power, e espraiada na academia, provavelmente estaria carregando dois problemas metodológicos que poderiam estar gerando os resultados díspares:

- P.1) o uso inadequado de modelos lineares; e,
- P.2) conclusões distorcidas pelo *viés de agregação*.

Quanto ao primeiro problema, em geral essa literatura tem usado uma econometria tradicional, tendo como variável dependente a diferença de votação de Lula entre 2006 e 2002, entre Unidades da Federação (UFs) ou municípios, e considerando como variáveis explicativas o percentual de famílias atendidas pelo Bolsa Família, o PIB per capita, a taxa de escolaridade etc. Então o uso de mínimos quadrados, ou uma variante do método, é inadequado porque a variável dependente é limitada (pertence ao intervalo $[0,1]$ ou $[-1,1]$, dependendo do modelo que se estrutura).

⁴⁶ Ver Carreirão (2007) para uma contextualização da discussão nos primeiros meses após a eleição.

Quanto ao *viés de agregação*, ou *falácia ecológica*, trata-se de um erro de inferência das escolhas (ou resultados) individuais a partir das estatísticas agregadas⁴⁷ (do grupo, do bairro, do município, da UF etc.). Nesse sentido, note que um problema de votação é originalmente um problema de escolha individual. Mas pelo escrutínio secreto não se sabe se determinado indivíduo votou ou não no candidato Lula (dado que reside em determinada localidade). Por conseqüência, se votou no candidato Lula, a princípio não se pode estimar o quanto essa escolha foi determinada pelo Bolsa Família. Assim, usando dados agregados, a estimativa de parâmetros que relacionam o Bolsa Família a votação possivelmente estará viesada.

Enfim, dado que o voto é secreto, não existem micro-dados para as eleições de forma a se operacionalizar uma econometria tradicional para mitigar o *viés de agregação*. Então uma metodologia mais adequada (para se estimar o impacto do Bolsa Família no resultado da eleição) envolveria alguma forma de escolha multinomial (entre os candidatos) e um método econométrico que envolvesse simulação.

Levando em conta as observações de Canêdo-Pinheiro (2009), o objetivo desse ensaio é ajudar a preencher a lacuna existente entre o resultado das eleições presidenciais de 2006 e os aspectos metodológicos até então usados para estudá-lo. Para tanto, usando uma estrutura *Mixed LOGIT*, pretende-se responder a seguinte provocativa: por quantos votos uma Bolsa Família poderia ser trocada?

Além desta introdução, o ensaio possui mais cinco seções. A seção 3.2 contextualiza o cenário das eleições presidenciais de 2006 e a discussão acadêmica que se sucedeu. A seção 3.3 discute os aspectos metodológicos que se propõe aplicar. As seções 3.4 e 3.5 apresentam a amostra e os resultados estimados, respectivamente. E, por fim, na última seção fazem-se as considerações finais.

3.2 O contexto das eleições presidenciais de 2006 e do debate acadêmico

As eleições presidenciais de 2006 tiveram sete candidatos⁴⁸:

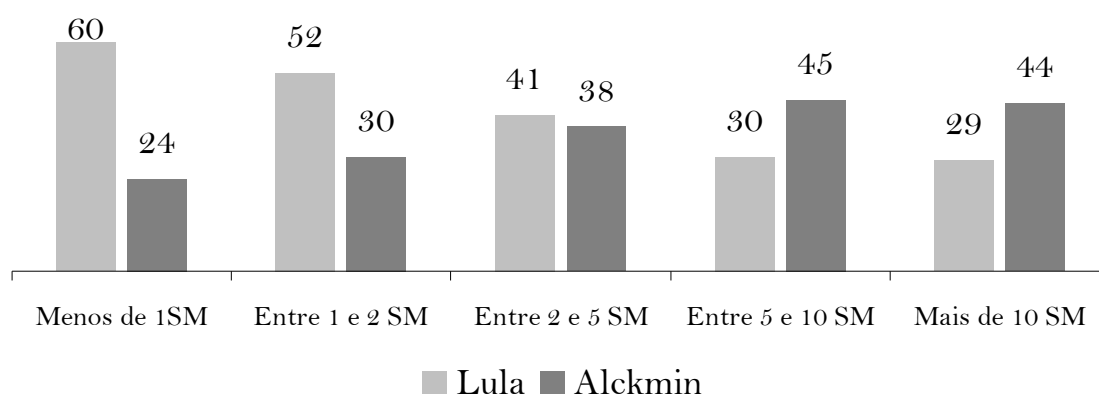
⁴⁷ A explicação do termo é atribuída a interpretação das estatísticas de suicídio da Europa no século XIX feita por Émile Durkheim (um dos precursores da Sociologia). Taxas mais altas eram observadas em países onde a igreja protestante predominava, assim, Durkheim concluiu e disseminou a idéia de que o suicídio era promovido pelo protestantismo. Ele definiu isso por “conclusão ecológica”, mas, obviamente, os países protestantes eram diferentes dos países católicos em muitos aspectos além da religião. Mais ainda, os dados agregados não faziam qualquer ligação individual entre o suicida e sua fé religiosa. Ironizando Durkheim, logo a comunidade acadêmica passou a denominar a “conclusão ecológica” por *falácia ecológica*. Stoker (1993) apontou que a má interpretação de Durkheim ainda se faz presente em muitos trabalhos acadêmicos, mesmo que de forma sutil, mas agora é elegantemente definida por *viés de agregação*. Glaeser, Sacerdote e Scheinkman (2003) também documentaram esse viés, dando-o uma roupagem de “multiplicador social”.

⁴⁸ Como bem lembra Carreirão (2007), é digno de nota mencionar as disputas internas do PMDB, o maior partido do Brasil, com o maior número de prefeitos eleitos e com alto poder de influência nacional. Havia, de um lado, os adeptos do lançamento de candidatura própria (com o ex-Governador fluminense Garotinho tendo vencido o Governador gaúcho Germano Rigotto na disputa interna ao partido); de outro, os adeptos do apoio à candidatura Lula (além de uma minoria favorável ao apoio a Alckmin). Após uma série de disputas internas, prevaleceu a decisão de não lançar candidatura própria, nem apoiar formalmente nenhum outro, deixando os partidários livres para apoiar qualquer candidato a presidente e fazer coligações variadas nos estados. Outro ponto a lembrar é que o candidato pelo Partido da Causa Operária (PCO), Rui Pimenta, teve sua candidatura impugnada pelo Tribunal Superior Eleitoral (TSE) – por não ter prestado contas de sua campanha presidencial em 2002.

1. o Presidente em exercício, Luís Inácio Lula da Silva, pela coligação dos Partido dos Trabalhadores (PT), Partido Comunista do Brasil (PC do B) e Partido Republicano Brasileiro (PRB);
2. o ex-Governador de São Paulo, Geraldo Alckmin⁴⁹, pela coligação do Partido da Social Democracia Brasileira (PSDB) com o Partido da Frente Liberal (PFL);
3. a Senadora Heloísa Helena, pela coligação dos Partido Socialismo e Liberdade (PSOL), Partido Socialista dos Trabalhadores Unificado (PSTU) e Partido Comunista Brasileiro (PCB);
4. o Senador Cristovam Buarque, pelo Partido Democrático Trabalhista (PDT);
5. o ex-Deputado Federal José Maria Eymael, pelo Partido Social Democrata Cristão (PSDC); e,
6. o empresário Luciano Bivar, pelo Partido Social Liberal (PSL); e,
7. a cientista política Ana Maria Rangel, pelo Partido Republicano Progressista (PRP);

Apesar de existirem sete candidatos, a campanha se fechou em torno dos dois primeiros da lista acima. O gráfico 1, abaixo, apresenta os resultados de uma pesquisa de intenção de voto feita pelo IBOPE⁵⁰ na última semana antes do primeiro turno da eleição (em outubro de 2006), separando o eleitorado por faixa de renda.

Gráfico 1 – Pesquisa de intenção de voto (%) por faixa de renda do entrevistado (em Salários Mínimos) – última semana antes do primeiro turno das eleições de 2006



Fonte: IBOPE.

No gráfico acima fica claro o que foi apresentado na introdução deste ensaio: a base eleitoral de Lula em 2006 estava na população de baixa renda (e menos instruída), e o

⁴⁹ Uma disputa acirrada aconteceu no PSDB, onde Alckmin acabou saindo candidato em detrimento a José Serra. A maioria dos estudos, incluindo Carreirão (2007) e Hunter e Power (2007), acabaram por concluir que a escolha de Alckmin (em vez de José Serra) como candidato do PSDB foi bastante favorável ao Presidente Lula: em primeiro lugar, porque, quanto às intenções de voto, a diferença entre Lula e Serra era menor do que a existente entre Lula e Alckmin, no momento da escolha do candidato tucano. Além disso, como Alckmin era menos conhecido nacionalmente, gastou-se muito tempo para apresentar-lhe ao eleitorado, antes de iniciar uma campanha negativa contra Lula (o que era necessário, dada a grande vantagem do presidente no início da campanha).

⁵⁰ Instituto Brasileiro de Opinião Pública e Estatística.

inverso acontecia com Alckmin. As principais hipóteses para o exposto estão ligadas ao Bolsa Família e aos escândalos de corrupção durante o primeiro mandato.

Os programas de transferência direta de renda pelo governo federal se iniciaram em 2001⁵¹, no mandato do Presidente Fernando Henrique Cardoso. No início eles eram uma miríade de programas sociais que beneficiavam cerca de 5 milhões de famílias, através, entre outros, de programas como o “Bolsa Escola”, vinculado ao Ministério da Educação, “Auxílio Gás”, vinculado ao Ministério de Minas e Energia e o “Cartão Alimentação”, vinculado ao Ministério da Saúde, cada um gerido por administrações diferentes. O Bolsa Família se consistiu na unificação e ampliação desses programas num único programa social, com cadastro e administração centralizados no Ministério do Desenvolvimento Social e Combate à Fome (MDS), criado em 2003 pelo Presidente Lula⁵².

Menos de quatro anos depois da criação do Bolsa Família, próximo das eleições de 2006, pelos dados do MDS mais de 11,1 milhões de famílias de baixa renda, equivalendo a cerca de 45,0 milhões de pessoas, recebiam R\$ 8,2 bilhões em transferências diretas via Bolsa Família ao ano, o que correspondia a 0,4% do PIB brasileiro. Ou seja, durante o primeiro mandato do Presidente Lula, houve um crescimento de mais de 120% no número de pessoas atendidas pelos programas sociais de transferência direta de renda.

Levando em conta isso e seu possível reflexo eleitoral, o Instituto Datafolha fez uma pesquisa no final de junho de 2006 fazendo a seguinte pergunta a uma amostra de 2.828 eleitores em todo o país: “ *você, alguém da sua família, ou alguém próximo a você participa de um programa do governo federal... do Bolsa Família?*”. E em seguida perguntou sobre a intenção de voto para o candidato Lula. Os resultados são apresentados abaixo, na Tabela 1:

Tabela 1 – Resultados da pesquisa do Datafolha de junho de 2006					
<i>“você, alguém da sua família, ou alguém próximo a você participa de um programa...?”</i>					
Resposta:	O entrevistado participa	Alguém da família participa	Alguém próximo participa	Total de respostas Sim	Total de respostas Não
Casos:	11%	27%	15%	42%	58%
<i>“você pretende votar no candidato Lula?”</i>					
Sim	62%	54%	55%	56%	39%

Fonte: Datafolha.

Os resultados apresentados acima apontam que 62% das pessoas que responderam participar de um programa iriam votar em Lula, e 56% dos que participavam ou tinham alguém próximo participando iriam votar em Lula. Em contraponto, esse número se reduzia a 39% para quem respondia não participar ou não ter nenhum conhecido participando do programa. Esse resultado e outros semelhantes a ele é que deram base a formulação da hipótese de que foi o aumento dos programas sociais de transferência

⁵¹ Tendo como inspiração o Bolsa Escola, um programa criado pela Prefeitura de Campinas em 1994 e logo em seguida aplicado também no Distrito Federal em 1995.

⁵² Detalhes em: <http://www.mds.gov.br/bolsafamilia/>

direta de renda (em especial o Bolsa Família) que atraiu o eleitorado de baixa renda para o Presidente Lula nas eleições de 2006.

Levando em conta também as denúncias de corrupção sobre o governo Lula, desde o início do episódio do “mensalão”⁵³ até o episódio do “dossiê”⁵⁴, no início de setembro de 2006 o Instituto Datafolha fez outra pesquisa questionando a uma amostra de 4.319 eleitores em todo o país o seguinte: “*you believe that there is corruption in the Lula government?*”, e se acredita, “*what is the degree of responsibility of the president?*”. E em seguida questionou o entrevistado sobre sua intenção de voto. Os resultados são apresentados abaixo, na Tabela 2:

Pergunta	“... <i>existe corrupção no governo Lula?</i> ”			“ <i>qual o grau de responsabilidade do presidente?</i> ”			
	Sim	Não	Não Sei	Muito	Pouco	Nenhum	Não sei
Resposta	75%	14%	11%	25%	37%	10%	28%
Razão entre as intenções de voto pró Lula em relação a Alckmin	1,03	1,76	-	0,73	1,20	1,27	-

Fonte: Datafolha.

O resultado da pesquisa apontou que $\frac{3}{4}$ do eleitorado acreditava nas denúncias de corrupção, mas apenas $\frac{1}{4}$ dos que acreditavam associava os episódios ao presidente de forma a inverterem suas intenções de voto de Lula para Alckmin. Até pessoas que respondiam acreditar que o presidente tinha algum grau de responsabilidade com a corrupção (mesmo que pouco) acabavam por dar 20% de vantagem a Lula sobre Alckmin.

Isso levou a formulação da segunda hipótese sobre a inversão da base eleitoral do presidente: as pessoas que estavam direta ou indiretamente sendo beneficiadas pelo Bolsa Família não tinham acesso a informação sobre as denúncias de corrupção, ou tinham e não acreditavam, ou tinham e não se importavam se eram ou não verdadeiras (pois vinham sendo beneficiadas pelo programa, como nunca antes foram na história do país). Em contraponto, os escândalos de corrupção afastaram de Lula o eleitorado mais instruído (no mínimo pela dúvida acerca da veracidade das denúncias).

Alguns analistas passaram a denominar isso por “hipótese do voto econômico”⁵⁵, mesclando o que se chamou de *hipóteses 1 e 2* apresentadas até aqui, qual seja: os mais pobres tiveram melhorias econômicas durante o primeiro mandato de Lula, e era só

⁵³ Esquema de compra de votos no Congresso Nacional denunciado pela revista *Veja* em 2005.

⁵⁴ Em setembro, Valdebran Padilha da Silva, do PT do Mato Grosso, e Gedimar Passos foram presos, em São Paulo, sob suspeita de intermediar a compra de documentos que mostrariam o suposto envolvimento de Serra e Alckmin com a “máfia dos sanguessugas” (esquema de corrupção em torno do superfaturamento de ambulâncias, envolvendo órgãos do governo federal, deputados federais, prefeituras e empresas). Com eles a polícia apreendeu cerca de R\$ 1,7 milhão. Gedimar afirmou à Polícia Federal que foi “contratado pela Executiva Nacional do PT” para negociar com a família Vedoin (que comandava o esquema dos “sanguessugas”) a compra de um dossiê contra os tucanos. Como resultado, foram afastados de seus cargos o presidente do PT (e coordenador da campanha do presidente Lula) Ricardo Berzoini e vários membros do PT e assessores diretos do presidente, além do coordenador de campanha de Aluísio Mercadante, candidato ao governo de São Paulo pelo PT.

⁵⁵ Ver detalhes em Carreirão (2007).

isso que lhes importava. Ao passo que os escândalos de corrupção só incomodavam a classe média.

A terceira hipótese da inversão da base eleitoral é mais amigável a biografia do presidente Lula, postulando que foram seus bons resultados nos âmbitos econômico e social que o conduziram a reeleição. De fato, uma rápida análise de alguns indicadores econômicos mostra, por exemplo, que no período do primeiro mandato as exportações bateram recordes históricos, de forma que a balança comercial, superavitária, foi suficiente, desde o primeiro ano, para registrar superávit na conta de transações correntes.

A melhora do desempenho dessa conta (US\$ 4,2 bilhões, US\$ 11,8 bilhões e US\$ 14,2 bilhões de 2003 a 2005, respectivamente), além de indicar que o saldo da balança comercial era mais do que suficiente para financiar o serviço da dívida, mostrava, ao mesmo tempo, menor dependência do país ao afluxo de capitais externos para zerar o balanço de pagamentos. Ao mesmo tempo, o volume da dívida externa pública e privada reduziu-se de US\$ 210,7 bilhões em dezembro de 2002 para US\$ 169,5 bilhões em dezembro de 2005 (de 45,9% do PIB para 21,3% do PIB, respectivamente, a menor registrada de 1975).

Além disso, ao final de 2005, o governo Lula pagou antecipadamente uma dívida (US\$ 15,5 bilhões) junto ao Fundo Monetário Internacional (FMI), que teria vencimento apenas em 2007. Isso rendeu várias manchetes internacionais a favor do governo.

No plano interno, a taxa de desemprego calculada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que estava em 10,5% em dezembro de 2002, caiu para 8,3% em dezembro de 2005. A taxa básica de juros anual (Selic), de 25% em dezembro de 2002, foi reduzida para 18,0% em dezembro de 2005 e para 15,75% em abril de 2006. Isso entre outros indicadores positivos que poderiam ser citados aqui para sustentar essa hipótese.

Todavia, nem o Bolsa Família e nem os bons resultados da política econômica foram suficientes para referendar a reeleição logo no primeiro turno. Quando o presidente Lula recebeu 48,60% dos votos válidos (Tabela 3, abaixo), algo surpreendente, pois a maioria das pesquisas de “boca de urna” apontava uma vitória sem necessidade de segundo turno⁵⁶.

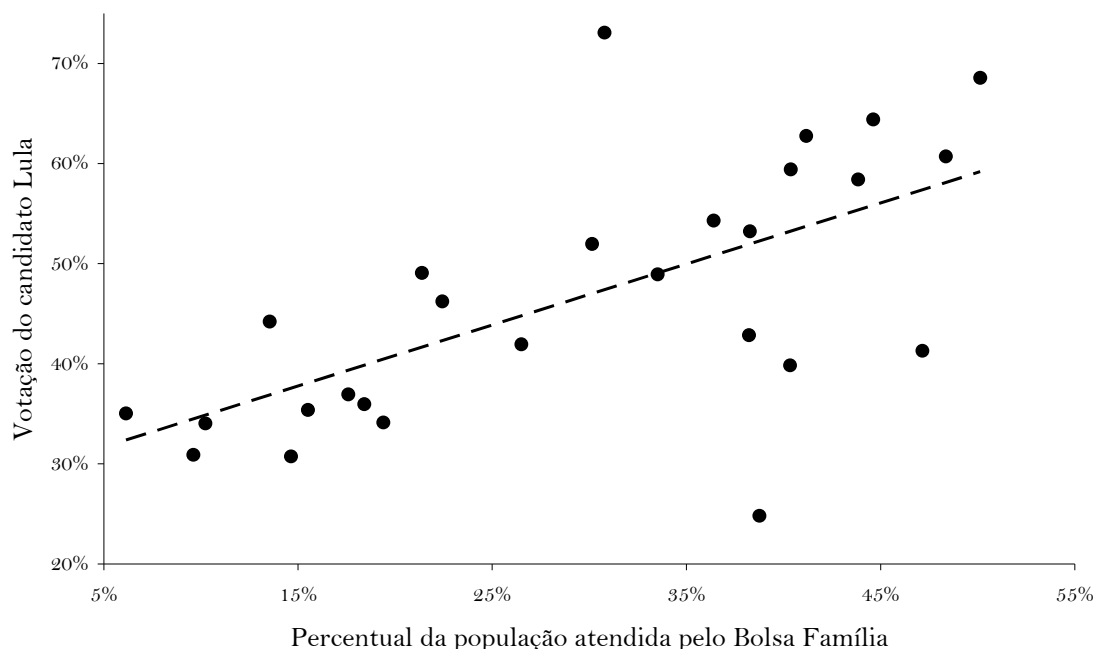
Candidato	Número de votos (milhões)	% dos votos totais	% dos votos válidos
Lula	46,66	44,51	48,60
Alckmin	39,97	38,13	41,64
Heloísa Helena	6,58	6,28	6,85
Cristóvão Buarque	2,54	2,42	2,65
Outros candidatos	0,25	0,24	0,26
Branco e Nulos	8,82	8,41	-
Total	104,82	100,00	100,00

Fonte: Tribunal Superior Eleitoral (TSE).

⁵⁶ A pesquisa do Datafolha, por exemplo, apresentava que a votação seria de 58%. Errou em quase 10%.

O Gráfico 2 apresenta a dispersão entre a votação do candidato Lula no primeiro turno de 2006 e o percentual da população atendida pelo Bolsa Família por UF em 2006.

Gráfico 2 – Dispersão entre a votação do candidato Lula no primeiro turno de 2006 e o percentual da população atendida pelo Bolsa Família por UF em 2006



Fonte: TSE e MDS.

E foi com o apelo de um gráfico semelhante ao exposto acima que dois pesquisadores estrangeiros, Wendy Hunter (Universidade do Texas) e Timothy Power (Universidade de Cambridge, Inglaterra), desencadearam toda uma discussão na academia brasileira sobre se o Bolsa Família foi ou não foi fundamental para a reeleição e a inversão da base eleitoral do presidente Lula em detrimento aos escândalos de corrupção e outras possibilidades.

A base da análise de Hunter e Power (2007) é uma seqüência de regressões simples cuja variável dependente é a variação da votação de Lula entre as eleições de 2006 e 2002 *versus* uma série de indicadores econômicos e sociais das UFs (as variáveis explicativas). Nas conclusões, eles foram enfáticos em argumentar que se o presidente Lula obteve uma maior aprovação nas regiões mais urbanizadas e desenvolvidas do Brasil em 2002, em 2006 sua base eleitoral migrou para as regiões menos desenvolvidas por conta do Bolsa Família e da indiferença das camadas sociais menos favorecidas em relação aos escândalos de corrupção.

Como relatou Abensur *et al* (2007), isso parece ter sido interpretado de forma distinta de acordo com as simpatias eleitorais de cada um. Para os simpatizantes do presidente Lula, tal padrão seria o resultado de um governo voltado para os cidadãos mais pobres do Brasil. Para outros, isso seria uma conseqüência de políticas paternalistas e eleitoreiras. Houve também aqueles que levantaram a possibilidade de ter acontecido uma dicotomia Norte-Sul, pois, supostamente, o primeiro mandato de Lula teria beneficiado muito pouco os estados das regiões Sudeste e Sul.

Abensur *et al* (2007) foram os primeiros a apontar que a análise de Hunter e Power (e a de outros estudos iniciais que surgiram logo em seguida, e foram posteriormente publicados) poderia estar falhando no aspecto metodológico. Pois a variável dependente pertencia a intervalos limitados, e o uso de mínimos quadrados não é adequado para esses casos, pois pode render valores ajustados que excedem os limites dos intervalos. Eles então ajustaram regressões semelhantes as de Hunter e Power pela metodologia *Beta* (de variável dependente censurada), mas não rejeitaram a hipótese de que o Bolsa Família foi fundamental para a reeleição.

Na seqüência da discussão, Soares e Terron (2008) aplicaram técnicas de econometria espacial para estudar o tema (com bases de dados em nível de municípios). E concluíram: “*Os resultados confirmam ainda que a participação do Bolsa Família sobre a renda local foi um determinante do novo contorno das bases geoeleitorais, e o fator de maior peso na explicação da votação municipal*”.

Na contramão dessa frase, Shikida *et al* (2009), também usando técnicas de econometria espacial e base de dados semelhante, concluíram: “*Nevertheless, the Bolsa Família Programme showed some evidence of positive impact on the election, but the results did not appear to be robust. Even if it were significant, the value of the estimator would be much smaller than necessary to explain Lula’s victory*”.

Já Zucco (2008) apontou evidências de que no Brasil o candidato a presidente ligado ao governo sempre obteve desempenho eleitoral relativamente melhor em regiões menos desenvolvidas, porque naturalmente essas regiões são mais dependentes do governo federal (isso deu margem a formulação da *hipótese 4* apresentada no início deste ensaio). Tal regularidade explicaria o desempenho de Lula nas regiões mais pobres, e o autor conclui: “*Nevertheless, adopting a more optimistic perspective, a programme such as the Bolsa Família, with its objective eligibility criteria and universal scope, can be interpreted as a sign of the increasing responsiveness of the government to the median voter, rather than as simply old fashioned clientalism*”.

Em outros estudos, Marques *et al* (2009) concluíram: “*Foi determinante para o resultado do pleito presidencial de 2006 o fato de o programa Bolsa Família abranger mais de 11 milhões de famílias e beneficiar quase 26% da população*”. E Licio *et al* (2009), na mesma linha, concluíram: “*Os resultados indicam um forte impacto de ser beneficiário do programa no voto em Lula e em avaliações positivas de seu desempenho*”.

Enfim, Zucco (2008) e Shikida *et al* (2009) encontram resultados que amenizam o suposto impacto populista na mudança no perfil na votação de Lula entre 2006 e 2002 por conta do Bolsa Família. Mas os outros autores são enfáticos nas conclusões de o programa foi determinante para isso.

Já Canêdo-Pinheiro (2009) observou que a discussão poderia estar carregando dois problemas metodológicos: 1) o uso inadequado de modelos lineares (já observado por Abensur *et al*, 2007); e, 2) conclusões distorcidas pelo *viés de agregação*. E para o *viés de agregação* ninguém teria atentado ainda, e é isso que poderia estar gerando os resultados díspares.

Detalhes sobre o *viés de agregação* serão apresentados na próxima sessão, aqui se descreverá brevemente a estratégia proposta por Canêdo-Pinheiro para se estudar o tema. Qual seja: usar uma variante do *Mixed LOGIT* como fizeram Heckelman (1997) para estudar os pleitos nos Estados Unidos e Glasgow (2001) na Inglaterra.

A idéia geral destes trabalhos é que cada eleitor, em cada localidade do país, tem uma utilidade desenhada pelas atitudes do governo que refletiram nele, na sua família e na sua localidade. A utilidade é uma variável latente, que quando ultrapassa determinado limite (diferente entre os eleitores) faz como que o indivíduo esteja ou não propenso a votar em determinado candidato. Mas essa propensão a votar depende de muitos fatores idiossincráticos, e um deles, no caso brasileiro, seria a opinião sobre o

programa Bolsa Família (algo como a seqüência de perguntas mostrada anteriormente na Tabela 1).

Usando essa estrutura, Canêdo-Pinheiro promoveu um exercício empírico usando dados municipais e conclui seu trabalho da seguinte forma: “*O que explicaria a mudança no padrão de votação de Lula nas eleições presidenciais de 2006: desempenho da economia ou a massificação do programa Bolsa Família? Ao contrário de boa parte da literatura anterior, a resposta é: nenhum dos dois. O impacto do programa de transferência de renda se mostrou bastante superior ao do crescimento da economia, mas parece não ser a principal explicação para a dramática migração de votos em direção do candidato do PT nas regiões menos desenvolvidas do Brasil*”. E, por fim, a conclusão de Canêdo-Pinheiro se alinha melhor com a conclusão de Zucco: quem é governo sempre terá a vantagem eleitoral sobre a população mais dependente do governo.

A análise que irá se propor adiante vai além do proposto por Canêdo-Pinheiro no seguinte sentido: ali o eleitor escolhia apenas entre votar ou não no candidato Lula, aqui a metodologia permitirá a escolha entre todos os candidatos ou invalidar o voto⁵⁷. Mais ainda, com uma álgebra simples se mostrará que é possível calcular quantos votos podem valer uma Bolsa Família.

3.3 A estratégia empírica

A estratégia empírica deste ensaio consiste em observar como o *viés de agregação* poderia estar gerando os resultados díspares nessa literatura, adaptar o *Mixed LOGIT* para contornar o problema e estimá-lo para inferir como possíveis mudanças no número de Bolsas Família distribuídas no primeiro mandato do presidente Lula poderiam afetar os resultados do pleito de 2006.

3.3.1 A questão do *viés de agregação*

A questão do *viés de agregação* é bem documentada em Stoker (1993). Com base em vários exemplos que o autor apresenta de como o viés pode se manifestar, aqui se adaptará uma de suas ilustrações. Para tanto, considere que em uma eleição existem N eleitores, indexados por i , e C candidatos, indexados por j . A opção $j = 0$ é invalidar o voto (votar branco ou nulo).

Considere também que cada eleitor i possui uma função utilidade por cada opção j , $U_{i,j}$. Essa utilidade é bastante abstrata, e trata-se de uma função de inúmeras opiniões do eleitor acerca das características observadas dos candidatos, um vetor $\zeta_{i,j}$.

Considere $I(\cdot)$ uma função indicador, que pode assumir os valores 0 ou 1. $I(\cdot) = 1$ indica escolher determinada opção. Então defina o número total de votos recebidos por

⁵⁷ Observe que a estrutura se repete: no primeiro ensaio os consumidores escolhiam modelos de automóveis a comprar; e no segundo ensaio as firmas escolhiam trabalhadores a contratar. Aqui os eleitores escolhem candidatos a votar.

um candidato j em uma eleição por⁵⁸: $V_j = \sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | r = 0, 1, \dots, j, \dots, C)$. E depois defina o total de votos válidos: $V = \sum_{j \neq 0} V_j$.

Pela regra eleitoral, computa-se a seguinte razão de votos válidos para um determinado candidato j :

$$(1) \quad v_j = \frac{\sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | r = 1, \dots, j, \dots, C)}{\sum_{j=1}^C \sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | r = 1, \dots, j, \dots, C)}$$

Se $v_j \geq 0,5$ o candidato j vence em primeiro turno.

A observação de Canêdo-Pinheiro (2009) é que os trabalhos discutidos na seção anterior usam uma econometria que é alguma variante de:

$$(2) \quad \bar{v}_j = \sum_{k=1}^K \beta_k X_{k,j} + u_j$$

Onde: \bar{v}_j é a razão de votos válidos do candidato j (nas UFs ou municípios) ou a diferença dessa razão entre as eleições de 2006 e 2002; os β 's são parâmetros e os X 's são variáveis explicativas (K é o número total das variáveis explicativas); e u é o erro.

Considere o caso simples onde em (2) está se modelando $\bar{v}_j = v_j$. O lado esquerdo de (1) e (2) é conhecido (é publicado pelo TSE, e pode ser desagregado até as zonas eleitorais). Enquanto se deseja estimar o lado direito das igualdades para se fazer inferência.

O que levou o lado esquerdo acontecer, em ambas as equações, foi o comportamento conjunto de percepções individuais (desagregadas) de todos os eleitores sobre os candidatos, os $U_i(\zeta_{i,j})$, que é desconhecido pelo pesquisador. Mas suponha que pelo menos o pesquisador conhece a distribuição conjunta dos termos idiossincráticos, dada por $F(\zeta)$, então se poderia estimar o seguinte resultado sem agregação:

$$E(v_j) = \int_{\zeta_1} \dots \int_{\zeta_N} \frac{\sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | r = 1, \dots, j, \dots, C)}{\sum_{j=1}^C \sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | r = 1, \dots, j, \dots, C)} \cdot dF(\zeta)$$

Todavia, ao se promover um exercício econométrico com dados agregados para a estrutura (2), o pesquisador está estimando o seguinte resultado com agregação:

$$E(v_j) = \sum_{k=1}^K \beta_k X_{k,j}$$

Então, uma forma de definir o *viés de agregação* é:

$$(3) \quad \sum_{k=1}^K \beta_k X_{k,j} - \int_{\zeta_1} \dots \int_{\zeta_N} \frac{\sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | \dots)}{\sum_{j=1}^C \sum_{i=1}^N I(U_i(\zeta_{i,j}) \geq U_i(\zeta_{i,r}) | \dots)} \cdot dF(\zeta)$$

⁵⁸ Por simplicidade de exposição, assume-se $U_i(\zeta_{i,j}) \neq U_i(\zeta_{i,r})$ para todo $j \neq r$.

Como Stoker mostra, são raros os casos onde esse viés pode ser medido com precisão (ou sua direção pode ser determinada). Mas o uso de dados mais desagregados (no limite: micro-dados) ou métodos de simulação podem mitigar a diferença (3).

Glaeser, Sacerdote e Scheinkman (2003) mostram que o *viés de agregação* pode ser visto como um efeito das interações sociais cuja percepção, por parte do pesquisador, se perde na agregação dos dados. Eles mostraram, em uma seqüência de três exercícios simples, que na possibilidade de existirem interações sociais, em que as ações de uma pessoa influenciam as ações de outra, as inferências dos β 's com dados agregados em estruturas semelhantes a (2) podem ser inadequadas.

No primeiro exercício, eles estudam como a agregação pode gerar um *viés* nos parâmetros estimados de um modelo observando os dados de colegas de quarto da Universidade da Califórnia. Em uma primeira regressão (por mínimos quadrados), usaram como variável dependente o número de colegas de quarto que participavam de uma fraternidade, e como variáveis explicativas: o número deles que afirmavam já beber cerveja na escola; uma *dummy* para dormitórios masculinos; a renda familiar média dos colegas de quarto; e outros controles para as suas notas. Na segunda regressão usam a mesma estrutura, mas fazendo a agregação por andar do dormitório. E na terceira fazem a agregação por prédio inteiro. Os resultados mostraram que conforme o nível de agregação aumenta, os parâmetros estimados se modificam em magnitudes significativas.

Nos outros dois exercícios, Glaeser, Sacerdote e Scheinkman usam bases de dados sobre criminalidade e mercado de trabalho. E mais uma vez observam que diferentes níveis de agregação dos dados levam a diferentes magnitudes nos parâmetros estimados.

Os resultados de uma eleição tendem a refletir certo nível de interação social (basta observar os dados da Tabela 1 para notar isso, apresentada anteriormente). Assim, a inferência sobre os dados agregados deve possuir um viés sobre os parâmetros estimados semelhantes aos documentados por Glaeser, Sacerdote e Scheinkman, o qual o pesquisador não conhece nem o tamanho e nem a direção. Por consequência, é arriscado, por exemplo, afirmar que o Bolsa Família foi ou não foi fundamental para a reeleição e a inversão da base eleitoral do presidente Lula nas últimas eleições com base em regressões sobre dados agregados.

Enfim, como bem argumentam Canêdo-Pinheiro (2009), Glasgow (2001) e Heckelman (1997), quando o voto é secreto se inviabiliza a construção de uma base de micro-dados para mitigar o *viés de agregação*. Logo, possivelmente a melhor estratégia para se estimar efeitos de determinadas variáveis sobre os resultados de uma eleição seja a simulação.

3.3.2 O papel do *Mixed* LOGIT

O papel do *Mixed* LOGIT será o de, primeiro, promover uma estimação do lado direito da equação (1)⁵⁹ e, segundo, buscar a resposta da questão que intitula este ensaio. O lado esquerdo da equação (1), é claro, será votação de um determinado candidato, mas controlando-se pela localidade.

O controle por localidade se faz necessário porque o objetivo do trabalho é entender o impacto do Bolsa Família na votação dos candidatos, e o número de Bolsas Família muda entre as localidades. Então, intrinsecamente, está se admitindo que o

⁵⁹ A esperança dela, para ser mais preciso.

efeito do programa sobre a votação dos candidatos é diferente entre diferentes localidades. Por conseqüência, mais do que se estimar quantos votos pode valer uma Bolsa Família, se estimará quantos votos pode valer em determinada localidade⁶⁰.

Para a estimação do lado direito da equação (1), propõe-se trabalhar com a seguinte função utilidade:

$$(4) \quad U_{i,j,\ell} = \beta_{0,\ell} + \beta_{1,i} \cdot M_j + \beta_{2,i} \cdot G_{j,\ell} + \beta_{3,i} \cdot B_\ell + \beta_{4,i} \cdot D_{Lula} \cdot B_\ell + \beta_{5,i} \cdot PIB_\ell + \xi_{j,\ell} + \varepsilon_{i,j,\ell}$$

Onde: $U_{i,j,\ell}$ é a medida latente do eleitor i em relação ao candidato j (entre J opções) dado que ele mora na localidade ℓ ; $j = 0$ é optar por votar branco ou nulo; $\beta_{0,\ell}$ é constante entre os eleitores da localidade ℓ , mas muda entre as localidades – esta estrutura é conseqüência de um painel que se montará entre votações e localidades, então $\beta_{0,\ell}$ será uma constante de efeito fixo⁶¹; $\beta_{1,i}, \dots, \beta_{5,i}$ são valorações individuais de cada eleitor; M_j é o orçamento de campanha do candidato j , uma medida para a força de sua propaganda eleitoral⁶²; $G_{j,\ell}$ é uma *dummy* que indica 1 se o governo da localidade ℓ é da base aliada do candidato j (e 0 no caso contrário); B_ℓ é o percentual da população atendida pelo Bolsa Família na localidade ℓ e D_{Lula} é uma *dummy* para o candidato Lula (a razão da *dummy* ficará clara adiante, quando se computar a derivada entre votação e Bolsa Família.); PIB_ℓ é o PIB per capita da localidade, uma tentativa de se controlar a renda média do eleitorado e seu reflexo na tomada de decisão; $\xi_{j,\ell}$ é um índice que resume as outras características do candidato j percebidas pelos eleitores da localidade, e não percebidas pelo pesquisador; e, por fim, $\varepsilon_{i,j,\ell}$ é o termo idiossincrático i.i.d. de Valor Extremo do Tipo I necessário para operacionalizar o *Mixed* LOGIT.

Como nos ensaios anteriores, assume-se que $\beta_{1,i}, \dots, \beta_{5,i}$ seguem distribuição $N(\bar{\beta}_k, \sigma_k)$, para $k = 1, \dots, 5$, e então se escreve:

$$(5) \quad U_{i,j,\ell} = \underbrace{\beta_{0,\ell} + \bar{\beta}_1 \cdot M_j + \bar{\beta}_2 \cdot G_{j,\ell} + \bar{\beta}_3 \cdot B_\ell + \bar{\beta}_4 \cdot D_{Lula} \cdot B_\ell + \bar{\beta}_5 \cdot PIB_\ell + \xi_{j,\ell}}_{=\delta_{j,\ell}} + \underbrace{\sigma_1 \cdot z_{1,i} \cdot M_j + \sigma_2 \cdot z_{2,i} \cdot G_{j,\ell} + \sigma_3 \cdot z_{3,i} \cdot B_\ell + \sigma_4 \cdot z_{4,i} \cdot D_{Lula} \cdot B_\ell + \sigma_5 \cdot z_{5,i} \cdot PIB_\ell}_{=\mu_{i,j,\ell}} + \varepsilon_{i,j,\ell}$$

⁶⁰ Talvez se pudesse pensar nisso em dois tipos de desagregação. A primeira, por conta da simulação, desagrega os dados por eleitores (simulados). A segunda, por conta da própria estrutura dos dados, desagrega os eleitores por localidade.

⁶¹ Outra observação: $\beta_{0,\ell}$ funciona como uma referência de utilidade comum entre os moradores de uma dada localidade, e, pela estrutura do *Mixed* LOGIT (sendo de praxe se normalizar a utilidade pela opção externa em zero), quanto menor o $\beta_{0,\ell}$ de uma localidade maior será a probabilidade de um eleitor daquela localidade optar por votar branco ou nulo.

⁶² Na base de dados do TSE, o orçamento de campanha é dado por candidato e não é detalhado por localidade. Por isso essa variável não será indexada por ℓ .

Onde os z_i 's possuem distribuição normal padrão, $\delta_{j,\ell}$ é uma medida comum da opção j a todos os eleitores da localidade ℓ e $\mu_{i,j,\ell}$ é uma medida da opção j que difere entre os eleitores i da localidade ℓ , mas pode ser estimada pelo pesquisador.

Normalizando a opção de invalidar o voto em $\delta_{0,\ell} + \mu_{i,0,\ell} = 0$, chega-se na probabilidade de que um dado eleitor i opte pela opção j dado está na localidade ℓ :

$$P_{i,j,\ell} = \frac{\exp(\delta_{j,\ell} + \mu_{i,j,\ell})}{1 + \sum_{r=1}^J \exp(\delta_{r,\ell} + \mu_{i,r,\ell})}$$

Por conseqüência, dada a distribuição de probabilidade $F(z_\ell)$ e tomando-se $P_{j,\ell}$ por $v_{j,\ell}$, a votação do candidato j na localidade ℓ será dada por⁶³:

$$(6) \quad v_{j,\ell} = \int_{z_\ell} P_{i,j,\ell} \cdot F(z_\ell)$$

Então a estrutura (6) é usada para se inferir (1). E, para buscar a resposta da questão que intitula este ensaio, note entender o que levou cada candidato ter uma maior (ou menor) votação por conta do Bolsa Família, em cada localidade ℓ , passa pelo computo da seguinte derivada:

$$(7) \quad \frac{\partial v_{j,\ell}}{\partial B_\ell} = \int_{z_\ell} (\bar{\beta}_3 + \sigma_3 \cdot z_{3,i} + (\bar{\beta}_4 + \sigma_4 \cdot z_{4,i}) \cdot D_{Lula}) \cdot B_\ell \cdot P_{i,j,\ell} \cdot (1 - P_{i,j,\ell}) \cdot F(z_\ell)$$

Agora fica clara a razão da presença da *dummy* D_{Lula} . Se em determinada localidade aumentos do Bolsa Família tendem a gerar aumentos na votação do candidato Lula ocorre: $\frac{\partial v_{Lula,\ell}}{\partial B_\ell} = \int_{z_\ell} (\bar{\beta}_3 + \sigma_3 \cdot z_{3,i} + \bar{\beta}_4 + \sigma_4 \cdot z_{4,i}) \cdot B_\ell \cdot P_{i,Lula,\ell} \cdot (1 - P_{i,Lula,\ell}) \cdot F(z_\ell) > 0$. E esse resultado dependerá fundamentalmente do sinal de $\bar{\beta}_4$ e da magnitude de σ_4 .

Então, dado que o uso do *Mixed LOGIT* deve mitigar o *viés de agregação*, ao se verificar $\frac{\partial v_{Lula,\ell}}{\partial B_\ell} > 0$ em todas as localidades estudadas, isto apresentará indicio de que o Bolsa Família foi fundamental na estruturação da base eleitoral do presidente Lula.

Agora, para se estimar quantos votos podem valer uma Bolsa Família, observe que, dada uma localidade, a derivada (7) pode ser aproximada por:

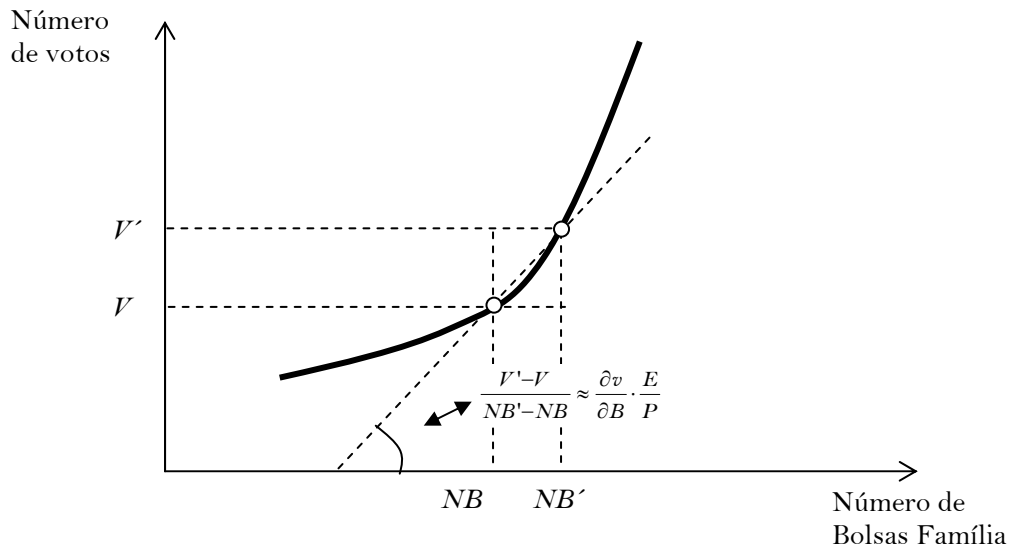
$$(8) \quad \frac{\partial v_j}{\partial B} \approx \frac{\frac{V'_j - V_j}{NB' - NB}}{P} = \frac{V'_j - V_j}{NB' - NB} \cdot \frac{1}{P} \quad \Rightarrow \quad \frac{V'_j - V_j}{NB' - NB} \approx \frac{\partial v_j}{\partial B} \cdot \frac{E}{P}$$

Onde: NB é o número de Bolsas Família (apóstrofe significa o número após uma variação); V é o número de votos recebidos por um determinado candidato; P é a

⁶³ z_ℓ é um vetor de dimensão 4 vezes o número de eleitores da localidade ℓ .

população da localidade; E é o número de eleitores; e, $B = NB/P$ e $v_j = V_j/E$ se observam nos dados do MDS e do TSE, respectivamente⁶⁴.

Assim, $(V'_j - V_j)/(NB' - NB)$ reflete a variação do número de votos do candidato j por unidade de Bolsa Família distribuída. Por consequência, cada Bolsa Família vale (em número de votos) aproximadamente a derivada dada por (7) multiplicada pelo percentual de eleitores da localidade (em relação à população). O diagrama abaixo ilustra a idéia para o caso de uma relação positivamente inclinada.



Mas é importante notar que esta é uma “medida no ponto”, isto é, quando se modifica o número de Bolsas Família distribuídas também se modifica a magnitude da variação do número de votos (podendo ser marginalmente crescente ou decrescente).

Por fim, note que (8) pode assumir valores positivos ou negativos, dependendo do valor que assumir a derivada (7). Se assumir valores negativos significa: votos perdidos por unidade de Bolsa Família existentes na localidade. Caso contrário a interpretação será de votos ganhos.

3.4 A amostra

Para se dar estrutura a especificação (4) utilizou-se as bases do TSE (para os dados de votação, orçamento de campanha etc.) e do MDS (para o percentual da população atendida pelo Bolsa Família em 2006 e o PIB per capita das localidades). Dada a grande carga computacional do *Mixed* LOGIT, optou-se por usar as localidades em nível de UFs mais o exterior⁶⁵ (em detrimento de municípios ou zonas eleitorais, que eram as outras possibilidades). Assim, totalizaram-se 28 localidades.

A Tabela 4, abaixo, mostra que o orçamento de campanha dos dois principais candidatos (no primeiro turno das eleições presidenciais de 2006) foi relativamente

⁶⁴ Note que: $\frac{\partial v}{\partial B} = \frac{\partial(V/E)}{\partial(NB/P)} = \lim_{\Delta(NB/P) \rightarrow 0} \frac{\Delta(V/E)}{\Delta(NB/P)} = \lim_{\Delta(NB/P) \rightarrow 0} \frac{(V' - V)/E}{(NB' - NB)/P}$.

⁶⁵ O número de votos do exterior representou 0,04% da votação total.

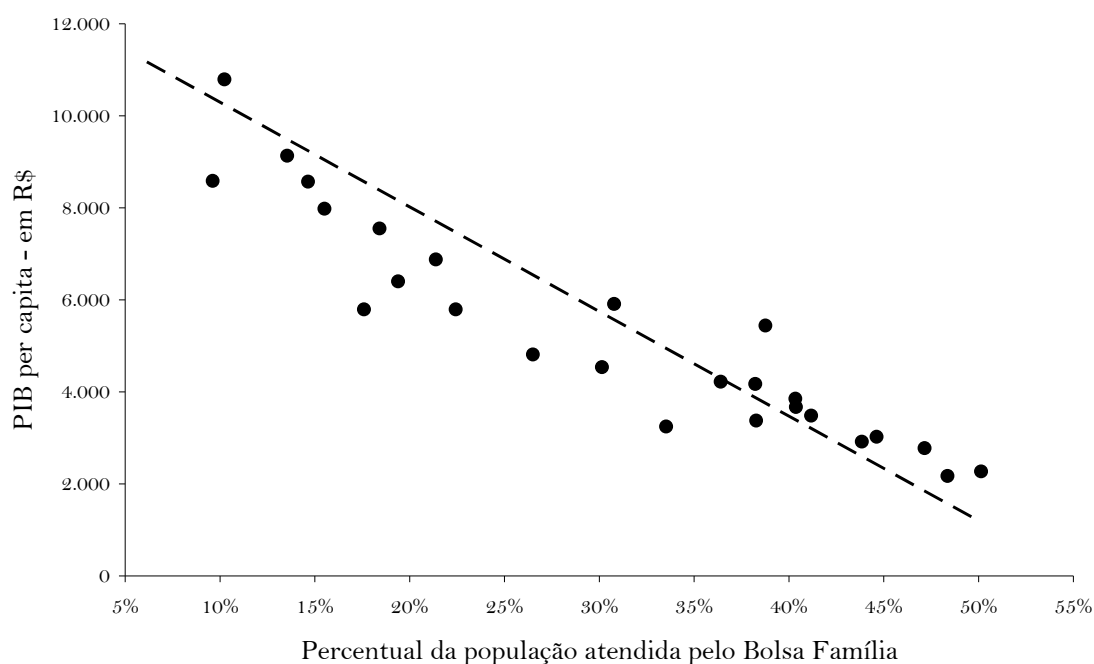
próximo: R\$ 89 milhões o de Lula; e R\$ 85 milhões o de Alckmin. O maior orçamento foi da candidata Ana Maria Rangel (R\$ 150 milhões) e o menor o da candidata Heloísa Helena (R\$ 5 milhões).

Tabela 4 – Orçamento de campanha (R\$ milhões) e Governadores aliados dos candidatos (por UF)		
Candidato	Orçamento	Governadores aliados
Lula	89	AC, MS e PI
Alckmin	85	BA, CE, GO, MA, MG, PA, PB, RO, SE, SP e TO
Heloísa Helena	5	-
Cristóvão Buarque	20	AP
José Maria Eymael	20	-
Luciano Bivar	60	RR
Ana Maria Rangel	150	-

Fonte: TSE.

A mesma tabela também mostra que Alckmin era o candidato com mais governadores aliados (11 no total – segundo a relação do TSE, o que não implicou necessariamente em apoio efetivo por parte destes governadores), enquanto Lula possuía três. Cristóvão Buarque e Luciano Bivar, ambos, possuíam um governador da base aliada. Os demais 11 governadores não participaram explicitamente de qualquer base aliada (muito embora pudessem demonstrar simpatia por um ou outro candidato).

Gráfico 3 – Dispersão entre o PIB per capita e o percentual da população atendida pelo Bolsa Família por UF em 2006



Fonte: MDS.

Já o gráfico 3, acima, apresenta a dispersão entre o PIB per capita e o percentual da população atendida pelo Bolsa Família por UF em 2006. Como era de se esperar, pelo próprio desenho do programa, existe uma relação inversa entre estas variáveis – locais com menores rendas tendem a ter um maior percentual de famílias atendidas pelo programa.

Os mapas apresentados abaixo buscam transmitir uma idéia da dispersão espacial do número de eleitores e da parcela da população atendida pelo Bolsa Família entre as UFs.

No Mapa 1 observa-se (pela pigmentação mais escura) que São Paulo é maior colégio eleitorais do país, com cerca de 24 milhões de eleitores (em 2006). Depois, Minas Gerais (11 milhões), Rio de Janeiro (9 milhões), Bahia e Rio Grande do Sul (ambos com aproximadamente 7 milhões) são as UFs com mais eleitores.

Mapa 1 – Número de eleitores por UF em 2006 – em milhões



Fonte: TSE.

E no Mapa 2 observa-se as maiores parcelas de população atendida pelo Bolsa Família encontram-se no Nordeste (entre 40 e 50%).

necessariamente mais votos. O que é o caso da campanha da candidata Ana Maria Rangel, que teve o maior orçamento da campanha (R\$ 150 milhões) e obteve apenas 126 mil votos (0,13% dos votos válidos).

Tabela 5 – Parâmetros estimados da especificação (5) para 1.000 eleitores simulados

Parâmetro	$\hat{\beta}$	Desvio Padrão	Estatística t	$\hat{\sigma}$	Desvio Padrão	Estatística t
Orçamento da Campanha	- 0,228	0,339	- 0,672	0,324	0,011	30,064
Governador Aliado	2,499	0,580	4,310	1,077	0,014	79,079
Bolsa Família	- 4,072	1,034	- 3,938	0,410	0,013	32,906
$D_{Lula} \times$ Bolsa Família	11,716	1,463	8,009	0,342	0,010	32,794
PIB ⁶⁸	0,869	0,038	22,584	0,594	0,014	44,150

Os números também apontam que se o governo local for da base aliada do candidato isso faz com que esse candidato tenda a ter maiores votações na localidade (o que parece ser bastante intuitivo). Mais ainda, para que um eleitor pondere de forma negativa o fato de o governo local ser da base aliada do candidato é necessário um $z_{2,i}$ de no mínimo $-2,320$. O que, pelas características da distribuição normal padrão, acontecerá em menos de 1% das vezes.

Agora observe os parâmetros associados ao Bolsa Família, e, por simplicidade de entendimento, pense no caso $z_{3,i} = z_{4,i} = 0$. Este caso reflete a ponderação média do eleitor dada por: $\beta_{3,i} + \beta_{4,i} \approx 11,7 - 4,1 = 7,6$ associada ao candidato Lula; e $\beta_{3,i} \approx -4,1$ associada aos outros candidatos. Ou seja, na média os eleitores dão ponderações positivas ao Bolsa Família associando o programa a Lula e ponderações negativas associando o programa aos outros candidatos.

Além disso, também note que o parâmetro de ponderação média associado ao PIB per capita é positivo, isto, pela estrutura de probabilidades derivada do *Mixed* LOGIT, indica que regiões com maiores níveis de renda apresentarão menores níveis de abstenção (ou, talvez, maiores níveis de conscientização política).

Interpretações mais profundas que a posta acima podem ser bastante complexas de se fazer, pois dependem da amplitude dos $\hat{\sigma}$'s e da interação dos pesos que todos os eleitores dão a todos os candidatos.

Para contornar essa complexidade e gerar uma medida de fácil interpretação, a Tabela 6, abaixo, apresenta os resultados estimados para a equação (8). Isto é, apresenta a estimativa de quantos votos cada um dos quatro principais candidatos

⁶⁸ Regredido em R\$ 1.000, pelo mesmo motivo apresentado na nota rodapé anterior.

ganhou ou perdeu durante o primeiro turno das eleições de 2006 por conta do Bolsa Família por UF (note que existe uma ordenação do maior para o menor número de votos ganhos por Lula). Possivelmente esta seja a interpretação mais simples e geral e que possa se demandar do *Mixed* LOGIT nesse tipo de ensaio.

Tabela 6 – Resultados estimados para a equação (8) – o valor de cada Bolsa Família em termos de votos por UF				
UF	Lula	Alckmin	Heloisa Helena	Cristovam Buarque
PB	0,42	- 0,21	- 0,10	- 0,02
RN	0,42	- 0,16	- 0,04	- 0,02
PI	0,41	- 0,18	- 0,04	- 0,02
AL	0,40	- 0,19	- 0,02	- 0,02
PE	0,38	- 0,15	- 0,03	- 0,02
CE	0,37	- 0,12	- 0,03	- 0,02
MA	0,37	- 0,13	- 0,03	- 0,02
SE	0,34	- 0,17	- 0,05	- 0,01
BA	0,33	- 0,13	- 0,03	- 0,02
TO	0,31	- 0,15	- 0,02	- 0,01
AC	0,29	- 0,17	- 0,03	- 0,01
AP	0,28	- 0,13	- 0,05	- 0,02
PA	0,27	- 0,18	- 0,08	- 0,02
RR	0,26	- 0,12	- 0,03	- 0,01
MG	0,22	- 0,11	- 0,03	- 0,02
ES	0,21	- 0,11	- 0,03	- 0,01
RO	0,21	- 0,10	- 0,03	- 0,01
AM	0,18	- 0,06	- 0,03	- 0,01
MT	0,17	- 0,10	- 0,02	- 0,01
MS	0,15	- 0,09	- 0,02	- 0,01
RJ	0,15	- 0,09	- 0,02	- 0,01
PR	0,15	- 0,08	- 0,02	- 0,01
GO	0,14	- 0,06	- 0,04	- 0,01
RS	0,14	- 0,09	- 0,02	- 0,01
SP	0,09	- 0,06	- 0,01	- 0,01
SC	0,09	- 0,05	- 0,01	0,00
DF	0,06	- 0,03	- 0,02	- 0,01

Pelos resultados apresentados acima, nos estados do Nordeste cada Bolsa Família gerou algo entre 0,3 e 0,4 voto para Lula. E perdas de algo entre 0,1 e 0,2 voto para Alckmin. Para os outros candidatos, em geral, essa relação se mostrou diminuta.

Dentre os estados com maior número de eleitores (São Paulo, Minas Gerais, Rio de Janeiro, Bahia e Rio Grande do Sul), na Bahia é onde se estimou a maior relação entre votos e Bolsas Família: cada unidade gerou 0,33 voto para Lula. Enfim, os resultados apontam que uma Bolsa Família pode gerar entre 0,1 e 0,4 voto, dependendo da localidade.

3.6 Conclusões

Como discutido anteriormente, nas eleições presidenciais de 2006 ocorreu uma clara mudança no perfil de votação do candidato Lula, quando sua base eleitoral migrou para as regiões menos desenvolvidas do país.

Para explicar esse fenômeno surgiram várias hipóteses, sendo que uma delas vem polarizando a literatura sobre o assunto. Qual seja: o aumento dos programas sociais de transferência de renda (em especial o Bolsa Família) teria atraído o eleitorado de baixa renda em um movimento que se definiu por “voto econômico”.

Buscando testar essa hipótese, o debate acadêmico tem mostrado resultados bastante díspares. Mas, como foi discutido neste ensaio, possivelmente a geração destes resultados díspares estaria no *viés de agregação* dos dados.

Considerando o fato de que o voto é secreto, e conseqüentemente não existem micro-dados para as eleições de forma a se operacionalizar uma econometria tradicional para mitigar o *viés de agregação*, este ensaio procurou ajudar a preencher uma lacuna existente na literatura usando uma estrutura *Mixed LOGIT* (e simular as escolhas dos eleitores). Especificamente, buscou-se responder a seguinte provocativa: por quantos votos uma Bolsa Família poderia ser trocada?

Concluiu-se então que, de fato, existem evidências de que o Bolsa Família afetou significativamente a votação do presidente Lula. Mas possivelmente este impacto tenha sido menor do que se esperava, porque o *viés de agregação* parece superestimar o efeito.

Por fim, estimou-se que cada Bolsa Família teria gerado entre 0,1 e 0,4 voto para Lula. Sendo que os maiores valores foram encontrados na região Nordeste.

CONSIDERAÇÕES FINAIS

Existem duas considerações finais a se fazer: uma em relação à técnica utilizada nesta tese; e outra em relação a avanços na Teoria da Escolha que podem vir a modificar a técnica futuramente.

Em termos gerais, o *Mixed* LOGIT se alinha muito mais com a realidade que os outros modelos de escolha discreta, pois cada indivíduo tem seu próprio julgamento de cada alternativa de escolha. A grande vantagem dessa técnica é possibilidade de identificação das fontes de heterogeneidade de preferências associados com as médias ($\bar{\beta}'s$) e as variâncias ($\sigma's$) dos pesos que cada agente dá as características do objeto de escolha.

Como argumentam Hensher e Greene (2003) e Berry e Pakes (2007), os principais desafios no desenvolvimento da técnica estão em melhorar a previsibilidade dos modelos quando os indivíduos são confrontados com mudanças no “ambiente de decisão”: os atributos das alternativas, a inclusão ou exclusão de alternativas, as características dos tomadores de decisão e outros efeitos contextuais.

Exercícios de mudança no “ambiente de decisão” foram conduzidos nos dois primeiros ensaios dessa tese: onde se acabou com a tributação sobre os automóveis; e onde em cada período surgiam diferentes tipos de postos de trabalho vagos. E, talvez, outro exercício pudesse ter sido feito no terceiro ensaio tentando prever o resultado das eleições se Serra tivesse sido candidato ao invés de Alckmin. Mas para que esse exercício pudesse ser executado seria necessária uma base de dados extremamente rica, e talvez impossível de ser elaborada.

O *Mixed* LOGIT certamente abre novas oportunidades para a investigação dos fenômenos comportamentais. Mas seu poder de contribuir com a Ciência Econômica está intrinsecamente ligado a qualidade e a quantidade das informações disponíveis ao pesquisador. O que, na realidade, é um problema de toda técnica de trabalho empírico.

Sobre os avanços na Teoria da Escolha, Ben-Akiva, McFadden e Train (2002) ponderam que técnicas como o *Mixed* LOGIT são baseadas na maximização direta de utilidade das alternativas. Mas talvez essa abordagem esteja incompleta. Porque quando o indivíduo escolhe algo, ele não está maximizando utilidade diretamente, mas na realidade ele está indiretamente maximizando a projeção da utilidade que ele terá depois de tomar a decisão de escolha. E entender e modelar isso pode ser extremamente complexo, e possivelmente aproximará a Economia e a Psicologia.

Enfim, a parceria entre a Teoria da Escolha e a técnica do *Mixed* LOGIT parece ser promissora, e deverá incorrer numa série de avanços nos próximos anos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abensur, T.; Cribari-Neto, F.; Menezes, T. (2007). **Impactos do Programa Bolsa Família nos Resultados das Eleições Presidenciais no Brasil em 2006**. Anais do XXXV Encontro Nacional de Economia.
- ANFAVEA (2006a). **Indústria Automobilística Brasileira: 50 Anos**. Disponível em: <http://www.anfavea.com.br/50anos.html>.
- ANFAVEA (2006b). **Novo Perfil da Siderurgia no Brasil**. Seminário apresentado pelo presidente Rogelio Golfarb em São Paulo, dia 21 de março de 2006.
- ANFAVEA (2009). **Anuário da Indústria Automobilística Brasileira edição 2009**. Disponível em: <http://www.anfavea.com.br/anuario.html>.
- Asano, Seki; Barbosa Ana Luiza N. H.; Fiúza, Eduardo P. S. (2004). **Optimal Commodity Taxes for Brazil Based on AIDS Preference**. RBE, Rio de Janeiro, 58(1):5-21. JAN/MAR 2004.
- Ballester, C.; Calvó-Armengol, A.; Zenou, Y. (2006). **Who's Who in Networks. Wanted: The Key Player**. Econometrica, Vol. 74, No. 5 (Sep. 1, 2006), pp. 1403-1417.
- Bartlett, R.; Miller, T. (1985). **Executive Compensation: Female Executives and Networking**. The American Economic Review, Vol. 75, No. 2 (May, 1985), pp. 266-270.
- Ben-Akiva, M.; McFadden, D.; Train, K. (2002). **Hybrid Choice Models: Progress and Challenges**. Marketing Letters, 13:3, pp. 163-175.
- Berry, S. (1994). **Estimating Discrete-Choice Models of Product Differentiation**. Rand Journal, 25(2), pp. 242-262.
- Berry, S., Levinsohn, J., and Pakes, A. (1995). **Automobile Prices in Market Equilibrium**. Econometrica, 63(4), pp. 841-890.
- Berry, S., Pakes, A. (2007). **The Pure Characteristics Demand Model**. International Economic Review, Vol. 48, No. 4, November 2007.
- Blanchard, O.; Diamond, P.; Hall, R.; Yellen, J. (1989). **The Beveridge Curve**. Brookings Papers on Economic Activity, Vol. 1989, No. 1, pp. 1-76.
- Blanchard, O.; Portugal, P. (2001). **What Hides behind an Unemployment Rate: Comparing Portuguese and U.S. Labor Markets**. The American Economic Review, Vol. 91, No. 1, pp. 187-207.
- Canêdo-Pinheiro, M. (2009). **Bolsa família ou desempenho da economia? Determinantes da reeleição de Lula em 2006**. Versão preliminar disponível para *download* em: <http://epge.fgv.br/pt/files/BolsaFam%C3%ADlia30Julho.pdf>.
- Calvó-Armengol, A. (2004). **Job contact networks**. Journal of Economic Theory, 115, pp. 191-206.

- Calvó-Armengol, A.; Jackson, M. (2007). **Networks in labor markets: Wage and employment dynamics and inequality**. *Journal of Economic Theory*, 132, pp. 27–46.
- Calvó-Armengol, A.; Patacchini, E.; Zenou, Y. (2009). **Peer effects and social networks in education**. *Review of Economic Studies*, forthcoming. Disponível para download em: http://www.ifn.se/web/Final_version_RESTUD.aspx.
- Calvó-Armengol, A.; Zenou, Y. (2005). **Job matching, social network and word-of-mouth communication**. *Journal of Urban Economics*, 57, pp. 500–522.
- Carreirão, Y. (2007). **A eleição presidencial brasileira de 2006: uma análise**. *Política e Sociedade*, v. 6, p. 91-116.
- Corcoran, M.; Datcher, L.; Duncan, G. (1980). **Most workers find jobs through word of mouth**. *Monthly Labor Review*, 103, pp. 33–35.
- Datcher, L. (1983). **The Impact of Informal Networks on Quit Behavior**. *The Review of Economics and Statistics*, Vol. 65, No. 3, pp. 491-495.
- Deaton, A. and J. Muellbauer (1980). **An Almost Ideal Demand System**. *American Economic Review*, 70, 312-326.
- DeNegri, J. (1998). **Elasticidade-Renda e Elasticidade-Preço da Demanda de Automóveis no Brasil**. Texto para Discussão N° 558 do IPEA.
- DeSouza, S. (2008). **Combining Prior Information and Data to Uncover the Parameters from the Random Coefficient Discrete-Choice Demand Model**. LAMES, Rio de Janeiro, 2008.
- FENABRAVE (2009). **Anuário Estatístico de 2008**. Disponível para *download* em www.fenabrave.org.br.
- Ferraz, C.; Fiúza, E.; e Motta, R. (2001). **Medindo os efeitos da regulação ambiental em mercados de oligopólio: o caso da poluição automotiva**. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 31, n. 3, p. 333-356, dez. 2001.
- Fiúza, E. (2002). **Automobile Demand And Supply In Brazil: Effects Of Tax Rebates And Trade Liberalization On Price-Marginal Cost Markups in the 1990s**. Texto para Discussão N° 916 do IPEA.
- Fontaine, F. (2008). **Why are similar workers paid differently? The role of social networks**. *Journal of Economic Dynamics & Control*, 32, pp. 3960–3977.
- Glasgow, G. (2001). **Mixed Logit Models for Multiparty Elections**. *Political Analysis*, v. 9, pp. 116-136.
- Glaeser, E; Sacerdote, B; Scheinkman, J. (2003). **The Social Multiplier**. *Journal of the European Economic Association*, vol. 1, No. 2-3, pp. 345-353.
- Hausman, J., Leonard, G., Zona J. (1994). **Competitive Analysis with Differentiated Products**. *Annales d’Economie et de Statistique*, 34, 159-180.
- Heckelman, J.; Power, T. (1997). **Determining who voted in historical elections: An aggregated Logit approach**. *Social Science Research*, v. 26, pp. 121-134.
- Hensher, D.; Greene, W. (2003). **The Mixed LOGIT model: The state of practice**. *Transportation*, v. 30, pp. 133–176.
- Holzer, H. (1988). **Search method used by unemployed youth**. *Journal of Labor Economics*, 6, pp. 1–20.

- Hunter, W.; Power, T. (2007). **Rewarding Lula: Executive Power, Social Policy, and the Brazilian Elections of 2006**. *Latin American Politics and Society*, v. 49, pp. 1-30.
- Huse, C. & Salvo, A. (2005). **Métodos empíricos em organização industrial e aplicações ao antitruste**. Seminário Estudos em Métodos Quantitativos Aplicados a Defesa da Concorrência e a Regulação Econômica. SDE, IPEA, ANPEC, Brasília.
- Ioannides, Y.; Loury, L. (2004). **Job Information Networks, Neighborhood Effects, and Inequality**. *Journal of Economic Literature*, Vol. 42, No. 4, pp. 1056-1093.
- Ioannides, Y.; Soetevent, A. (2006). **Wages and Employment in a Random Social Network with Arbitrary Degree Distribution**. *The American Economic Review*, Vol. 96, No. 2, pp. 270-274.
- Lancaster, K., (1966). **A New Approach to Consumer Theory**. *Journal of Political Economy*, 74, 132-157.
- Licio, E.; Rennó, L.; Castro, H. (2009). **Bolsa Família e Voto na Eleição Presidencial de 2006: em busca do elo perdido**. *Opinião Pública*, vol. 15, nº 1, Junho, 2009, pp. 31-54.
- Marques, R.; Leite, M.; Mendes, A.; Ferreira, M. (2009). **Discutindo o papel do Programa Bolsa Família na decisão das eleições presidenciais brasileiras de 2006**. *Revista de Economia Política*, v. 29, p. 114-132.
- Mas-Collel, A.; Whinston, M.; e Green, J. (1995). **Microeconomic Theory**. Oxford University Press.
- McFadden, D. (1981). **Econometric Models of Probabilistic Choice**, in C. Manski and D. McFadden (Eds), *Structural Analysis of Discrete Data*.
- McFadden, D. (2001). **Economic Choices**. *The American Economic Review*, Vol. 91, No. 3 (Jun., 2001), pp. 351-378.
- McFadden, D.; Train, K (2000). **Mixed MNL Models for Discrete Response**. *Journal of Applied Econometrics*, Vol. 15, No. 5, pp. 447-470.
- Miller, R. (2000). **The Road Ahead for the U.S. Auto Industry**. Office of Automotive Affairs, International Trade Administration, U.S. Department of Commerce.
- Montgomery, J. (1991). **Social Networks and Labor-Market Outcomes: Toward an Economic Analysis**. *The American Economic Review*, Vol. 81, No. 5, pp. 1408-1418.
- Montgomery, J. (1994). **Weak Ties, Employment, and Inequality: An Equilibrium Analysis**. *The American Journal of Sociology*, Vol. 99, No. 5, pp. 1212-1236.
- Nelder, J., Mead (1965). **A Simplex Method for Function Minimization**. *Computer Journal*, 7, 308-313.
- Nevo A. (2000a) **A Practitioner's Guide to Estimation of Random-Coefficients LOGIT Models of Demand**. *Journal of Economics & Management Strategy*, 9, pp.513-548.
- Nevo A. (2000b). **Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry**. *Rand Journal of Economics*, 31, 395-421
- Nevo A. (2001). **Measuring Market Power in the Ready-to-Eat Cereal Industry**. *Econometrica*, 69(2) pp.307-342.

- Pakes, A. (1994). **Dynamic Structural Models, Problems and Prospects: Mixed Continues Discret Controls and Market Iterations**, em *Advances in Econometrics: The Sixth World Congress of the Econometric Society*, vol II, New York, Cambridge Press, pp. 171-260.
- Petrin, A. (2002). **Quantifying the benefits of New Products: The Case of the Minivan**. *Journal of Political Economy*. 110 (4), pp. 705-729.
- Petrongolo, B.; Pissarides, C. (2001). **Looking into the Black Box: A Survey of the Matching Function**. *Journal of Economic Literature*, Vol. 39, No. 2, pp. 390-431.
- Pissarides, C. (2000). **Equilibrium Unemployment Theory**. The MIT Press, Second Edition.
- Portugal, P. (2003). **Emprego, Salários e Condições de Trabalho**. Em: *Portugal Social*, Instituto Nacional de Estatística, pp. 91-112.
- Rees, A. (1966). **Information Networks in Labor Markets**. *The American Economic Review*, Vol. 56, pp. 559-566.
- Scheinerman, E. (2005). **Mathematics: A Discrete Introduction**. Brooks Cole.
- Shikida, C.; Monasterio, L.; Araujo Jr., A.; Carraro, A.; Damé, O. (2009). **“It is the economy, companheiro!”: an empirical analysis of Lula's re-election based on municipal data**. *Economics Bulletin*, Vol. 29, No. 2, pp. 977-992.
- Small, K.; Rosen, S. (1981). **Applied Welfare Economics with Discrete Choice Models**. *Econometrica*, Vol. 49, No. 1 (Jan., 1981), pp. 105-130.
- Soares, G.; Terron, S. (2008). **Dois Lulas: a geografia eleitoral da reeleição (explorando conceitos, métodos e técnicas de análise geoespacial)**. *Opinião Pública*, v. 14, pp. 269-301.
- Stoker, T. (1993). **Empirical Approaches to the Problem of Aggregation Over Individuals**. *Journal of Economic Literature*, v. 31, p. 1827-1874.
- Topa, G. (2001). **Social Interactions, Local Spillovers and Unemployment**. *Review of Economic Studies*, 68, pp. 261-295.
- Train, K. (2003). **Discrete Choice Methods With Simulation**. Cambridge Univ. Press.
- Wahba, J.; Zenou, Y. (2005). **Density, social networks and job search methods: Theory and application to Egypt**. *Journal of Development Economics*, 78, pp. 443- 473.
- Wooldridge, J. (2002). **Econometric Analysis of Cross Section and Panel Data**. The MIT Press.
- Zucco, C. (2008). **The President's 'New' Constituency: Lula and the Pragmatic Vote in Brazil's 2006 Presidential Elections**. *Journal of Latin American Studies*, v. 40, pp. 29-39.

ANEXO I

A probabilidade de um consumidor i escolher um modelo j em detrimento a qualquer modelo r ou a opção externa é:

$$P_{ij} = \Pr(U_{ij} \geq U_{ir}) = \Pr(V_{ij} + \varepsilon_{ij} \geq V_{ir} + \varepsilon_{ir}) = \Pr(\varepsilon_{ir} \leq V_{ij} - V_{ir} + \varepsilon_{ij}) \quad ; \quad \forall r = 0, \dots, J$$

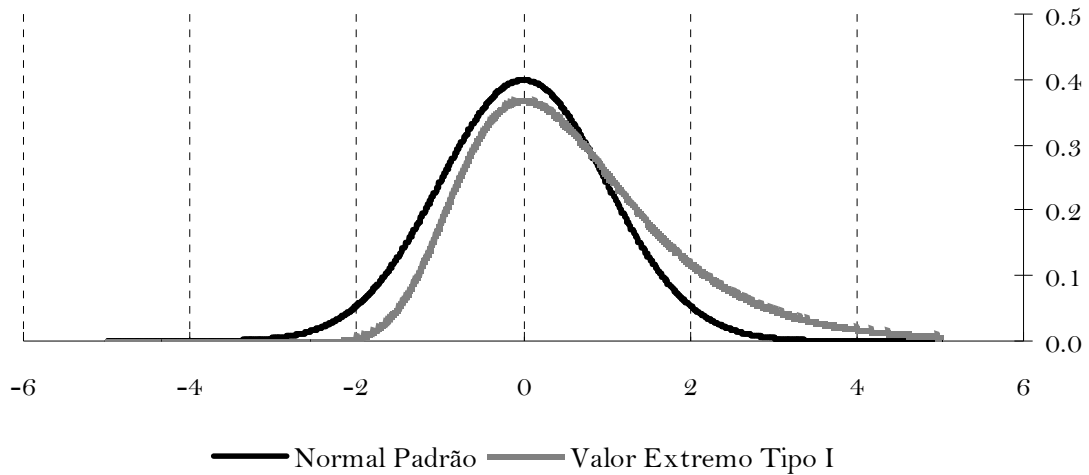
Por hipótese, as funções densidade e cumulativa de probabilidade de ε são:

$$f(\varepsilon_{ij}) = \exp(-\varepsilon_{ij} - \exp(-\varepsilon_{ij})) \quad ; \quad F(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij})) \quad ; \quad -\infty < \varepsilon_{ij} < +\infty$$

Então, pela condição i.i.d.:

$$\begin{aligned} P_{ij} &= \int \int \dots \int_{\varepsilon_{ir} \leq V_{ij} - V_{ir} + \varepsilon_{ij}} \prod_{r=0}^J \exp(-\varepsilon_{ir} - \exp(-\varepsilon_{ir})) \times d\varepsilon_{ir} \\ P_{ij} &= \int_{-\infty}^{+\infty} \left[\exp(-\exp(-\varepsilon_{ij})) \times \prod_{r \neq j} \exp(-\exp(-(V_{ij} - V_{ir} + \varepsilon_{ij}))) \right] \times d\varepsilon_{ij} \\ P_{ij} &= \int_{-\infty}^{+\infty} \left[\exp(-\exp(-\varepsilon_{ij})) \times \exp(-\sum_{r \neq j} \exp(-(V_{ij} - V_{ir} + \varepsilon_{ij}))) \right] \times d\varepsilon_{ij} \\ P_{ij} &= \int_{-\infty}^{+\infty} \left[\exp(-\sum_{r=0}^J \exp(-(V_{ij} - V_{ir} + \varepsilon_{ij}))) \right] \times d\varepsilon_{ij} \\ P_{ij} &= \int_{-\infty}^{+\infty} \left[\exp(-\exp(-\varepsilon_{ij})) \times \sum_{r=0}^J \exp(V_{ir} - V_{ij}) \right] \times d\varepsilon_{ij} \\ P_{ij} &= \frac{\left[\exp(-\exp(-\varepsilon_{ij})) \times \sum_{r=0}^J \exp(V_{ir} - V_{ij}) \right]_{\varepsilon_{ij} \rightarrow -\infty}^{\varepsilon_{ij} \rightarrow +\infty}}{\sum_{r=0}^J \exp(V_{ir} - V_{ij})} \\ P_{ij} &= \frac{(1 - 0)}{\exp(-V_{ij}) \times \sum_{r=0}^J \exp(V_{ir})} \\ P_{ij} &= \frac{\exp(V_{ij})}{\sum_{r=0}^J \exp(V_{ir})} \quad c.q.d. \end{aligned}$$

A figura abaixo sobrepõe as funções densidade de probabilidade de uma Normal Padrão e uma Valor Extremo Tipo I. Note que a Valor Extremo Tipo I não é simétrica, tampouco tem média zero (na realidade tem média 0,5772157).



ANEXO II

Os códigos que seguem foram escritos para o *software* Gauss 6.0 e descrevem apenas os passos 1 a 4 (que é o cerne do programa).

```

/* Declare a matriz x, separando os vetores que se mantêm
constantes no tempo dos que não se mantêm */
xc = missrv(cte~popu,0);
xnc = missrv(flex~hpeso~trend,0);
x = lambda.*(xc~xnc);

/* Declare os parâmetros de base da rotina */
k = cols(x);
n = 1000; /* n é o mesmo que C */
eta = 0.6;
rndseed 7654321;
v = rndn(k,n);
vp = rndn(1,n);
sigma0 = {...}; /* preencha os starts */
alfa0 = ...; /* preencha os starts */
teta0 = sigma0|alfa0;

/* Declare o uso do algoritmo Nelder-Mead */
library nmead;
_nmd_PolyCoef = {0.5, 2, 1};
_nmd_Tol = 1e-5;
_nmd_MaxIters = 5000;
_nmd_Maxtime = 1e+5;
_nmd_IterInfo = 1;
{teta, retcode} = Nmead(&gmm, teta0);

```



```

/* Declare o procedimento que computará os passos 1 a 4 */
proc gmm(teta);
local delta, sigma, sigmam, alfa, e, mu, a, b, omega, ro,
outbase, out05, out06, out07, out08, p05, p06, p07, p08,
omega05, omega06, omega07, omega08, pm05, pm06, pm07, pm08,
oute05, oute06, oute07, oute08, s005, s006, s007, s008, outside,
share, res, mdcs, ti, mdelta, deltap, mxnc, xncp, betanc, c,
qsi, axc, axc08, xc08, betac, qsic, aqsi, aqsi05, aqsi06,
aqsi07, aqsi08, qsi05, qsi06, qsi07, qsi08, az, az05, az06,
az07, az08, z05, z06, z07, z08, cm05, cm06, cm07, cm08, w05,
w06, w07, w08;

```

```

/* Aqui inicia o looping externo */
sigma = teta[1:k,1];
sigmam = ((sigma')*.x)*v;
alfa = teta[k+1,1];
e = exp(alfa*vp).*missrv(p,0);
mu = sigmam-e;

```

```

/* Aqui inicia o looping interno,
o PASSO 1*/
load delta[264,1] = C:\...\delta.txt;
for i(1,1e+1000,1);
a = missrv(lambda.*exp(delta+mu),0);
b = (ones(m,1).*eye(t))'a;
omega = a./(ones(m*t,1)+ones(m,1).*b);
ro = (1/n)*sumc(omega');

```

O estimador de (12):

$$\rho^t = \frac{1}{C} \times \begin{bmatrix} \sum_{i=1}^c \frac{\exp(\mu_{i_1} + \delta_1)}{1 + \sum_{j=1}^n \exp(\mu_{ij} + \delta_i)} \\ \sum_{i=1}^c \frac{\exp(\mu_{i_2} + \delta_2)}{1 + \sum_{j=1}^n \exp(\mu_{ij} + \delta_i)} \\ \vdots \\ \sum_{i=1}^c \frac{\exp(\mu_{i_n} + \delta_n)}{1 + \sum_{j=1}^n \exp(\mu_{ij} + \delta_i)} \end{bmatrix}$$

```

outbase = ano~p~omega;
out05 = delif(outbase,outbase[:,1]./= 2005);
out06 = delif(outbase,outbase[:,1]./= 2006);
out07 = delif(outbase,outbase[:,1]./= 2007);
out08 = delif(outbase,outbase[:,1]./= 2008);
p05 = missrv(out05[:,2],0);
p06 = missrv(out06[:,2],0);
p07 = missrv(out07[:,2],0);
p08 = missrv(out08[:,2],0);
omega05 = out05[:,3:(n+2)];
omega06 = out06[:,3:(n+2)];
omega07 = out07[:,3:(n+2)];

```

```

omega08 = out08[., 3:(n+2)];
pm05    = sumc(p05.*omega05);
pm06    = sumc(p06.*omega06);
pm07    = sumc(p07.*omega07);
pm08    = sumc(p08.*omega08);
oute05  =
(exp(alfa*vp)') .* (pm05.*(1-sumc(missrv(omega05,0))));
oute06  =
(exp(alfa*vp)') .* (pm06.*(1-sumc(missrv(omega06,0))));
oute07  =
(exp(alfa*vp)') .* (pm07.*(1-sumc(missrv(omega07,0))));
oute08  =
(exp(alfa*vp)') .* (pm08.*(1-sumc(missrv(omega08,0))));
s005    = 1-(1/(eta*n))*sumc(oute05);
s006    = 1-(1/(eta*n))*sumc(oute06);
s007    = 1-(1/(eta*n))*sumc(oute07);
s008    = 1-(1/(eta*n))*sumc(oute08);
outside = s005|s006|s007|s008;

```

O estimador da opção externa dado por (15):

$$\hat{s}_0 = 1 - \frac{1}{\eta \times C} \times \sum_{i=1}^C \left(\hat{\alpha} \times u_i \times \left(\sum_{j=1}^J s_{ij} p_j \right) \times \left(1 - \sum_{j=1}^J s_{ij} \right) \right)$$

```

share    = share0.*(1-ones(m,1) .* .outside);
res      = missrv(ln(share)-ln(miss(ro,0)),0);
delta    = missrv(delta,0)+res;
if abs(res) < 1.e-5;
break;
endif;
endfor;
/* Aqui termina o looping interno, o PASSO 1*/

/* Aqui inicia a estimação de qsi, o PASSO 2*/
mdcs     = eye(m) .* .ones(t,1);
ti       = sumc(lambda.*mdcs);
mdelta   = (sumc(lambda.*mdcs.*delta))./ti;
deltap   = delta-(mdelta.*.ones(t,1));
mxnc     = zeros(m,cols(xnc));
xncp     = zeros(rows(xnc),cols(xnc));
for j(1,cols(xnc),1);
mxnc[.,j] = (sumc(lambda.*mdcs.*missrv(xnc[.,j],0)))./ti;
xncp[.,j] = missrv(xnc[.,j],0)-(mxnc[.,j] .* .ones(t,1));
endfor;
betanc   = inv(xncp'xncp)*(xncp'deltap);
c        = mdelta-mxnc*betanc;
qsi      = missrv(delta-lambda.*(xnc*betanc+c.*.ones(t,1)),0);
/* Aqui termina a estimação de qsi, o PASSO 2*/

/* Declare a matriz de instrumentos do PASSO 3, tomando lambda
como um vetor que indica a presença de missings */

```

```

zip = missrv(ip_cte~ip_popu~ip_flex~ip_hppeso~ip_trend,0);
zir = missrv(ir_cte~ir_popu~ir_flex~ir_hppeso~ir_trend,0);
z   = lambda.*(x~zip~zir);

/* Declare a seqüência do PASSO 4 */
aqsi      = ano~qsi;
aqsi05    = delif(aqsi,aqsi[:,1]./=2005);
aqsi06    = delif(aqsi,aqsi[:,1]./=2006);
aqsi07    = delif(aqsi,aqsi[:,1]./=2007);
aqsi08    = delif(aqsi,aqsi[:,1]./=2008);
qsi05     = aqsi05[:,2];
qsi06     = aqsi06[:,2];
qsi07     = aqsi07[:,2];
qsi08     = aqsi08[:,2];
az        = ano~z;
az05      = delif(az,az[:,1]./=2005);
az06      = delif(az,az[:,1]./=2006);
az07      = delif(az,az[:,1]./=2007);
az08      = delif(az,az[:,1]./=2008);
z05       = missrv(az05[:,2:(cols(z)+1)],0);
z06       = missrv(az06[:,2:(cols(z)+1)],0);
z07       = missrv(az07[:,2:(cols(z)+1)],0);
z08       = missrv(az08[:,2:(cols(z)+1)],0);
cm05      = (1/m)*sumc(z05.*qsi05);
cm06      = (1/m)*sumc(z06.*qsi06);
cm07      = (1/m)*sumc(z07.*qsi07);
cm08      = (1/m)*sumc(z08.*qsi08);
w05       = (diagrv(eye(rows(cm05)),(1/m)*sumc((z05.*qsi05-
cm05').^2))).^0.5;
w06       = (diagrv(eye(rows(cm06)),(1/m)*sumc((z06.*qsi06-
cm06').^2))).^0.5;
w07       = (diagrv(eye(rows(cm07)),(1/m)*sumc((z07.*qsi07-
cm07').^2))).^0.5;
w08       = (diagrv(eye(rows(cm08)),(1/m)*sumc((z08.*qsi08-
cm08').^2))).^0.5;
retp      (cm05'inv(w05)*cm05+cm06'inv(w06)*cm06+cm07'inv(w07)*cm07+
cm08'inv(w08)*cm08);
endp;
/* Aqui termina o looping externo e o PASSO 4*/

```

ANEXO III

Essa análise visa comparar os preços recebidos pelo produtor com e sem tributação. A tabela adiante apresenta, por modelo da amostra, para 2008, o preço pago pelo consumidor com tributação, o recebido pelo produtor ($(1-\tau_j) \times p_j$) com tributação e o preço simulado sem tributação ($p_j^{pós}$). O preço médio recebido pelo produtor com tributação seria de R\$ 31,35 mil, o qual aumentaria para R\$ 32,54 mil após a desoneração tributária (uma diferença, na média, de R\$ 1,19 mil).

Por outro lado, como pode ser visto na tabela 12 do texto, a redução média estimada para o preço pago pelo consumidor seria da ordem de R\$ 9,54 mil.

Modelo	Preço – R\$ mil – 2008			(Preço sem tributação) – (Preço recebido pelo produtor com tributação)	Vectra Sedan	66,61	42,64	49,12	6,48
	Preço pago pelo consumidor com tributação	Preço recebido pelo produtor com tributação	Preço sem tributação						
Doblo	56,63	39,64	42,10	2,46	Zafira	71,13	45,52	39,99	-5,53
Idea	50,35	35,24	37,53	2,29	Civic	75,87	53,11	58,00	4,89
Palio > 1.0	30,74	21,52	21,81	0,29	Fit	51,23	35,86	37,03	1,17
Palio 1.0	28,78	21,02	21,84	0,82	206	39,98	27,99	28,16	0,17
Palio Weekend	46,26	32,38	34,49	2,11	307	56,64	39,65	42,13	2,48
Punto	45,39	31,77	32,89	1,12	206 SW	49,65	34,75	37,10	2,35
Sienna > 1.0	42,55	29,79	30,50	0,71	C3	47,48	33,24	34,47	1,23
Sienna 1.0	31,74	23,18	23,99	0,81	C4 Pallas	69,02	44,17	49,70	5,53
Stilo	57,15	40,01	42,64	2,63	Picasso	63,29	44,30	46,35	2,05
Uno Mille	24,02	17,53	17,86	0,33	Clio Sedan > 1.0	45,49	31,84	34,13	2,29
Fiesta Sedan > 1.0	37,37	26,16	26,57	0,41	Clio Sedan 1.0	39,75	29,03	31,04	2,01
Fiesta Sedan 1.0	32,91	24,03	24,10	0,07	Clio 1.0	27,34	19,96	20,01	0,05
Fiesta > 1.0	40,47	28,33	29,11	0,78	Kangoo	47,85	33,50	34,55	1,05
Fiesta 1.0	36,15	26,4	27,16	0,76	Logan > 1.0	36,68	25,68	25,87	0,19
Focus Hatch	47,37	33,16	34,13	0,97	Logan 1.0	29,64	21,64	21,81	0,17
Focus Sedan	52,37	36,66	38,01	1,35	Megane	61,70	43,19	45,03	1,84
Ka > 1.0	31,80	22,26	20,93	-1,33	Sandero > 1.0	38,09	26,66	26,88	0,22
Ka 1.0	25,19	18,39	19,10	0,71	Sandero 1.0	31,69	23,13	23,70	0,57
Astra Hatch	50,92	32,59	33,46	0,87	Scenic	60,11	42,08	40,84	-1,24
Astra Sedan	52,76	33,77	34,64	0,87	Corolla	66,74	46,72	50,41	3,69
Celta 1.0	28,10	20,52	21,08	0,56	Corolla Fielder	72,45	50,71	54,15	3,44
Classic	29,61	21,62	22,31	0,69	Fox > 1.0	37,73	26,41	26,98	0,57
Corsa Hatch > 1.0	37,35	26,14	27,38	1,24	Fox 1.0	33,44	24,42	25,42	1,00
Corsa Hatch 1.0	31,31	22,86	23,76	0,90	Golf	53,34	37,34	39,08	1,74
Corsa Sedan > 1.0	35,70	24,99	25,87	0,88	Gol > 1.0	30,36	21,24	20,98	-0,26
Meriva	51,39	35,97	39,48	3,51	Gol 1.0	29,46	21,50	22,10	0,60
Prisma	33,22	23,26	23,25	-0,01	Parati	42,67	29,87	30,86	0,99
Vectra Hatch	65,09	41,66	40,11	-1,55	Polo Hatch	45,44	31,81	33,25	1,44
					Polo Sedan	51,49	36,04	38,12	2,08
					Spacefox	49,44	34,61	36,48	1,87

ANEXO IV

Tome por base a equação (4) do ensaio 2:

$$\Pr_j = \int \frac{\exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,j} + \xi_j\right) + \left(\sigma_c \cdot z_{c,\tau} + \sum_{k=1}^K \sigma_k \cdot d_{k,j} \cdot z_{k,\tau}\right)\right]}{1 + \sum_{r=0}^{2^K} \exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,r} + \xi_r\right) + \left(\sigma_c \cdot z_{c,\tau} + \sum_{k=1}^K \sigma_k \cdot d_{k,r} \cdot z_{k,\tau}\right)\right]} \cdot dF(z)$$

Um estimador de (4) para Q postos é:

$$\Pr_j = \frac{1}{Q} \cdot \sum_{q=1}^Q \frac{\exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,j} + \xi_j\right) + \left(\sigma_c \cdot z_{c,q} + \sum_{k=1}^K \sigma_k \cdot d_{k,j} \cdot z_{k,q}\right)\right]}{1 + \sum_{r=0}^{2^K} \exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,r} + \xi_r\right) + \left(\sigma_c \cdot z_{c,q} + \sum_{k=1}^K \sigma_k \cdot d_{k,r} \cdot z_{k,q}\right)\right]}$$

Usando (3):

$$\bar{m}_j = \frac{1}{Q \cdot (1 + 1/\lambda)} \cdot \sum_{q=1}^Q \frac{\exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,j} + \xi_j\right) + \left(\sigma_c \cdot z_{c,q} + \sum_{k=1}^K \sigma_k \cdot d_{k,j} \cdot z_{k,q}\right)\right]}{1 + \sum_{r=0}^{2^K} \exp\left[\left(\bar{c} + \sum_{k=1}^K \bar{\beta}_k \cdot d_{k,r} + \xi_r\right) + \left(\sigma_c \cdot z_{c,q} + \sum_{k=1}^K \sigma_k \cdot d_{k,r} \cdot z_{k,q}\right)\right]}$$

Usando a agregação $J = \sum_k^{K-1} 2^{k-1} \cdot d_k$, note que até a *dummy* $K-1$ a seqüência de posições se mantém. Então:

$$\bar{m}_J = \frac{\sum_{q=1}^Q \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,J} + \hat{\sigma}_c \cdot \tilde{z}_{c,q} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,J} \cdot \tilde{z}_{k,q}\right] \cdot \frac{(\exp(\hat{\beta}_K + \hat{\sigma}_K \cdot \tilde{z}_{K,q} + \hat{\xi}_{j'}) + \exp(\hat{\xi}_{j''}))}{Q \cdot (1 + 1/\lambda)}}{1 + \sum_{R=0}^{2^{K-1}} \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,R} + \hat{\sigma}_c \cdot \tilde{z}_{c,q} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,R} \cdot \tilde{z}_{k,q}\right] \cdot (\exp(\hat{\beta}_K + \hat{\sigma}_K \cdot \tilde{z}_{K,q} + \hat{\xi}_{r'}) + \exp(\hat{\xi}_{r''}))}$$

Agora note que se o tipo agregado J não tiver um sub-tipo j' ocorre:

$$\bar{m}_J = \frac{\sum_{q=1}^Q \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,J} + \hat{\sigma}_c \cdot \tilde{z}_{c,q} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,J} \cdot \tilde{z}_{k,q}\right] \cdot \frac{\exp(\hat{\xi}_{j''})}{Q \cdot (1 + 1/\lambda)}}{1 + \sum_{R=0}^{2^{K-1}} \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,R} + \hat{\sigma}_c \cdot \tilde{z}_{c,q} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,R} \cdot \tilde{z}_{k,q}\right] \cdot \exp(\hat{\xi}_{r''})}$$

E se o tipo agregado J não tiver um sub-tipo j'' ocorre:

$$\bar{m}_J = \frac{\sum_{q=1}^Q \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,J} + \hat{\sigma}_c \cdot \tilde{z}_{c,q} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,J} \cdot \tilde{z}_{k,q}\right] \cdot \frac{\exp(\hat{\beta}_K + \hat{\sigma}_K \cdot \tilde{z}_{K,q} + \hat{\xi}_{j'})}{Q \cdot (1 + 1/\lambda)}}{1 + \sum_{R=0}^{Q^{K-1}} \exp\left[\hat{c} + \sum_{k=1}^{K-1} \hat{\beta}_k \cdot d_{k,R} + \hat{\sigma}_c \cdot \tilde{z}_{c,q} + \sum_{k=1}^{K-1} \hat{\sigma}_k \cdot d_{k,R} \cdot \tilde{z}_{k,q}\right] \cdot \exp(\hat{\beta}_K + \hat{\sigma}_K \cdot \tilde{z}_{K,q} + \hat{\xi}_{r'})}$$

Então basta posicionar adequadamente γ' e γ'' (e substituir Q por V).

ANEXO V

Abaixo se listam os $\beta_{0,\ell}$ estimados (referentes ao terceiro ensaio). Lembrando que eles funcionam como uma referência de utilidade comum entre os moradores de uma dada localidade, e, pela estrutura do *Mixed* LOGIT (sendo de praxe se normalizar a utilidade pela opção externa em zero), quanto menor o $\beta_{0,\ell}$ de uma localidade maior será a probabilidade de um eleitor daquela localidade optar por votar branco ou nulo.

SP	-1,562	AC	-1,112
MG	-1,548	RJ	-1,111
GO	-1,530	RO	-1,089
PI	-1,416	MT	-1,065
BA	-1,394	SC	-1,038
TO	-1,393	RS	-1,021
PB	-1,391	PA	-0,996
SE	-1,301	ES	-0,935
MA	-1,289	PR	-0,925
CE	-1,273	AM	-0,797
PE	-1,191	AP	-0,581
AL	-1,131	RR	-0,573
RN	-1,126	DF	-0,447
MS	-1,114	Exterior	-0,029