



Universidade de Brasília – UnB  
Faculdade de Economia, Administração, Contabilidade e Gestão Pública – FACE  
Departamento de Economia – ECO

MATHEUS JOSÉ SILVA DE SOUZA

**Uma Introdução Compreensiva  
à Modelagem Estrutural na Economia**

Brasília

2017

MATHEUS JOSÉ SILVA DE SOUZA

**Uma Introdução Compreensiva  
à Modelagem Estrutural na Economia**

Monografia apresentada ao Departamento de Economia – ECO – como requisito parcial à obtenção do título Bacharel em Economia.

Orientador: Prof. Dr. Daniel Oliveira Cajueiro.

Brasília  
2017



# Agradecimentos

Agradeço, mormente, à Deus, sem o qual nem eu estaria aqui nem aos demais aos quais estou agradecendo. Agradeço a meu pai e minha mãe por *tudo*. Não vou descrever o *tudo* pois levaria algumas muitas páginas, mas eles sabem *tudo* que fizeram para viabilizar este investimento arriscado. Ademais, agradeço a meu irmão, por ter sido um bom irmão (às vezes).

Agradeço, é claro, ao Prof. Daniel Cajueiro pela ótima orientação deste trabalho.

Agradeço ao Prof. Vinicius do Departamento de Administração da Universidade de Brasília, que apesar de velho, não está ultrapassado e está constantemente reagindo. Agradeço, finalmente, a todos os outros que não me lembrei diretamente enquanto escrevia este agradecimento mas que certamente fizeram parte desta história espetacular.



# Resumo

A modelagem estrutural na Economia tem sido extensivamente difundida, uma vez que trata problemas econométricos a partir de hipóteses econômicas, conferindo um maior grau de relevância e melhor capacidade de previsão aos modelos. Os benefícios desse tipo de abordagem são diversos: maior proximidade da teoria com a realidade dos dados; possibilidade de o pesquisador capturar a influência das preferências dos agentes sobre o comportamento esperado; o fato de os parâmetros estimados serem invariantes para uma mesma estrutura; e, conseqüentemente, a possibilidade de simular políticas contrafactuais. No entanto, todos esses benefícios são acompanhados de alguns custos e dificuldades de modelagem do problema ou implementação computacional do modelo para estimação dos parâmetros. Por essa razão, este trabalho foi elaborado na forma de um tutorial compreensivo que auxilia na formulação de modelos estruturais. Para tanto, foi elaborada uma revisão de literatura que introduz seus principais fundamentos e destaca sua relevância e foi formulado e implementado um modelo simples de escolha dinâmica e discreta, além de providos alguns recursos para análise dos resultados.

***Palavras-chave:*** Econometria Estrutural, Escolha Dinâmica e Discreta, Máxima Verossimilhança, Python.



# Abstract

The Structural Economics got spread around the academic environment due to a new approach on econometric research grounded on economic hypothesis, which grants more relevance and better forecast power to the models. The payoffs of Structural Modeling are many: the theory gets more closer to the data observed; makes possible to capture the impact of agents preferences on the expected behaviour; the estimation of invariant parameters for a certain structure; and, consequently, the power to conduct counterfactual policy analysis. However, all these benefits came with some costs, like a harder modeling or a more complex computational implementation of the model to estimate the parameters. Therefore, this work was made like an comprehensive tutorial which aids the making of structural work on Economics. Hence, it comprises a literature review which discuss the structural econometrics' fundamentals and importance and formulates and implements a dynamic discrete choice model, as well provide some tools to analyze the results.

**Keywords:** Structural Econometrics, Dynamic Discrete Choice, Maximum Likelihood, Python.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Revisão de Literatura</b>	<b>7</b>
2.1	Importância da modelagem estrutural de problemas econômicos . . . . .	7
2.2	Aspectos técnicos da formulação de modelos na econometria estrutural . . . . .	13
2.3	Identificação de modelos estruturais e análise de políticas contrafactuais . . . . .	16
<b>3</b>	<b>Estimação de Modelos Estruturais</b>	<b>19</b>
3.1	Formulação do modelo . . . . .	19
3.2	Base de dados . . . . .	27
3.3	Implementação do algoritmo . . . . .	28
<b>4</b>	<b>Resultados do algoritmo e análise dos resultados</b>	<b>35</b>
<b>5</b>	<b>Conclusão</b>	<b>41</b>
	<b>Apêndice A Demonstração do Teorema do ponto fixo de Banach</b>	<b>43</b>
	<b>Apêndice B Derivação da função logit dinâmica</b>	<b>45</b>



# Capítulo 1

## Introdução

A Teoria Econômica desenvolvida desde o surgimento da Economia Política no século XV até a contemporaneidade é, como toda ciência, fundamentada em hipóteses acerca do comportamento do objeto de estudo que, no caso da economia, é essencialmente o homem, tomado como um ser racional. Diferentes axiomas e métodos conduzem a diferentes conclusões acerca daquilo que é analisado, o que contribui para o desenvolvimento de diversas teorias dentro de uma mesma área de pesquisa. Desenvolvida uma teoria, geralmente, busca-se o respaldo dela na realidade observável, o que destaca as teorias que mais se adequam à realidade e aquelas com menor grau de ajuste a ela.

Na etapa em que a realidade é requerida para validar ou não uma teoria econômica, a econometria é o campo que fornece o ferramental para conhecer o seu grau de adequação aos fatos observados. Originalmente definida como sendo a síntese de conhecimentos econômicos aos métodos estatísticos, a econometria tem se tornado, no entanto, o estudo estatístico de variáveis econômicas, tratando a maioria dos fenômenos como ocorrências de curto prazo e descartando as constantes mudanças por que passam as economias, o que afeta as relações entre as variáveis. De maneira geral, é como se os pressupostos de uma teoria fossem omitidos no modelo estatístico.

Para exemplificar como usualmente são tratados problemas econométricos, considere o caso de seleção adversa no mercado de seguros. A teoria prevê que, dado que o agente tem pleno conhecimento do seu perfil de risco e o principal não tem essa informação, maior será a cobertura do contrato adquirida por um agente com alto risco de utilizar este seguro. Para testar essa conclusão, Cohen e Siegelman (2009) sugerem uma regressão linear em que o risco é a variável dependente e a cobertura do seguro é uma das variáveis explicativas.

Esse tipo de análise investiga se existe uma correlação entre ambas as variáveis e se a modelagem linear é satisfatória para aproximar o efeito que a cobertura de seguro escolhida por um indivíduo tem sobre o nível de risco de um indivíduo. Essa forma de abordagem é, atualmente, referida na literatura como forma reduzida ou modelagem semi-experimental. Neste aspecto, diferencia-se da abordagem puramente experimental em virtude da natureza dos

dados econômicos, que dificulta a tarefa de isolar o efeito de uma variável sobre outra, isto é, impossibilidade de se conduzir experimentos aleatórios.

Se, para uma determinada amostra, esse modelo fornece estimativas consistentes para os parâmetros e os testes estatísticos conferem robustez a ele, todas as hipóteses da teoria de seleção adversa, completamente implícitas ao modelo econométrico, são validadas. Caso contrário, o modelo é rejeitado para aquela amostra, o que leva a crer que aquelas hipóteses não são adequadas para uma determinada conjuntura.

Note que um estudo desse tipo é não mais que a avaliação de curto prazo das relações estabelecidas entre variáveis para uma determinada amostra no tempo e, quando executado individualmente, parece distanciar o conteúdo econômico da análise. Neste aspecto, salienta-se que uma forma mais fiel pela qual se pode testar a adequação de uma teoria aos dados da realidade é através da incorporação das hipóteses dessa teoria na especificação do modelo estatístico, isto é, adição de uma estrutura ao modelo econométrico. Essa é a premissa da modelagem estrutural na economia.

Neste âmbito, um problema de seleção adversa semelhante é modelado por Jeziorski, Krasnokutskaya e Ceccarini (2017), que fazem uma análise estrutural do mercado de seguros. Primordialmente, essa nova abordagem é caracterizada pelas hipóteses estruturais de que parte o modelo, isto é, especificações *ex ante* do comportamento esperado dos agentes com base na Teoria Econômica. Os autores consideram que os agentes escolhem, em cada período, uma cobertura básica ou estendida de seguro e que podem modificar seu perfil de risco no início de cada período, desde que incorram em um determinado custo para manter esse perfil. Argumentam que essas hipóteses permitem ao modelo incorporar uma elevação no preço do seguro em virtude do número recente de acidentes sofridos e a possibilidade de o agente aprender com a sua experiência.

Note, portanto, que a primeira diferença fundamental entre a modelagem estrutural e a modelagem semi-experimental da economia é essa: ao passo que a primeira torna explícitas as hipóteses do modelo econômico no modelo econométrico, definindo funções de utilidade e relações de comportamento esperado para o agente no intuito de verificar se as conclusões do modelo se sustentam, a última parte, sobretudo, das conclusões de uma teoria para verificar se suas hipóteses se sustentam para uma determinada conjuntura. A modelagem estrutural é, portanto, o resgate da economia na econometria. Assim, métodos estatísticos passam a servir à Teoria Econômica e não o contrário.

Dada a notoriedade da modelagem estrutural para o desenvolvimento da economia, destaca-se a extrema relevância deste trabalho, elaborado na forma de um tutorial compreensivo cuja pretensão é não somente salientar os ganhos que a especificação econômica dos modelos estatísticos traz, como também introduzir técnicas básicas de modelagem estrutural e estimação dos parâmetros desses modelos.

No Capítulo 2, buscou-se fazer uma revisão acerca da literatura de modelos estruturais, destacando as origens, os argumentos em prol da incorporação e explicitação de hipóteses econômicas a um modelo, os custos associados às técnicas de modelagem estrutural e os benefícios advindos delas. Para discutir a origem da modelagem estrutural, são tratados os modelos de equações simultâneas, que permitem analisar uma economia na presença de endogenia entre algumas variáveis.

No Capítulo 3 é detalhado o modelo de Rust (1987), artigo seminal que desenvolve uma classe de modelo estrutural muito relevante, baseada em um modelo dinâmico de escolha discreta (DDC<sup>1</sup>). Neste trabalho foram analisadas as implicações formais da incorporação de hipóteses econômicas na especificação de um modelo econométrico e como tratá-las, isto é, buscou-se desenvolver o modelo de forma compreensiva, bem como implementá-lo.

Optou-se pelo embasamento em Rust (1987) por esse ser um dos primeiros estudos que desenvolve um trabalho de estimação estrutural e formulação de um DDC em tempo infinito, muito recorrente neste tipo de problema. O intuito é, por conseguinte, verificar se as decisões individuais de um agente coincidem com o ótimo esperado pela teoria econômica, isto é, se ele é ou não maximizador de utilidade. No entanto, para implementação do modelo, foi considerada uma linguagem de programação mais moderna que aquela utilizada por Rust (1987). Ademais, o algoritmo conta com algumas modificações no intuito de diminuir sua complexidade e aprimorar sua eficiência.

Na Seção 3.3 é implementado e estimado o modelo desenvolvido em Python. Finalmente, no Capítulo 4 são apresentados os resultados e exploradas mais algumas possibilidades desse tipo de modelagem, como a simulação de políticas contrafactuais.

---

<sup>1</sup>Do inglês *Dynamic Discrete Choice*.



# Capítulo 2

## Revisão de Literatura

### 2.1 Importância da modelagem estrutural de problemas econômicos

Modelos estruturais partem da Teoria Econômica para derivar relações de interesse entre as variáveis de uma base de dados, incorporando supostos sobre o comportamento do agente na especificação do modelo. Isto significa que a partir de hipóteses acerca da racionalidade dos agentes, como por exemplo a maximização de utilidade, formulam-se modelos econométricos com aspectos econômicos inerentes, o que não somente confere à análise dos dados maior proximidade ao mundo observável como também torna a pesquisa mais relevante.

Suponha que um pesquisador tem a sua disposição uma base de dados de séries de tempo para o consumo em uma região. Considere que  $c_t$ ,  $y_t$  e  $p_t$  denotam, respectivamente, o consumo, a renda e o preço de um bem no período  $t$ . Um modelo comumente formulado é:

$$c_t = \alpha_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 p_t + \varepsilon_t \quad (2.1)$$

em que  $\varepsilon_t$  denota um ruído branco. Embora não seja puramente atóxico, uma vez que existe uma intuição econômica por trás da Eq. (2.1), isto é, o consumo de um agente ou de uma economia é colocado como função das rendas auferidas no período corrente, no período anterior e dos preços correntes, essa modelagem é, sobretudo, descritiva e assume uma relação linear entre as variáveis.

O caráter linear não representa um grande problema, já que em modelos de regressão múltipla podem ser incorporadas “não-linearidades”. No entanto, essa modelagem capta o comportamento do consumo com mudanças exógenas, representadas pelas variáveis no lado direito na Eq. (2.1). Assim, não é razoável supor, por exemplo, que os preços configurem uma variável exógena do modelo, exceto no curto prazo, quando os preços apresentam maior nível de resistência e menor flexibilidade.

Modelos como este são, portanto, úteis em análises de curto prazo essencialmente, já que não existe um caráter dinâmico na relação entre as variáveis, isto é, *ceteris paribus*, um aumento de uma unidade na renda  $y$  sempre levará a uma variação de  $\beta_1$  no consumo, para  $t = 1, 2, \dots$ . O que se vê extensivamente na literatura, no entanto, é a generalização para o longo prazo de modelagens de curto prazo, que negligenciam a mudança na influência de uma variável sobre outra, isto é, as mudanças estruturais na economia. Para se obter parâmetros que sejam relativos ao longo prazo o modelo precisa incorporar hipóteses que traduzam aspectos estruturais da economia. Frisch (1933) sumariza a deficiência da modelagem semi-experimental no seguinte trecho:

*“No amount of statistical information, however complete and exact, can by itself explain economic phenomena. If we are not to get lost in the overwhelming, bewildering mass of statistical data, we need the guidance of a powerful theoretical framework. Without this no significant interpretation of our observations will be possible.”*

Neste sentido, se uma das hipóteses do modelo econométrico é que, no tempo  $t_0$ , o indivíduo maximiza o seu consumo no período corrente e nos períodos seguintes quando descontados por um fator qualquer, que é um comportamento perene do agente racional, nota-se que o modelo tem a pretensão de capturar aspectos estruturais da economia. Essa hipótese, em suma, é fundamentada na concepção de Friedman (1957), o qual afirma que os agentes tendem a suavizar o consumo, uma vez que o consumo volátil geraria utilidade negativa.

Assim, a modelagem estrutural surge como uma alternativa mais robusta à econometria, permitindo que os modelos captem mudanças estruturais na economia em análise, bem como recuperar parâmetros estruturais de longo prazo, tornando esse tipo de modelo mais adequado para previsão e simulação de situações não verificadas anteriormente.

Entenda mudança estrutural como uma alteração na estrutura de uma economia, que afeta as decisões agregadas e, por conseguinte, as decisões individuais dos agentes no longo prazo. Uma mudança estrutural significativa, por exemplo, ocorreu no período pós Segunda Grande Guerra em alguns países asiáticos, como Hong Kong, Singapura, Taiwan e Coreia do Sul, que ficaram conhecidos como Tigres Asiáticos.

Em poucas linhas, verificou-se a alocação de mão de obra de setores de baixa produtividade para setores de mais alta produtividade, o que conduziu a economia a maiores taxas de crescimento do produto e uma elevação da capacidade industrial. Nestas condições, economias com baixa produtividade tornaram-se industrializadas e, portanto, houve uma mudança estrutural nestes países, modificando a forma com que a renda dos agentes ou os preços dos bens relacionavam-se com o consumo, por exemplo. Este estudo da produtividade e crescimento dos Tigres Asiáticos foi feito por Young (1994).

Quando se fala em estrutura de uma economia, busca-se captar relações de longo prazo e invariantes no tempo entre as variáveis, dadas as hipóteses feitas sobre o comportamento esperado dos agentes. Nevo e Whinston (2010) argumentam que a econometria estrutural tem o intuito primário de, com base nas observações de um processo estocástico, recuperar parâmetros estruturais da economia, que são invariantes, como tecnologia de firmas e preferências dos consumidores e permitem avaliar a eficiência de políticas de decisão. Seguente a essa etapa, sob a hipótese de que os parâmetros estimados são perenes, o modelo, sublinhado pela teoria econômica, é usado para prever as mudanças na economia a partir da aplicação de diferentes políticas, tenham elas ocorrido anteriormente ou não.

Segundo Reiss e Wolak (2007), o economista estrutural deve argumentar de forma convincente que o modelo é invariante para mudanças na economia, isto é, capta eventuais mudanças estruturais na economia. Uma das formas pela qual pode-se argumentar é que, uma vez estimados os parâmetros comportamentais de um agente racional maximizador de utilidade de forma consistente, garantimos que as mudanças estruturais estão sendo capturadas e que esses parâmetros são relativos ao longo prazo também, pois, seja lá qual for a estrutura da economia, o agente estará buscando maximizar seus ganhos.

Segundo Hausman (1983, p. 396) a principal diferença entre a modelagem estrutural e a modelagem reduzida é a presença de mais de uma variável endógena. Existem muitas outras divergências que separam ambas abordagens, mas, de fato, essa diferença destacada pelo autor é crucial e foi ela quem fomentou a formulação de modelos de equações simultâneas (SEM<sup>1</sup>), muito empregado em estudos sobre oferta e demanda, em que são considerados endógenos os preços e as quantidades de um bem. Discutir alguns aspectos desse tipo de modelo auxilia na compreensão de alguns aspectos gerais da modelagem estrutural.

Suponha que  $q_t$  denota a quantidade de um bem no mercado no instante  $t$  e que  $p_t$  seja o preço correspondente. Ademais, denote por  $w_t$  a renda de um consumidor, que é uma variável exógena. Assuma que essas variáveis são relacionadas conforme:

$$q_t = \phi_0 + \phi_1 p_t + \phi_2 w_t + \varepsilon_t \quad (2.2)$$

$$q_t = \psi_0 + \psi_1 p_t + u_t \quad (2.3)$$

em que a Eq. (2.2) é uma função de demanda e a Eq. (2.3) é proveniente da oferta de mercado. Ambas equações são ditas estruturais. Notadamente esse problema tem uma natureza endógena, que inviabiliza o uso do método de Mínimos Quadrados Ordinários (MQO), já que não se verifica a hipótese fundamental de exogeneidade dos regressores. Como MQO é um método relativamente simples de ser implementado, já fica evidente que a modelagem estrutural, desde

---

<sup>1</sup>Do inglês: *Simultaneous Equations Models*

sua origem com equações simultâneas, apresenta maior grau de complexidade em termos de implementação e estimação do modelo.

Note que, em um SEM, pode-se escrever as variáveis endógenas em termos das variáveis exógenas. No caso acima, usamos a condição de equilíbrio de mercado, que iguala as quantidades demandada e ofertada. Nessas condições:

$$p_t = \pi_0 + \pi_1 w_t + \nu_t \quad (2.4)$$

em que  $\pi_0 = \frac{\phi_0 - \psi_0}{\psi_1 - \phi_1}$ ,  $\pi_1 = \frac{\phi_2}{\psi_1 - \phi_1}$  e  $\nu_t = \varepsilon_t - u_t$ , que é um termo de erro composto. Substituindo a Eq. (2.4) na Eq. (2.3) obtemos:

$$q_t = \pi_2 + \pi_3 w_t + z_t \quad (2.5)$$

em que os coeficientes  $\pi_2 = \frac{\psi_1 \phi_0 - \psi_0 \phi_1}{\psi_1 - \phi_1}$ ,  $\pi_3 = \frac{\psi_1 \phi_2}{\psi_1 - \phi_1}$  e o termo de erro  $z_t = \frac{\psi_1 \varepsilon_t - \phi_1 u_t}{\psi_1 - \phi_1}$ . As Eq. (2.4) e Eq. (2.5) são chamadas de equações reduzidas, que expressam a relação das variáveis endógenas do modelo com as variáveis exógenas, dada a relação de endogenia entre as variáveis que o modelo busca explicar. Por essa razão, é dito que um modelo na forma reduzida pode ser obtido de um modelo estrutural mais geral e que a estimação de parâmetros de uma equação na forma reduzida é um processo intermediário à estimação estrutural.

Os coeficientes  $\pi_0$ ,  $\pi_1$ ,  $\pi_2$  e  $\pi_3$  são chamados de multiplicadores de impacto porque captam variações de curto prazo sobre as variáveis explicadas pelo modelo. Nessas condições, fica evidente que um modelo que estima uma regressão conforme a Eq. (2.1) não recupera os parâmetros estruturais do modelo a partir das estimativas obtidas para os multiplicadores de impacto e é, por conseguinte, um modelo de curto prazo somente.

Outro tipo de modelagem estrutural muito comum e que será o foco desse trabalho são os modelos que trabalham com hipóteses comportamentais dos agentes na especificação do modelo econométrico, como por exemplo a consideração de que os agentes atuam no sentido de maximizar uma função objetivo. Esse tipo de formulação requer, muitas vezes, métodos não lineares de estimação e a caracterização das variáveis de escolha e dos estados possíveis do problema. Geralmente, o problema de maximização intertemporal do consumidor é descrito como sendo:

$$\begin{aligned} \max_{y_t, p_t} \quad & E \left[ \sum_{t=0}^{\infty} \phi^t c_t(y_t, p_t) \right] \\ \text{s.a} \quad & y_0 = w \\ & y_t = c_t + s_t \end{aligned} \quad (2.6)$$

em que considera-se o indivíduo como ser racional e maximizador de utilidade. É comum que seja incluída alguma restrição sobre o consumo em relação à riqueza inicial  $w$  ou que haja um

termo explícito da poupança  $s_t$  em cada período. Na Eq. (2.6),  $\phi \in [0, 1]$  é um fator de desconto que capta o quão o agente valoriza sua utilidade em períodos de tempo futuros em relação ao momento atual.

Nessas condições, torna-se especialmente relevante, discutir a interpretação dos parâmetros estimados tanto por modelos estruturais quanto por modelos semi-experimentais. No caso em que as variáveis em Eq. (2.1) são logaritmizadas, por exemplo, os coeficientes da regressão linear adquirem caráter de elasticidade, isto significa que, por exemplo,  $\beta_1$  seria interpretado como a elasticidade renda da demanda do bem. Nesse sentido, torna-se menos plausível a utilização de um modelo semi-experimental para fazer previsões de longo prazo, já que estaríamos, implicitamente, supondo que  $\beta_1$  é constante ao longo do tempo, sejam lá quais mudanças estruturais ocorrerão na economia.

Muitos estudos foram feitos sobre o comportamento da elasticidade renda dos bens. Um dos principais e mais difundidos foi elaborado por Ernst Engel em 1857, que ficou conhecido como curva de Engel. Engel argumentou que à medida que a renda de uma família crescia no tempo, o consumo de um determinado bem crescia cada vez mais até certo ponto, a partir do qual, cresceria cada vez menos com o acréscimo da renda, até atingir um ponto de saturação. Em outras palavras, a elasticidade renda da demanda não seria constante à proporção em que ocorrem mudanças estruturais na economia. Essa é uma motivação adicional para o desenvolvimento de modelos estruturais.

No caso em que são obtidas elasticidades de um modelo estrutural, o econometrista pode argumentar que determinadas estimativas desses parâmetros de elasticidade são compatíveis com a estrutura em que foi formulado o modelo, isto é, o arcabouço das hipóteses comportamentais feitas sobre o agente. Segundo Reiss e Wolak (2007, p. 4289), essa é uma vantagem da modelagem estrutural, tendo em vista que podem ser obtidas elasticidades de acordo com diversas especificações da função de utilidade, por exemplo. Muitas vezes, quando se fala em modelos de escolha intertemporal, a elasticidade comumente analisada é a elasticidade de substituição intertemporal, que mede a variação no consumo corrente dada uma variação da taxa de juros real da economia.

A crítica de Lucas (1976), como ficou conhecida a tese de Robert Lucas, argumenta que não se pode tentar prever o comportamento de um sistema conhecendo-se apenas a história – os eventos passados – desse sistema. Esta análise não somente trouxe implicações importantes para a macroeconomia, mormente, em relação ao arcabouço de expectativas normativas, como também, e talvez mais fortemente, para a econometria.

Note que, se um pesquisador tem uma base de dados e deseja estimar (2.1), ele acredita que, se satisfeitas algumas condições sobre os estimadores obtidos para (2.1), com base em observações anteriores acerca do consumo, renda e preços, será possível estimar consistentemente quanto o agente consumirá em  $t + k$ ,  $k = 1, 2, 3, \dots$

Nesse sentido, Lucas (1976) chama a atenção para as deficiências de um modelo semi-experimental descritivo, isto é, os estimadores obtidos nesse caso pressupõe que os agentes se comportam exatamente como comportaram nos períodos anteriores (com os mesmos hábitos de consumo e renda). Isto implica que não há mudanças estruturais na economia, ou seja, a influência de uma variável explicativa sobre a variável dependente não muda no tempo.

A previsão a partir das estimativas de um modelo semiexperimental é apropriada para  $k$  pequeno, ou seja, no curto prazo, quando é razoável supor que não há mudanças estruturais significativas na economia. Recordando da literatura acerca dos SEM, os parâmetros das equações reduzidas são chamados de multiplicadores de impacto, que captam efeitos de curto prazo das variáveis exógenas sobre as variáveis endógenas.

A modelagem estrutural, à luz da crítica de Lucas, é, portanto, a forma pela qual a econometria atinge a excelência, ou seja, une a teoria econômica aos conceitos estatísticos. Isso se deve ao fato de que modelos estruturais são invariantes para um sistema. Uma vez invariantes, alguns autores, como Su e Judd (2012), afirmam que a principal vantagem desses modelos é a capacidade que o pesquisador adquire de conduzir análise de políticas contrafactuais. Isso significa que se pode prever de forma mais adequada o caminho a ser seguido por uma economia caso alguns de seus elementos sejam alterados.

Notadamente, percebe-se que a Teoria Econômica deve ser essencialmente empregada na formulação de modelos que têm a pretensão de serem relevantes. Na literatura acerca da estimação por variáveis instrumentais, por exemplo, um bom instrumento é aquele que é altamente correlacionado com uma variável explicativa e não correlacionado com o termo de erro. Pode-se testar facilmente a correlação de um instrumento com uma variável explicativa através de uma regressão simples. No entanto, a ausência de correlação do instrumento com o termo de erro não é trivialmente garantida e a Teoria Econômica provém hipóteses pertinentes para isso.

Para Keane (2010), a razão que justifica o advento de modelos ateóricos é a quantidade de trabalho envolvida em trabalhos de econometria estrutural, salientando que é tentador produzir artigos mais rapidamente utilizando modelos que sejam destituídos da Teoria Econômica. De fato, modelos estruturais mais sofisticados utilizam-se de um ferramental mais complexo, isto é, elaboração de um modelagem de escolha dinâmica no tempo à luz da teoria da escolha individual, em que o agente é maximizador de utilidade.

Nevo e Whinston (2010) ainda chamam atenção para os ganhos de uma modelagem estrutural na análise do bem estar. Segundo os autores, muitos estudos são realizados para identificar o efeito dos movimentos de preços na economia, como por exemplo, variações sobre quantidade e inflação. No entanto, somente uma modelagem estrutural sobre a luz da Teoria Econômica permite avaliar o bem estar dos consumidores após uma flutuação de preços ocasionada por uma simulação de uma política econômica, a partir de um modelo estrutural.

Os econométristas que advogam em favor das formas reduzidas salientam que o grande número de hipóteses que são feitas acerca dos agentes prejudica a qualidade e a validade de um modelo estrutural. Nesse sentido, Keane (2010) e Wolpin (2013) salientam uma das diferenças entre a modelagem semiexperimental e a modelagem estrutural: não é a quantidade de hipóteses feitas que as diferencia, mas sim a quantidade delas que é explicitada pelo modelo.

Os autores destacam que em modelos semi-experimentais muitas hipóteses subjacentes são tornadas implícitas. Keane (2010) ainda destaca que o econométrista semi-experimental não está preocupado em como as variáveis do modelo foram geradas ou escolhidas pelo agente. A exogeneidade de um instrumento, por exemplo, não pode ser formalmente provada, de tal forma que mesmo os bons instrumentos consolidados na literatura são exógenos em razão de argumentações a partir da Teoria Econômica que não é, porém, explicitada.

No que concerne às principais aplicações, a modelagem estrutural tem sido bastante empregada recentemente para análise de dados relacionados à economia do trabalho, organização industrial, *marketing* (Arcidiacono e Miller (2013)). Alguns exemplos são: Keane e Wolpin (1997) utilizam um modelo estrutural para estudar decisões de investimento em capital humano; Rust e Phelan (1997) analisam as decisões tomadas durante a idade ativa que afetam os benefícios com aposentadoria e Foss e Host (2011) estimam um modelo estrutural para auxiliar na implementação ótima de uma usina hidrelétrica.

## 2.2 Aspectos técnicos da formulação de modelos na econometria estrutural

A econometria estrutural utiliza extensivamente processos de decisão de Markov na formulação de problemas de decisão intertemporal, uma vez que são muito úteis para capturar o comportamento do agente ao longo do tempo e auxiliam na obtenção de uma solução na forma de uma regra ótima de decisão. Segundo Rust (1994), cadeias de Markov não somente provêm uma teoria de como os agentes deveriam se comportar como também pode proporcionar bons modelos empíricos de como os tomadores de decisão se comportam no mundo real.

Um processo de decisão de Markov é um processo estocástico em que a probabilidade de transição de um estado para o outro não depende da história inteira do processo (conjunto de observações em todos os estados anteriores), mas depende unicamente do estado atual. Mais formalmente, seja  $S$  um conjunto de estados possíveis e  $A_t$  um conjunto de variáveis de controle – são ditas variáveis de controle aquelas que o agente pode escolher –, com  $x_t \in S$  sendo o estado e  $a_t \in A_t$  sendo a escolha do agente, ambos indexados pelo no tempo  $t = 1, 2, \dots, T$ . Se a probabilidade de transição do estado  $t$  para  $t + 1$  é do tipo  $p(x_{t+1}|a_1, x_1, \dots, a_t, x_t)$  dizemos que um processo é markoviano se:

$$p(x_{t+1}|a_1, x_1, \dots, a_t, x_t) = p(x_{t+1}|a_t, x_t) \quad (2.7)$$

Essa propriedade de um processo markoviano é conveniente não somente por ser intuitiva e bastante razoável como também por simplificar a implementação computacional dos procedimentos de estimação.

No que diz respeito aos objetivos das modelagens reduzida e estrutural, Rust (1994) salienta que dado, um processo estocástico discreto de tempo  $T$ , que pode ser expresso como  $\{a_t, x_t\}_{t=1}^T$ , um modelo na forma reduzida está interessado em conhecer o processo estocástico que mais se aproxima das realizações, disponível numa base de dados, ao passo que uma estimação estrutural não busca somente descobrir a forma do processo, mas também estimar as primitivas que geraram as variáveis endógenas, como as probabilidades de transição, a função *payoff* e os fatores de desconto. Neste sentido, a modelagem estrutural permite estimar parâmetros comportamentais e não somente multiplicadores de impacto de uma variável exógena sobre uma variável dependente.

Embora a abordagem estrutural apresente muitos benefícios, quando comparada à econometria semi-experimental, isso não significa que não há um custo associado. Como sugere Keane (2010), essa atividade não é simples e requer maior aprofundamento teórico na área de pesquisa. Além dos aspectos referentes a formulação do modelo, para estimação de um modelo conforme na Eq. (2.6), entre outros procedimentos que serão explicitados mais adiante, existe um enorme custo computacional que dificulta o uso de métodos estatísticos mais eficientes e poderosos, como salientam Su e Judd (2012).

O algoritmo NFXP (*nested fixed point algorithm*), proposto por Rust, estima os parâmetros do modelo a partir da função de verossimilhança dos dados, a qual depende de um ponto fixo de uma contração, isto é, para cada palpite dos parâmetros estruturais o algoritmo faz o cômputo do ponto fixo e, na mesma iteração calcula a função de máxima verossimilhança. O procedimento se repete até que a função de máxima verossimilhança seja maximizada. É, portanto, natural deduzir que o algoritmo NFXP é computacionalmente custoso, já que existem dois *loops* extensos e um deles é interno ao outro.

Algumas alternativas foram propostas para obter um algoritmo mais eficiente que o NFXP. Segundo Judd e Su (2006), para diminuir a complexidade do algoritmo, o problema poderia ser formulado de forma que o cômputo do ponto fixo da contração fosse incorporado ao conjunto de restrições. Nessas condições, para cada palpite, o ponto fixo seria calculado apenas uma vez e não resolvido através de um processo iterativo.

Os autores, inicialmente, chamaram essa abordagem de JS, que tem uma formulação relativamente simples: escolher os parâmetros estruturais de forma a maximizar a função de verossimilhança sujeita à restrição de que as variáveis endógenas sejam consistentes com o equilíbrio definido pelos parâmetros estruturais. Essa abordagem foi posteriormente chamada

de MPEC (*mathematical program with equilibrium constraints*) e é melhor descrita por Su e Judd (2012).

No entanto, a complexidade do problema de otimização aumenta consideravelmente através do método MPEC, já que a inclusão de um novo conjunto de variáveis no problema de maximização aumenta a dimensão do problema, sendo que as estimativas dos parâmetros relativos à maioria desses parâmetros são irrelevantes (geralmente chamados *nuisance parameters*). Em linhas gerais, a restrição de equilíbrio estabelece  $n$  restrições, em que  $n$  é a quantidade de estados possíveis do problema, mas os resultados mais relevantes do problema são apenas as estimativas dos parâmetros estruturais e não dos  $n$  multiplicadores de Lagrange.

Num problema descrito de maneira geral pela Eq. (2.6), utiliza-se a programação dinâmica para obter uma solução que determine as escolhas dos agentes em um período em função do estado no mesmo período e, mais ainda, que o estado em  $t$  seja expresso como uma resposta ao estado e a decisão em  $t - 1$ . Programação dinâmica é uma forma de tratar problemas de natureza sequencial que, sobre certas hipóteses permite obter soluções fechadas e bem definidas para problemas complexos. Em suma, o problema é visto de uma forma recursiva.

Mais formalmente, se  $x$  indica uma variável de estado e  $d$  uma variável de decisão, a solução será dada por  $d_t = v(x_t)$  e  $x_t = h(x_{t-1}, d_{t-1})$ . A partir dessa representação, podemos representar a Eq. (2.6) em formato recursivo. Seja o consumo  $c_t$  a variável de decisão, para todo  $t = 1, 2, \dots$  e  $(y_t, p_t)$  as variáveis de estado. Nessas condições,  $c_t = v(y_t, p_t)$  e  $(y_t, p_t) = h((y_{t-1}, p_{t-1}), v(y_{t-1}, p_{t-1}))$ . Escrevemos, então:

$$\begin{aligned} \max_{y_{t-1}, p_{t-1}} \quad & E \left[ \sum_{t=1}^{\infty} \phi^t v(h((y_{t-1}, p_{t-1}), v(y_{t-1}, p_{t-1}))) \right] \\ \text{s.a} \quad & y_0 = w \\ & y_t = c_t + s_t \end{aligned} \tag{2.8}$$

Note que, a partir da estrutura recursiva da Eq. (2.8), derivamos a regra ótima de consumo a partir do estado inicial  $(y_0, p_0)$ .

Na Eq. (2.8), a natureza recursiva do problema é expressa na maximização, no tempo  $t_0$ , de uma função nas decisões de consumo  $c_t$  tomadas pelo agente ao longo do tempo, de tal forma que cada decisão seja afetada pela decisão tomada no período anterior. Neste caso, a regra de decisão ótima a ser definida pode ser obtida a partir da maximização de subproblemas menores, que é o princípio de recursão, que permite desenvolver a equação de Bellman, que será mais aprofundada na seção seguinte. Em suma, no caso geral, dada a condição inicial, denotada pelo estado inicial  $x_0$ , pode-se obter uma política que maximize a função objetivo.

De acordo com Ljungqvist e Sargent (2000, pp. 88–89) são três os principais métodos empregados para resolução de problemas desta natureza: iteração da função valor, iteração da função política e o método de tentativa e erro para função valor, isto é, supõe-se uma solução

para o problema e verifica-se se ela de fato é ótima. O primeiro método, que é o utilizado por Rust (1987), de iteração da função valor, parte do princípio de que, dada a condição inicial  $x_0$  e que  $V_{t=0} = 0$ , pode-se obter a política ótima a partir da iteração sobre  $V_j$  até que convirja para um determinado valor. Note que, em problemas desse tipo, surge naturalmente a ideia de ponto fixo, que será aprofundada na seção seguinte.

No método de iteração da função política, sumariza Sutton e Barto (1998, pp. 91–94), define-se, primeiramente, uma função política inicial arbitrária, que mapeia cada decisão possível a uma probabilidade, dado um determinado estado. Em seguida, calcula-se a função valor para essa política anteriormente definida e, a partir desse valor obtido, atualiza-se a função política. O método segue de forma iterativa até que a regra de decisão ótima convirja.

## 2.3 Identificação de modelos estruturais e análise de políticas contrafactuais

Para argumentar que os parâmetros estimados do modelo são de fato estruturais e invariantes, será discutida o grau de identificação do modelo. O problema de identificação diz respeito à capacidade de um modelo de recuperar os parâmetros estruturais. No caso de um SEM, essa recuperação se dá a partir de equações na forma reduzida geralmente. No modelo de equações simultâneas descrito pelas Eq. (2.2) e Eq. (2.3), a partir dos coeficientes das formas reduzidas  $\pi_0$ ,  $\pi_1$ ,  $\pi_2$  e  $\pi_3$ , obtém-se quatro equações, mas os parâmetros estruturais a serem determinados são cinco, o que inviabiliza obter estimativas para todos os parâmetros estruturais.

Note, no entanto, que combinando as expressões de  $\pi_1$ ,  $\pi_3$ , pode-se escrever  $\psi_1 = \frac{\pi_3}{\pi_1}$  e, a partir de algumas manipulações algébricas sobre  $\pi_0$ ,  $\pi_2$ , obtém-se que  $\psi_0 = \pi_2 - \pi_0\psi_1$ . Com isso, podemos recuperar  $\psi_0$  e  $\psi_1$ , que são justamente os parâmetros estruturais da função de oferta na Eq. (2.3). Diz-se, por conseguinte, que a equação de oferta do sistema está identificada, ao passo que a equação de demanda não está identificada. Nos SEM, existem algumas condições para precisar o grau de identificação das equações descritas pelo modelo, como as condições de ordem e de posto, no entanto, elas não serão aprofundadas neste trabalho.

Para o caso de modelos estruturais do tipo DDC, que será desenvolvido no próximo capítulo, destaca-se que, segundo Rust (1994, p. 3125), na ausência de hipóteses fortes sobre as primitivas do modelo,  $(u, p, \beta)$ , o modelo é não-parametricamente não-identificado<sup>2</sup>, o que significa que existem infinitas primitivas que se adequem aos dados de forma consistente. Nessas condições, não se pode recuperar estimativas únicas para os parâmetros estruturais.

São consideradas suposições fortes, as hipóteses de separabilidade aditiva da função de utilidade do agente e a hipótese de independência condicional (a serem detalhadas na próxima

<sup>2</sup>Do inglês: *nonparametric not identified*.

seção), a especificação da função de utilidade dos agentes e a caracterização das probabilidades de transição entre os estados. Sob essas hipóteses, pode-se recuperar os parâmetros do modelo, o que significa que o modelo está identificado para tais especificações.

Note que, se diferentes restrições sobre o modelo, isto é, sobre as primitivas do modelo, implicam em diferentes parâmetros estruturais, fica evidente a sensibilidade de modelos DDC em relação a essas especificações. Segue que, já que a análise de políticas contrafactuais tem o intuito de verificar a mudança sobre os parâmetros do modelo mediante mudanças estruturais sobre as primitivas que o geraram, a análise dessas políticas é identificada.

Esse resultado é sumarizado por Kalouptsi, Scott e Souza-Rodrigues (2017). Em suma, mesmo que modelos DDC sejam, sobre hipóteses genéricas, não-parametricamente não-identificados e tornem-se identificados sobre hipóteses feitas com relação às primitivas, a análise do contrafactuais desses modelos está quase sempre identificada, uma vez que sobre condições menos fortes os contrafactuais estão identificados.

No que concerne à análise de políticas contrafactuais, essa é uma das principais vantagens da modelagem estrutural na economia, isto é, a possibilidade de se analisar realidades que nunca tenham sido verificadas. Nessas condições, pode-se estudar o comportamento de um agente ou economia quando inseridos numa realidade criada por determinada política econômica. A ideia geral por de trás da análise contrafactual é expressa como: Se um evento  $A$  ocorrer, outro evento  $B$  ocorrerá.

Segundo Balke e Pearl (2013) e Pearl (2013),  $A$  é dito antecedente contrafactual e não corresponde ao que foi observado na realidade e  $B$  é dito consequência contrafactual, isto é, aquilo que se espera que ocorra, uma vez que  $A$  ocorreu. Para clarificar, tome  $A$  como sendo a sentença “*O banco central eleva a taxa de juros básica da economia*” e tome a sentença “*O consumo agregado será reduzido*” como  $B$ . Nessas condições, se o banco central eleva a taxa de juros básica da economia, o consumo agregado será reduzido. Esse tipo de análise fomenta a criação de políticas econômicas, pois note que se  $B$  é indesejável, naturalmente uma política desejável será aquela que previne  $A$ .

Conduzir políticas contrafactuais a partir de modelos semi-experimentais gera uma inconsistência na análise como um todo, a qual é bem detalhada por Balke e Pearl (2013). Formulado e estimado um modelo como na Eq. (2.1), a previsão de uma política que atue sobre os regressores assume que tanto as realizações passadas e futuras dessas variáveis sejam exógenas. No entanto, mesmo sendo razoável supor que as futuras observações sejam determinadas fora do modelo, o mesmo não vale para as realizações passadas.

No exemplo de Balke e Pearl (2013, p.2), políticas de taxa são endógenas a um modelo econômico, dado que o governo é quem define os níveis de tributos e ele está sujeito a vários indicadores da economia e pressões diversas. No entanto, a taxa torna-se exógena nos modelos de forma reduzida quando se quer investigar o comportamento de outras variáveis, como

consumo ou produto, sujeito à variações nos impostos. Já os modelos estruturais, salientam Norets e Tang (2013), são cruciais para a análise dos efeitos de diferentes políticas sobre os parâmetros estruturais.

## Capítulo 3

# Estimação de Modelos Estruturais

### 3.1 Formulação do modelo

Nesta seção, será apresentado o modelo de controle ótimo em tempo infinito formulado por Rust (1987), no intuito de verificar se um agente toma suas decisões de forma ótima. Primeiramente, serão abordadas as derivações mais relevantes feitas por Rust (1987). O objetivo será tornar mais clara e intuitiva a explicação do modelo para que as equações obtidas possam ser implementadas para estimar os parâmetros relevantes.

Rust (1987) considera que o agente Harold Zurcher, superintendente de manutenção da *Madison Metropolitan Bus Company*, decide realizar ou não manutenção nos ônibus que param na estação. Os principais pressupostos do modelo são:

- O agente não é míope, isto é, Zurcher decide se realiza ou não a manutenção dos ônibus de forma que maximize sua utilidade esperada e descontada por um fator  $\beta \in (0, 1)$ ;
- $S_t$  denota o espaço de estados possíveis, em que  $x_t \in S_t$  é uma variável de estado no tempo  $t = 1, 2, \dots$ . A variável de estado relevante será a distância acumulada no odômetro de um veículo e, portanto, consideraremos  $S_t = S, \forall t$ ;
- $D_t(x_t)$  é o espaço de decisões passíveis de serem adotadas pelo agente no tempo  $t$ . Mais ainda,  $i_t \in D_t$  é a regra de decisão que denota se Zurcher decide realizar manutenção: quando  $i_t = 1$  há troca do motor e quando  $i_t = 0$  não é realizada manutenção. Por esta razão,  $D_t(x_t) = \{0, 1\}, \forall t \in \mathbb{N}$ ;
- O agente tem uma função de utilidade  $u(\cdot)$  sobre os custos nos quais incorre no decorrer do tempo. Esses custos são função da quilometragem acumulada dos veículos e da decisão do agente de realizar ou não a manutenção. Mais adiante será dado a forma funcional de  $u(\cdot)$ .

A regra de decisão que define se o agente realizará ou não a troca do motor de um ônibus no tempo  $t$  é do tipo  $i_t = \delta(x_t, \varepsilon_t, \theta)$ , isto é, a decisão de realizar ou não manutenção depende da distância acumulada do veículo no tempo  $t$ ,  $x_t$ , conhecida tanto pelo econometrista como pelo agente, de um conjunto de variáveis  $\varepsilon_t$  desconhecido pelo econometrista, mas conhecido pelo agente, e de um vetor de parâmetros estruturais desconhecidos  $\theta$  que deverá ser estimado. Em suma,  $\varepsilon_t$  é um termo de erro independente e identicamente distribuído (iid) que confere a relevância do estudo, pois caso não existisse estaria sendo investigada uma relação determinística.

Para melhor entendimento, o vetor de parâmetros pode ser visto como  $\theta = (\theta_1, \theta_2, \theta_3)$ , em que  $\theta_1$  é um vetor de parâmetros que caracterizam a função de utilidade do agente,  $\theta_2$  é um vetor de parâmetros que caracterizam a probabilidade de transição da variável  $\varepsilon$  e  $\theta_3$  é um vetor de parâmetros que caracterizam a probabilidade de transição da variável  $x$ .

Existem, então, duas variáveis de estado. Nesse sentido,  $\{(i_t, x_t) : i_t = \delta(x_t, \varepsilon_t)\}$  é um processo estocástico de controle. Mais ainda, o processo estocástico é regenerativo, o que significa dizer que é reiniciado toda vez que o agente decide fazer a manutenção do motor de um ônibus, quando é assumido que o ônibus esteja tão bom quanto se estivesse novo. Assim,  $i_t = 1$  implica que  $x_{t+1} = 0$ .

O problema de Zurcher é, portanto, sintetizado pela seguinte função objetivo:

$$V_\theta(x_0, \varepsilon_0) = \max_{i \in D(x)} E \left[ \sum_{t=0}^{\infty} \beta^t [u(x_t, i_t, \theta) + \varepsilon_t] \right] \quad (3.1)$$

Na Eq. (3.1), está expressa a hipótese de que o agente, em um estado inicial  $(x_0, \varepsilon_0)$  maximiza sua utilidade em cada estado posterior descontada por um fator  $\beta$ . Mas não somente isso, Zurcher maximiza sua utilidade esperada, em razão de existir uma distribuição de probabilidade sobre os estados, a qual será detalhada mais adiante. Na Eq. (3.1) é usada a hipótese de separabilidade aditiva que implica que  $u(x_t, i_t, \varepsilon_t) = u(x_t, i_t) + \varepsilon(i_t)$ .

Pelo Princípio da Otimalidade de Bellman, podemos escrever a Eq. (3.1) de forma recursiva, isto significa que a solução ótima de um problema de otimização é composta por soluções ótimas de subproblemas, que são menores e menos complexos. Supondo que o ponto inicial do processo seja  $(x_0, i_0, \varepsilon_0)$  seja dado, reescreve-se a Eq. (3.1) como:

$$V_\theta(x_0, \varepsilon_0) = \max_{i \in D(x)} \left[ u_0(x_0, i_0) + \varepsilon_0 + E \left[ \sum_{t=1}^{\infty} \beta^t [u(x_t, i_t, \theta) + \varepsilon_t] \right] \right] \quad (3.2)$$

Agora, quer-se reescrever a esperança do lado direito da Eq. (3.2) em termos de  $V_\theta(x_0, \varepsilon_0)$ . Note que modificando o índice do somatório e evidenciando o fato de desconto  $\beta$ , tem-se que:

$$E \left[ \sum_{t=1}^{\infty} \beta^t [u(x_t, i_t, \theta) + \varepsilon_t] \right] = E \left[ \sum_{t=0}^{\infty} \beta^{t+1} [u(x_{t+1}, i_{t+1}, \theta) + \varepsilon_{t+1}] \right] = \beta E \left[ \sum_{t=0}^{\infty} \beta^t [u(x_{t+1}, i_{t+1}, \theta) + \varepsilon_{t+1}] \right] \quad (3.3)$$

O termo que aparece na ltima igualdade s pode ser computado a partir da distribuico induzida pelo estado anterior  $(x_t, \varepsilon_t)$ . Assim, definindo:

$$EV_\theta(x_t, \varepsilon_t) := E \left[ \sum_{t=0}^{\infty} \beta^t [u(x_{t+1}, i_{t+1}, \theta) + \varepsilon_{t+1}] \right] \quad (3.4)$$

 obtido que:

$$V_\theta(x_0, \varepsilon_0) = \max_{i \in D(x)} [u(x_0, i_0, \theta) + \varepsilon_0 + \beta EV_\theta(x_0, \varepsilon_0)] \quad (3.5)$$

O termo  $EV_\theta(x_t, \varepsilon_t)$ , portanto, representa o somatrio das utilidades esperadas do indivduo em cada momento do tempo ponderado pela probabilidade de o indivduo alcanar cada um desses estados, dado que ele parte de  $(x_0, \varepsilon_0)$ . Recursivamente, todo o problema poder ser expresso em termos do ponto de partida inicial do processo, conforme a Eq. (3.5). No caso geral,  $EV_\theta(x_t, \varepsilon_t)$   calculado como uma integral nas probabilidades de transio, isto :

$$EV_\theta(x_t, \varepsilon_t) := \int_0^{\infty} V_\theta(x_{t+1}, \varepsilon_{t+1}) \cdot p(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i_t, \theta) \quad (3.6)$$

No entanto, o interesse real desse trabalho  para o caso discreto. Assim, sob algumas hipteses sobre as probabilidades de transio pode-se simplificar a Eq. (3.6) e tornar sua implementaco mais simples. Em suma, no caso em que a varivel  $x$   retirada de um conjunto finito, podemos expressar Eq. (3.6) como uma multiplicaco pela esquerda da matriz de transio do problema, que ser detalhada no final desta seo.

A equaco de Bellman, escrita na Eq. (3.5), caracteriza a modelagem de um problema de *Dynamic Discrete Choice* (DDC) e prev o comportamento de Zurcher enquanto agente racional maximizador de utilidade. O intuito, agora,  conhecer o vetor de parmetros estruturais  $(\beta, \theta)$ . Para tanto, utiliza-se uma base de dados do tipo  $(i_1, i_2, \dots, i_T, x_1, x_2, \dots, x_T)$  para verificar se as decises de Zurcher coincidem com o timo esperado pela teoria econmica, dado o modelo de controle timo formulado. Naturalmente,  necessrio obter a funo de mxima verossimilhana dos dados.

Para compor a funo de verossimilhana da amostra, convm, primeiramente, detalhar as funes de probabilidades de transio do modelo, decritas na Eq. (3.6). Sob a hiptese de independncia condicional assumida por Rust (1987), a densidade condicional de  $x_t, \varepsilon_t$  satisfaz a:

$$p(x_t, \varepsilon_t | x_{t-1}, \varepsilon_{t-1}, i_{t-1}, \theta) = \pi(\varepsilon_t | x_t, \theta_2) \cdot q(x_t | x_{t-1}, i_{t-1}, \theta_3) \quad (3.7)$$

Perceba que  $\varepsilon$   um rudo, um conjunto de variveis no observveis pelo pesquisador, de tal forma que a funo densidade a ela associada possa ser negligencivel. Assim, para caracterizar completamente a probabilidade de transio  $q$ , na Eq. (3.7),  necessrio descrever

a probabilidade de o indivíduo optar pela decisão  $i_t$ , que é do tipo  $P(i_t|x_t, \theta)$  dada pela função logit dinâmica.

Para definir  $P(i_t|x_t, \theta)$  faz-se necessária a hipótese de que as observações de  $\varepsilon_t$ , a parcela não observável da função de utilidade do agente, sejam independentes entre si e provenientes de uma distribuição de valor extremo do tipo I ou distribuição de Gumbel. Sabe-se que a probabilidade de o agente optar por fazer a reposição do motor ( $i_t = 1$ ) é a mesma probabilidade de essa substituição lhe gerar um nível maior de utilidade do que não fazer a manutenção ( $i_t = 0$ ), isto é:

$$\begin{aligned} P(i_t|x_t, \theta) &= \text{Prob}(u_1 + \varepsilon_1 > u_0 + \varepsilon_0) \\ &= \text{Prob}(\varepsilon_0 < \varepsilon_1 + u_1 - u_0) \end{aligned} \quad (3.8)$$

Algumas manipulações algébricas sobre a Eq. (3.8), apresentadas no Apêndice B, permitem deduzir que:

$$P(i_t|x_t, \theta) = \frac{\exp\{u(x, i, \theta) + \beta EV_\theta(x, i)\}}{\sum_{i' \in D(x)} \exp\{u(x, i', \theta) + \beta EV_\theta(x, i')\}} \quad (3.9)$$

As vantagens de uma modelagem conforme a Eq. (3.9) são, segundo Train (2002, pp. 50–53), não somente os fatos de que as probabilidades estão necessariamente entre 0 e 1 e que somam 1, quando consideradas todas as alternativas, mas também o fato de poderem captar mudanças das preferências dos agentes, uma vez que a função de utilidade é explícita em seu cômputo. Nessas condições, a função de verossimilhança será dada por:

$$l(i_1, i_2, \dots, i_T, x_1, x_2, \dots, x_T|x_0, i_0, \theta) = \prod_{t=1}^T P(i_t|x_t, \theta)q(x_t|x_{t-1}, i_{t-1}, \theta_3) \quad (3.10)$$

Como a estimação dos parâmetros é feita em dois estágios separados, como será visto mais adiante, escreve-se a Eq. (3.10) como:

$$l(i_1, i_2, \dots, i_T, x_1, x_2, \dots, x_T|x_0, i_0, \theta) = l_1 \cdot l_2 \quad (3.11)$$

em que:

$$l_1(i_1, i_2, \dots, i_T, x_1, x_2, \dots, x_T|x_0, i_0, \theta) = \prod_{t=1}^T q(x_t|x_{t-1}, i_{t-1}, \theta_3) \quad (3.12)$$

e:

$$l_2(i_1, i_2, \dots, i_T, x_1, x_2, \dots, x_T|x_0, i_0, \theta) = \prod_{t=1}^T P(i_t|x_t, \theta) \quad (3.13)$$

Quer-se, ento, estimar os parmetros da funo de mxima verossimilhana na Eq. (3.10). Para tanto, sero discutidas, primeiramente, as hipteses feitas sobre as probabilidade de transio, j que essas podem ser estimadas de forma descomplicada. Por outro lado, para o clculo da probabilidade de deciso expressa na Eq. (3.9), torna-se necessrio obter o valor da funo  $EV_\theta(x, i)$ , que  um ponto fixo nico de uma contrao.

A probabilidade de transio do processo markoviano  retirada de uma funo densidade  $g(\cdot)$  conforme:

$$q(x_t|x_{t-1}, i_{t-1}, \theta_3) = \begin{cases} g(x_t - x_{t-1}, \theta_3) & \text{se } i_{t-1} = 0 \\ g(x_t - 0, \theta_3) & \text{se } i_{t-1} = 1 \end{cases} \quad (3.14)$$

Convm, portanto, detalhar a funo de distribuio  $g(\cdot)$ . Assuma que  $x_t$   retirado de uma seqncia crescente nos ndices  $\{x^1, x^2, \dots, x^H\}$ , em que  $x^{h'} - x^h = \alpha \cdot (h' - h)$ , isto , o estado de cada nibus num determinado perodo  retirado de uma distribuio de estados igualmente espaadas por  $\alpha$ . Rust toma  $\alpha = 5000$  e  $H = 90$ . Essa escolha  motivada pelo fato de que, na amostra, no existem veculos com quilometragem superior a 450.000 em nenhum momento do tempo.

Nestas condioes, Rust escolhe modelar  $g$  usando uma distribuio trinomial da seguinte forma:

$$g(x_t - x_{t-1}, \theta_3) = \begin{cases} \theta_{30}, & \text{se } x_t - x_{t-1} \in [0, 5000) \\ \theta_{31}, & \text{se } x_t - x_{t-1} \in [5000, 10000) \\ \theta_{32}, & \text{se } x_t - x_{t-1} \in [10000, +\infty) \end{cases} \quad (3.15)$$

Essa escolha reduz o nmero de parmetros a serem estimados, mas vale salientar que o nmero de parmetros a ser estimado  flexvel e pode mudar facilmente de acordo com a natureza do problema. Como  $g$   uma funo de densidade,  $\theta_{32}$  pode ser descrito como uma combinao linear de  $\theta_{30}$  e  $\theta_{31}$ .

Feita esta especificao, a estimaco dos parmetros em  $\theta_3$  pode ser feita de forma paramtrica ou no-paramtrica. Para este trabalho, as probabilidades de transio foram estimadas a partir das frequncias observadas das realizaoes de  $g$ , a partir da Eq. (3.12) sob a especificao feita na Eq. (3.15). Note que a estimaco de  $\theta_3$  no exige o cmputo do ponto fixo  $EV_\theta$ , o que permite que uma implementao mais simples.

As estimativas para as probabilidades compooa a matriz de transio do processo. Dado que o conjunto de estados possveis  discretizado em 90 intervalos, a probabilidade de transio  $p(dy|x, i, \theta_3)$  pode ser denotada por uma matriz estocstica de ordem  $n = 90$ . Usando a hiptese da distribuio trinomial, podemos escrever  $\theta_{3j} = P\{x_{t+1} = x_t + j|x_t, i_t\}$ ,  $j = 0, 1, 2$ . Nessas condioes, quando o agente no realiza reposio, isto ,  $i_t = 0, \forall t$ , a matriz de Markov do processo,  $T_0$ , pode ser escrita como:

$$T_0 = \begin{bmatrix} \theta_{30} & \theta_{31} & \theta_{32} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \theta_{30} & \theta_{31} & \theta_{32} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \theta_{30} & \theta_{31} & \theta_{32} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \theta_{30} & \theta_{31} & \theta_{32} \\ 0 & 0 & \cdots & 0 & 0 & 0 & \theta_{30} & 1 - \theta_{30} \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.16)$$

Será visto agora como que a matriz de transição na Eq. (3.16) pode ser usada para simplificar a Eq. (3.6), de tal forma que a nova equação obtida para  $EV_\theta(x)$  seja aquela que de fato é implementada pelo algoritmo de estimação dos demais parâmetros do modelo. Para tanto, serão introduzidos alguns aspectos técnicos relevantes.

**Definição:** Seja  $(S, d)$  um espaço métrico. Diz-se que uma aplicação contínua  $T : S \rightarrow S$  é uma contração de módulo  $\alpha$  se,  $\exists \alpha \in [0, 1)$  tq

$$d(T(x), T(y)) \leq \alpha d(x, y), \forall x, y \in S \quad (3.17)$$

Pensando em  $d = d : S \times S \rightarrow \mathbb{R}$  como sendo a distância entre dois pontos de um espaço  $S$ , uma contração é um operador que aproxima os pontos a ela pertencentes. Mais ainda, se  $x \in S$  é tal que  $T(x) = x$ ,  $x$  é dito ponto fixo de  $T$ .

**Teorema 1** (*Teorema do ponto fixo de Banach*): Se  $T : S \rightarrow S$  é uma contração e  $S$  é um espaço completo, existe um ponto fixo  $x$  para  $T$  e ele é único.

A demonstração do Teorema 1 é apresentada no Apêndice A. Nessas condições, Rust (1994, pp. 3093–3094) mostra que é possível definir uma contração  $\Gamma$  em um espaço de funções completo, de forma similar a Eq. (3.5), de tal maneira que  $EV_\theta(x)$  seja o seu ponto fixo. Isso implica que podemos calcular  $EV_\theta(x)$  e estimar os parâmetros estruturais do problema.

Agora, será caracterizado  $EV_\theta(x, i)$ , para que seja possível calculá-lo. Para tanto, especifica-se, primeiramente, o formato da função de utilidade. Considere:

$$u(x_t, i_t, \theta_1) + \varepsilon(i_t) = \begin{cases} -RC - c(0, \theta_1) + \varepsilon_t(1), & \text{se } i_t = 1 \\ -c(x_t, \theta_1) + \varepsilon_t(0), & \text{se } i_t = 0 \end{cases} \quad (3.18)$$

Em que  $RC$  é um parâmetro a ser estimado pelo modelo que denota o custo total de reposição do motor do ônibus e  $c(x_t, \theta_1)$  é uma função diferenciável e crescente em  $x_t$  que expressa custos de manutenção dos veículos quando o estado é  $x_t$ .

Por outro lado, o termo  $V_\theta(x, \varepsilon)$ , expresso na Eq. (3.5), é dado pelo máximo de duas funções, já que  $i_t \in D(x_t) \equiv \{0, 1\}$ . Assim, denote por  $y$  e  $j$ , respectivamente, o estado e a escolha no período atual respectivamente e por  $x$  e  $i$ , respectivamente, o estado e a escolha

correspondentes ao perodo anterior. Pode-se, ento, escrever:

$$V_\theta(y, \varepsilon) = \max_{j \in D(x)} [u(y, j, \theta) + \varepsilon + \beta EV_\theta(y, \varepsilon, j)] \quad (3.19)$$

O intuito das proximas equaces ser expurgar o termo de erro da Eq. (3.19), para que o computo de seu valor esperado possa ser feito atravs de um processo iterativo sobre uma nica varivel de estado, que  a quilometragem. Note que, dados  $x$  e  $i$ ,  $y$  e  $\varepsilon$  podem assumir uma srie de valores com probabilidades distintas. Tratam-se, portanto, de variveis aleatrias. Nessas condies, aplicando na Eq. (3.19) o operador de esperana condicional com respeito s variveis de estado  $y$  e  $\varepsilon$  condicional nas observaes anteriores de  $x$  e  $i$ , denotado por  $E_{y,\varepsilon}[\cdot|x, i] = E_{y,\varepsilon|x,i}[\cdot]$ , obtm-se:

$$E_{y,\varepsilon}[V_\theta(y, \varepsilon)|x, i] = E_{y,\varepsilon|x,i} \left[ \max_{j \in D(x)} [u(y, j, \theta) + \varepsilon + \beta EV_\theta(y, \varepsilon, j)] \right] \quad (3.20)$$

Note que na Eq. (3.20) h uma esperana conjunta. Para trat-la mais formalmente,  feita a seguinte definio.

**Definio:** Sejam  $X$  e  $Y$  variveis aleatrias contnuas e  $f(x, y)$  a funo de distribuico conjunta. A funo de distribuico marginal de  $f$  com relao a  $x$   dada por:

$$f(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad (3.21)$$

e analogamente definida para  $y$ . Mais ainda, se  $\lambda(x, y)$   uma aplicao a  $\mathbb{R}$  em que  $x$   retirado de  $X$  e  $y$   retirado de  $Y$ , a esperana de  $\lambda$  com respeito a  $y$  condicional em  $x$   dada por:

$$E_{y|x}[\lambda(x, y)] = E_y[\lambda(x, y)|x] = \int_{-\infty}^{+\infty} \lambda(x, y) \cdot f(x, y) dy \quad (3.22)$$

e analogamente definida para  $y$ .

A funo de distribuico marginal informa, no caso da Eq. (3.21), a probabilidade de a varivel aleatria  $X = x$ . Por essa razo,  obtida atravs da integrao da densidade conjunta sobre  $y$ , dado que para que  $X = x$ ,  $Y$  pode assumir uma infinidade de valores.

Voltando  Eq. (3.20) e usando a Eq. (3.22) e o fato de que  $f(y, \varepsilon)$   a funo de densidade conjunta das variveis de estado, temos que a esperana de  $V_\theta(y, \varepsilon)$   do tipo:

$$E_{y,\varepsilon}[V_\theta(y, \varepsilon)|x, i] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(y, \varepsilon|x, i) \cdot V_\theta(y, \varepsilon) d\varepsilon dy \quad (3.23)$$

Repare que  $\int_{-\infty}^{+\infty} f(y, \varepsilon) \cdot V_\theta(y, \varepsilon) d\varepsilon = E_\varepsilon[V_\theta(y, \varepsilon)|y] = E_{\varepsilon|y}[V_\theta(y, \varepsilon)]$ , pois a integral  avaliada unicamente no diferencial de  $\varepsilon$ . Assim, fica fcil perceber que a Eq. (3.23) pode ser reescrita na forma:

$$E_{y,\varepsilon|x,i}[V_\theta(y, \varepsilon)] = E_{y|x,i} [E_{\varepsilon|y,x,i}[V_\theta(y, \varepsilon)]] \quad (3.24)$$

Definindo  $EV_\theta(x, i) := E_{y,\varepsilon}[V_\theta(y, \varepsilon)|x, i]$ , tem-se, então que:

$$EV_\theta(x, i) = E_{y|x,i} \left[ E_{\varepsilon|y,x,i} \left[ \max_{j \in D(x)} [u(y, j, \theta) + \varepsilon + \beta EV_\theta(y, \varepsilon, j)] \right] \right] \quad (3.25)$$

Usando o fato de que  $\varepsilon$  é proveniente de uma distribuição de valor extremo do tipo I, podemos fazer a seguinte aproximação:

$$E_{\varepsilon|y,x,i} \left[ \max_{j \in D(x)} [u(y, j, \theta) + \varepsilon + \beta EV_\theta(y, \varepsilon, j)] \right] = \log \left[ \sum_{j=0,1} \exp\{u(y, j, \theta) + \beta EV_\theta(y, \varepsilon, j)\} \right] \quad (3.26)$$

Essa aproximação é derivada da Análise Convexa, que não será aprofundada neste trabalho, pois foge ao propósito principal, mas é disponibilizada com mais detalhes em Chiong, Galichon e Shum (2016). Essa aproximação é dita LogSumExp (LSE), isto é, dado  $\{x_1, \dots, x_n\}, x_i \in \mathbb{R}, \forall i$  tem-se que  $\max(x_1, \dots, x_n) \approx LSE(x_1, \dots, x_n) := \log(\exp(x_1) + \dots + \exp(x_n))$ . Essa aproximação torna-se especialmente conveniente porque faz com que a implementação seja feita de maneira mais simples, já que dispensa um processo iterativo em que seria computada a função duas vezes e a partir daí o valor máximo seria obtido. Nessas condições, a Eq. (3.25) torna-se:

$$EV_\theta(x, i) = E_{y|x,i} \left\{ \log \left[ \sum_{j=0,1} \exp\{u(y, j, \theta) + \beta EV_\theta(y, \varepsilon, j)\} \right] \right\} \quad (3.27)$$

$$= \int_y \log \left[ \sum_{j=0,1} \exp\{u(y, j, \theta) + \beta EV_\theta(y, \varepsilon, j)\} \right] p(y|x, i, \theta_3) dy \quad (3.28)$$

Uma vez que a variável de decisão apresenta apenas dois valores, escreve-se a Eq. (3.28) na forma extensa a seguir:

$$EV_\theta(x) = \int_y \log [\exp\{-c(y, \theta_1) + \beta EV_\theta(y)\} + \exp\{-RC - c(0, \theta_1) + \beta EV_\theta(0)\}] p(y|x, i, \theta_3) dy \quad (3.29)$$

Substituindo o resultado obtido na Eq. (3.29) na Eq. (3.9) tem-se que a probabilidade de escolha condicional, fazendo  $i_t = 0$  e fazendo a razão do numerador e do denominador da Eq. (3.9) por  $\exp\{u(x, 0, \theta) + \beta EV_\theta(x, 0)\}$ , será escrita como:

$$P(i_t = 0|x_t, \theta) = \frac{1}{1 + \exp\{c(x, \theta_1) - c(0, \theta_1) - RC + \beta EV_\theta(0) - \beta EV_\theta(x)\}} \quad (3.30)$$

Assim,  $P(i_t = 1|x_t, \theta) = 1 - P(i_t = 0|x_t, \theta)$ . Para implementação do segundo passo da estimação de parâmetros, basta calcular o ponto fixo  $EV_\theta(x)$  para cada iteração dos parâmetros. Calculado o ponto fixo, pode-se estimar os demais parâmetros a partir da Eq. (3.13).

Pode-se, agora, escrever a Eq. (3.29) em termos da matriz de transição expressa na Eq. (3.16). Sejam  $c, rc, EV \in \mathbb{R}^{90}$ , tais que:

$$c = \begin{bmatrix} c(0, \theta_1) \\ c(1, \theta_1) \\ \vdots \\ c(89, \theta_1) \end{bmatrix}, rc = \begin{bmatrix} RC \\ RC \\ \vdots \\ RC \end{bmatrix}, EV = \begin{bmatrix} EV_\theta(0, \theta_1) \\ EV_\theta(1, \theta_1) \\ \vdots \\ EV_\theta(89, \theta_1) \end{bmatrix} \quad (3.31)$$

As representações acima tornam-se válidas para o caso discreto em que a variável de estado é discretizada em 90 intervalos. Nessas condições, a Eq. (3.29) será representada por:

$$EV = T_0 \cdot \log [\exp\{-c + \beta EV\} + \exp\{-rc - c_1 + \beta EV_1\}] \quad (3.32)$$

Nessas condições, obtemos que a  $i$ -ésima posição do vetor  $EV$  será descrita como:

$$\begin{aligned} EV(i) = & \theta_{30} \log [\exp\{-c(i|\theta_1) + \beta EV(i)\} + \exp\{-rc - c(0|\theta_1) + \beta EV(0)\}] + \\ & \theta_{31} \log [\exp\{-c(i+1|\theta_1) + \beta EV(i+1)\} + \exp\{-rc - c(0|\theta_1) + \beta EV(0)\}] + \\ & \theta_{32} \log [\exp\{-c(i+2|\theta_1) + \beta EV(i+2)\} + \exp\{-rc - c(0|\theta_1) + \beta EV(0)\}] + \end{aligned} \quad (3.33)$$

$EV(i)$  representa a esperança da máxima utilidade do agente dado que sua posição inicial é  $i$ . Note que a Eq. (3.33) é justamente o caso discreto da Eq (3.29). Nessas condições, a Eq. (3.32) é aquela que de fato será implementada na seção seguinte para estimação dos parâmetros.

## 3.2 Base de dados

Com relação aos dados utilizados por este trabalho, destaca-se, primeiramente, que não são os mesmos utilizados por Rust (1987), uma vez que os dados originais estão disponíveis no formato da linguagem GAUSS, atualmente pouco utilizada, e não puderam ser convertidos em algum outro formato que preservasse todas as suas características.

Nessas condições, os dados utilizados são provenientes de uma base em formato semelhante à original, isto é, contém dados do tipo painel, que são conjuntos de séries temporais para

104 de ônibus (discriminados em 4 grupos), que totalizam 8260 observações. As informações relevantes disponíveis são modelo e ano do ônibus, se houve ou não reposição em dado momento e quilometragem acumulada. Vale destacar que não foram obtidas mesmas quantidades de dados para cada ônibus. O tratamento dos dados para estimação dos parâmetros estruturais é descrito na próxima seção.

### 3.3 Implementação do algoritmo

A partir deste ponto, fica claro perceber que o vetor de parâmetros a ser estimado é  $\theta = (\beta, \theta_1, RC, \theta_3)$ , em que  $\beta$  denota o parâmetro de desconto no problema de maximização de Harold Zurcher,  $\theta_1$  denota um subconjunto de parâmetros da função de utilidade do agente (especificamente da função de custo de manutenção),  $RC$  representa o custo de reposição do motor do ônibus e  $\theta_3$  denota um subconjunto de parâmetros das probabilidades de transição de Markov. O parâmetro  $\beta$ , no entanto, não será estimado, uma vez que ele é tomado como uma primitiva do problema. Isso significa que ele deve ser especificado anteriormente pelo pesquisador para analisar, por exemplo, a variação dos parâmetros estruturais à medida que o agente valoriza mais ou menos sua utilidade futura.

Agora, será implementada a solução para o problema do agente Harold Zurcher, derivado na Seção anterior, no intuito de estimar o vetor de parâmetros  $\theta$ . No entanto, optou-se por utilizar a linguagem Python em vez da linguagem GAUSS utilizada por Rust. Isso se deu em virtude de Python ser uma linguagem mais moderna e avançada, que é orientada a objetos e permite maior clareza e simplicidade do algoritmo. O Alg. (1) sumariza a implantação de todas as etapas descritas anteriormente.

---

**Alg. 1:** Algoritmo para estimação dos parâmetros através do método NFXP.

---

**Dados:** Séries de tempo com as características de cada ônibus que passa pela estação central.

**Resultados:** Estimativas dos parâmetros estruturais do modelo.

- 1 Importação e organização dos dados;
  - 2 Estimação do vetor  $\theta_3$  das probabilidades de transição;
  - 3 **while**  $l_2^\tau < l_2^{\tau+1}$  **do**
  - 4     **while**  $T(EV) \neq EV$  **do**
  - 5         | O vetor em (3.32) é calculado;
  - 6     **end**
  - 7     A função de verossimilhança é calculada a partir do ponto fixo obtido;
  - 8     O valor da função de verossimilhança em (3.13) é comparada ao valor obtido na iteração anterior.
  - 9 **end**
  - 10 Maximizada a função de máxima verossimilhança, os parâmetros são retornados.
-

No Alg. (1), a linha 3   onde se inicia o processo iterativo em que a funo de m xima verossimilhana na Eq. (3.13)   maximizada. Vale destacar que  $\tau$  indica a iterao corrente do processo. Note que para cada palpite dos par metros  $(RC, \theta_1)$ , o ponto fixo   calculado atrav s de um processo iterativo e, seguinte a isto,   obtido o valor da funo de verossimilhana. O procedimento somente   encerrado quando a funo de verossimilhana   maximizada. Fica evidente, portanto, que o procedimento tem um maior grau de complexidade do que aquele exigido para estimaco de MQO.

Agora ser  descrito o processo de importao e ajuste dos dados. No Alg. (2), na linha 1, os dados s o importados do arquivo *dad.mat*. Os dados s o dispon veis numa matriz de seis colunas: a primeira cont m o ID do  nibus, a segunda cont m o modelo, a terceira cont m o ano da compra, a quarta o m s da compra, a quinta indica se houve ou n o reposio e a  ltima indica a dist ncia registrada no od metro. Vale destacar que a primeira coluna   indexada por 0 em Python.

Nas linhas 3–4, os dados para os 4 grupos de  nibus dispon veis s o concatenados numa  nica matriz, *data\_base*, e nas linhas 5–7, s o adicionadas 3 colunas   matriz.

---

**Alg. 2:** Importao e gerao de dados que alimentam o algoritmo de otimizao.

---

```

1  mat = scipy.io.loadmat('dad.mat')
2
3  data_base = np.vstack((mat['d203'], mat['d50'],
4                        mat['d530875'], mat['d870']))
5  data_base = np.hstack((data_base,
6                        np.zeros((data_base.shape[0], 3),
7                                dtype=data_base.dtype)))
8
9  for j in range(0,data_base.shape[0]-2):
10     if data_base[j,4] == 1:
11         data_base[j,6] = data_base[j+1,5]
12     else:
13         if data_base[j+1,0] == data_base[j,0]:
14             data_base[j,6] = data_base[j+1,5] - data_base[j,5]
15         else:
16             data_base[j,6] = -1
17
18  for i in range(0,data_base.shape[0]-1):
19     if 0 <= data_base[i,6] < 5000:
20         data_base[i,7] = 0
21     elif 5000 <= data_base[i,6] < 10000:
22         data_base[i,7] = 1
23     elif data_base[i,6] >= 10000:
24         data_base[i,7] = 2
25     elif data_base[i,6] == -1:
26         data_base[i,7] = 3
27
28  for i in range(0,data_base.shape[0]):
29     data_base[i,8] = (data_base[i,5]/5000)+1

```

---

A primeira coluna adicionada   preenchida com a diferena das quilometragens (linhas 9–16), sendo levadas em conta a troca do ID do  nibus na amostra e se houve ou n o reposio

no período anterior. A segunda coluna registra a frequência dos das diferenças de quilometragem conforme a distribuição  $g$  (linhas 18-26), isto é, em qual intervalo da Eq. (3.15) os valores se situam, e a última coluna contém a distância acumulada de forma discretizada (linhas 28–29). Note que o teste das linhas 25–26, quando a diferença das quilometragens for -1 diz respeito ao momento em que o ID do ônibus muda na matriz, valor este gerado na linha 16.

O primeiro passo a ser executado após a importação e organização dos dados é a estimação de  $\theta_3$  de ordem  $3 \times 1$  (conforme a Eq. (3.15)), o vetor que contém as probabilidades de transição, que pode ser feita de forma direta a partir da função de máxima verossimilhança parcial em Eq. (3.12).

---

**Alg. 3:** Composição da função de verossimilhança dos dados relativos às probabilidades de transição e estimação de  $\theta_3$ .

---

```

1  class theta_3_estimation(GenericLikelihoodModel):
2
3      def __init__(self, data):
4          self.endog = data[:,4]
5          self.exog = data[:,8]
6          self.frequency = itemfreq(data[:,[7]])
7          super(theta_3_estimation, self).__init__(self.endog, self.exog)
8
9      def nloglikeobs(self, params):
10         log_func = (-self.frequency[0,1]*np.log(params[0])
11                    -self.frequency[1,1]*np.log(params[1])
12                    -self.frequency[2,1]*np.log(1-params[0]-params[1]))
13         return log_func
14
15     def fit(self, start_params = [0.3,0.3], maxiter=10000, maxfun=5000):
16         return super(theta_3_estimation, self).fit(start_params=start_params,
17                                                    maxiter=maxiter,
18                                                    maxfun=maxfun)

```

---

No Alg. (3) é criada a classe *theta\_3\_estimation* para estimação de  $\theta_3$ . A classe herda métodos da classe *GenericLikelihoodModels* (super classes ou *mother classes* são declaradas como um parâmetro quando uma nova classe é criada, conforme a linha 1), do pacote *statsmodels.base.model*, que contém rotinas para estimação de parâmetros por máxima verossimilhança. A super classe *GenericLikelihoodModels* exige que sejam especificadas as variáveis endógenas (na linha 4 é especificada a variável endógena como sendo a decisão de repor ou não o motor) e exógenas (na linha 5 é especificada a variável exógena como sendo a variável de estado discretizada). No entanto, essa é uma mera condição para tornar possível o uso dos métodos da superclasse.

Um apontamento torna-se relevante. O método `__init__`, que é o construtor da classe, é iniciado automaticamente quando se instancia um objeto dela. Por essa razão ele é comumente usado para definir atributos da classe. Em virtude disso, usa-se a função *super* na linha 7, que chama o construtor da super classe para definir seus atributos como atributos da classe filha no

momento em que um objeto é instanciado. Em suma, um objeto de uma classe é um conjunto de dados e de métodos.

O método *nloglikeobs* retorna  $-\log(\cdot)$  da função de verossimilhança da amostra, baseada na frequência observada da variável  $X_t := x_t - x_{t-1}$ ,  $X_t \in \{0, 1, 2\}$  (gerada na linha 6); o método *fit* maximiza a função retornada por *nloglikeobs* através de um procedimento iterativo.

---

**Alg. 4:** Método *p\_transition\_f* da classe *theta\_3\_estimation*.

---

```

1 def p_transition_f(self):
2     self.model_e = self.fit()
3     print(self.model_e.summary())
4     p_transition = np.zeros((1,3))
5     p_transition[0,0]=self.model_e.params[0]
6     p_transition[0,1]=self.model_e.params[1]
7     p_transition[0,2]=1-self.model_e.params[0]-self.model_e.params[1]
8     return p_transition

```

---

O último método da classe *theta\_3\_estimation*, *p\_transition\_f*, apresentado no Alg. (4), chama a função *fit* (linha 2) e retorna o vetor transposto das probabilidades de transição estimadas, formato esse que será utilizado por procedimentos da próxima classe. Todos esses são métodos de uma classe e, portanto, para serem executados é necessário, primeiro, instanciar um objeto desta classe.

---

**Alg. 5:** Construtor da classe *dynamicmodel*, na qual ocorre a contração do ponto fixo e estimação dos parâmetros estruturais.

---

```

1 class dynamicmodel(theta_3_estimation):
2
3     def __init__(self, data, costs_f, beta):
4         self.obj = theta_3_estimation(data)
5         self.trans_prob = self.obj.p_transition_f()
6         self.cost_f = costs_f
7         self.bet = beta
8         super(dynamicmodel, self).__init__(data)

```

---

No Alg. (5), é apresentado o construtor da classe *dynamicmodel* que é filha da classe *theta\_3\_estimation* e, por conseguinte, herda métodos da classe *GenericLikelihoodModels*. Na linha 4 é instanciado um objeto da classe *theta\_3\_estimation*, para que seja utilizado um método da super classe no construtor. O atributo *trans\_prob* (linha 5) recebe um vetor transposto com as probabilidades de transição. As probabilidades de transição contidas em *trans\_prob* irão compor a matriz de transição, a ser definida mais adiante.

O Alg. (6) contém o método *dynamic\_est* da classe *dynamicmodel*, o qual realiza a contração do ponto fixo expressa na Eq. (3.29). Nas linhas 2–5 são criados, respectivamente, o vetor das probabilidades de transição, o vetor com todos os valores possíveis da variável de

**Alg. 6:** Contração do ponto fixo.

---

```

1 def dynamic_est(self, params):
2     theta_3 = self.trans_prob
3     x = np.transpose(np.array([range(0,89)]))
4     tr_temp = params[0]
5     theta_11 = params[1]
6     self.c = np.zeros((90,1))
7     self.tr = np.zeros((90,1))
8     ev = np.zeros((90,1))
9     for x in range(1,91):
10        self.c[x-1,0]=self.cost_f(theta_11, theta_12, x)
11        self.tr[x-1,0]=tr_temp
12    T_Pi = np.zeros((90,90))
13    for i in range(0,88):
14        T_Pi[i,i] = theta_3[0,0]
15        T_Pi[i,i+1] = theta_3[0,1]
16        T_Pi[i,i+2] = theta_3[0,2]
17    T_Pi[88,88] = theta_3[0,0]
18    T_Pi[88,89] = theta_3[0,1]
19    T_Pi[89,89] = 1
20    ev1 = np.zeros((90,1))+100
21    aux = ev
22    while np.max(np.abs(ev1-aux)) > 1e-8:
23        ev1 = np.dot(T_Pi, np.log(np.exp(-self.c+self.bet*ev)+
24                                np.exp(-self.tr-self.c[0,0]+self.bet*ev[0,0])))
25        aux = ev
26        ev = ev1
27    return ev

```

---

estado, o primeiro parâmetro a ser estimado (custo de reposição), o segundo parâmetro a ser estimado. Nas linhas 6, 7, 9, 10 e 11 são criados o vetor com o custo incorrido em cada estado e o vetor de ordem 90 preenchido com o custo de reposição a ser estimado (trata-se de um vetor para facilitar a implementação).

Na linha 8 é declarado o vetor  $ev$ , também de ordem 90. Note que, a partir da Eq. (3.29),  $EV_\theta$  é colocado como função de  $x$  somente (excetuando-se a dependência nos parâmetros a serem estimados), bem como a função de custo  $c(\cdot, \theta_1)$ . Nas linhas 12–19, a matriz de transição é formada.

Nas linhas 20–26 do Alg. (6), ocorre a contração do ponto fixo de fato. A Eq. (3.29) é implementada nas linhas 23–24. Seja  $\gamma$  a iteração na qual o algoritmo calcula  $EV_\theta$ , partindo do palpite inicial de que  $EV_\theta^{\gamma=1} = 0$ , assim definido na linha 8, Os valores  $EV_\theta(x)$ ,  $x = 0, 2, \dots, 89$  são calculados até que a máxima diferença  $EV(x)_\theta^\gamma - EV(x)_\theta^{\gamma-1}$  convirja para zero.

No Alg. (7), é implementado o problema descrito pela Eq. (3.13). Note que, no método *nloglikeobs*, que retorna a função de verossimilhança na Eq. (3.13), o ponto fixo é calculado para cada atualização do palpite inicial sobre os parâmetros (linha 2). Nas linhas 3–4, é implementada a Eq. (3.30) e nas linhas 5–10 é criada uma coluna adicional no vetor de probabilidades com o complementar da primeira coluna, que representa a probabilidade de haver reposição no estado da linha da matriz.

---

**Alg. 7:** Composição da função de verossimilhança dos dados relativos à decisão do agente e otimização sujeita à restrição de equilíbrio dada pela contração do ponto fixo  $EV_{\theta}$ .

---

```

1  def nloglikeobs(self, params):
2      fixed_point = self.dynamic_est(params)
3      p_choice_vector = 1/(1+np.exp(self.c-self.bet*fixed_point-self.tr-
4                               self.c[0,0]+self.bet*fixed_point[0,0]))
5      p_choice_vector = np.hstack((p_choice_vector,
6                                  np.zeros((p_choice_vector.shape[0], 1),
7                                             dtype=p_choice_vector.dtype)))
8
9      for i in range(0,90):
10         p_choice_vector[i,1] = 1-p_choice_vector[i,0]
11
12         log_func = 0
13         for i in range(0,self.endog.shape[0]):
14             log_func = log_func - np.log(p_choice_vector[self.exog[i]-1,self.endog[i]])
15         return log_func
16
17  def fit(self, start_params=[3,100,50], maxiter=10000, maxfun=5000):
18      return super(theta_3_estimation, self).fit(start_params=start_params,
19                                                maxiter=maxiter, maxfun=maxfun)

```

---

Por fim, nas linhas 12–15 a função de máxima verossimilhança é composta e maximizada pelo método *fit*, proveniente da classe mais geral *GenericLikelihoodModel*. Optou-se por essa distribuição de funções em virtude de ela ser flexível e permitir maximizar a função de máxima verossimilhança dos dados (que deve ser retornada pela função *nloglikeobs*) através de uma série de métodos, permitindo identificar aquele com melhor performance. Por padrão, o método utilizado é o Nelder-Mead, que foi o que teve melhor desempenho para os dados.

Trata-se de um método de otimização sem restrições que não usa as derivadas da função objetivo. O método é fundamentado num processo de busca iterativa pelo mínimo local da função objetivo. De acordo com Mathews e Fink (2004, p. 430) o método é efetivo e computacionalmente compacto.



## Capítulo 4

# Resultados do algoritmo e análise dos resultados

Nesta seção são apresentados os *outputs* do algoritmo implementado na Seção 3.3, bem como são analisadas as mudanças sobre os parâmetros estruturais quando consideradas algumas políticas contrafactuais incidentes sobre a especificação das primitivas do modelo. Seguindo a esta etapa, é estimada a demanda do agente pelo bem de capital em análise.

Na Tab. (4.1) está disposta a variação das estimativas dos parâmetros de acordo com a especificação da função de utilidade do agente, que depende, essencialmente da forma da função de custo  $c(x, \theta_1)$ . As estimativas da Tab. (4.1) foram obtidas quando considerado um fator de desconto intertemporal de  $\beta = 0,9$ , que caracteriza um agente que valoriza bastante a sua utilidade em períodos futuros. Por outro lado, na Tab. (4.2) são apresentadas as estimativas obtidas dos parâmetros para cada especificação da função de custo quando o fator de desconto intertemporal é 0,1.

Todas as estimativas assinaladas com asterisco são estatisticamente diferentes de zero ao nível de significância de 1%. Note, primeiramente, que as estimativas das componentes do vetor  $\theta_3$ , são as mesmas para quaisquer primitivas do problema, uma vez que a estimação das probabilidades de transição entre os estados é feita na primeira etapa do procedimento, que é executada de forma independente da segunda.

Akaike (1974) sugere que entre dois modelos cujos parâmetros são estimados pelo método da máxima verossimilhança, será considerado mais adequado aos dados aquele que apresentar menor estatística AIC (*Akaike Information Criterion*). Esse também é o critério para avaliar a qualidade de dois modelos pela estatística BIC (*Bayesian Information Criterion*). Nessas condições, as Tab. (4.1) e (4.2) sugerem que, para  $\beta = 0,9$ ,  $c_4$  se adequa melhor aos dados e que para  $\beta = 0,1$ ,  $c_3$  se adequa melhor aos dados.

Note, no entanto, que todos os modelos apresentam graus satisfatórios de qualidade, com todos os parâmetros significantes ao nível de 1%. Isso mostra que agentes ou economias diferentes

$c(x, \theta_1)$	$\theta_{30}$	$\theta_{31}$	$RC$	$\theta_{11}$	$\theta_{12}$	LV ( $l^2$ )	AIC ( $l^2$ )	BIC ( $l^2$ )
$c_1(x, \theta_1) = \theta_{11}x$	0,8870* (0,003)	0,1121* (0,003)	6,5598* (0,309)	0,0051* (0,001)	–	-324,60	651,2	658,2
$c_2(x, \theta_1) = \theta_{11}\sqrt{x}$	0,8870* (0,003)	0,1121* (0,003)	8,3715* (0,560)	0,0704* (0,009)	–	-317,31	636,6	643,6
$c_3(x, \theta_1) = \theta_{11}x + \theta_{12}\sqrt{x}$	0,8870* (0,003)	0,1121* (0,003)	3,0284* (0,232)	0,0174* (0,002)	-0,1587* (0,020)	-367,61	737,2	744,2
$c_4(x, \theta_1) = x^{\theta_{11}}$	0,8870* (0,003)	0,1121* (0,003)	15,3323* (0,773)	0,1352* (0,013)	–	-313,13	628,3	635,3

Tabela 4.1: Estimativas obtidas para os parâmetros do modelo de acordo com a especificação da utilidade do agente, para um fator de desconto  $\beta = 0,9$  (erros-padrão entre parênteses).

$c(x, \theta_1)$	$\theta_{30}$	$\theta_{31}$	$RC$	$\theta_{11}$	$\theta_{12}$	LV ( $l^2$ )	AIC ( $l^2$ )	BIC ( $l^2$ )
$c(x, \theta_1) = \theta_{11}x$	0,8870* (0,003)	0,1121* (0,003)	6,2173* (0,255)	0,0348* (0,004)	–	-328,14	658,3	665,3
$c(x, \theta_1) = \theta_{11}\sqrt{x}$	0,8870* (0,003)	0,1121* (0,003)	7,9944* (0,476)	0,4952* (0,062)	–	-320,02	642,0	649,1
$c(x, \theta_1) = \theta_{11}x + \theta_{12}\sqrt{x}$	0,8870* (0,003)	0,1121* (0,003)	24,8127* (4,162)	-0,3511* (0,082)	5,2056* (0,110)	-299,52	601,0	608,1
$c(x, \theta_1) = x^{\theta_{11}}$	0,8870* (0,003)	0,1121* (0,003)	8,9582* (0,389)	0,3787* (0,021)	–	-318,28	638,6	645,6

Tabela 4.2: Estimativas obtidas para os parâmetros do modelo de acordo com a especificação da utilidade do agente, para um fator de desconto  $\beta = 0,1$  (erros-padrão entre parênteses).

reagem de forma diferente ao ambiente no qual estão inseridos, o que pode ser capturado por um modelo estrutural.

Nas Fig. (4.1) e (4.2) são expressas as probabilidades de o agente efetuar a reposição do motor de um ônibus dada a quilometragem acumulada do veículo em um dado momento. Foram traçados quatro curvas distintas, cada uma referente a cada especificação da função custo do agente.

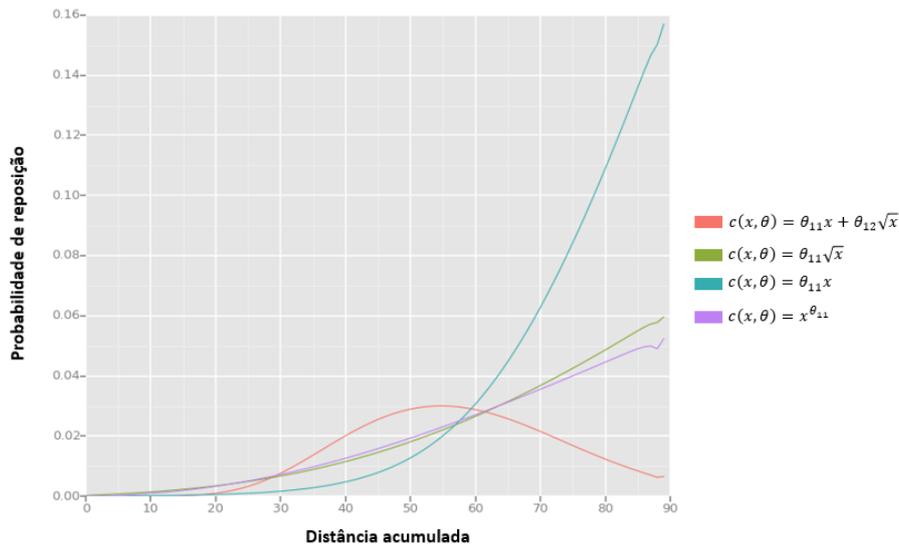


Figura 4.1: Probabilidade de reposição do motor do ônibus condicional à distância acumulada do veículo, discretizada em 90 intervalos e tomando  $\beta = 0,9$ .

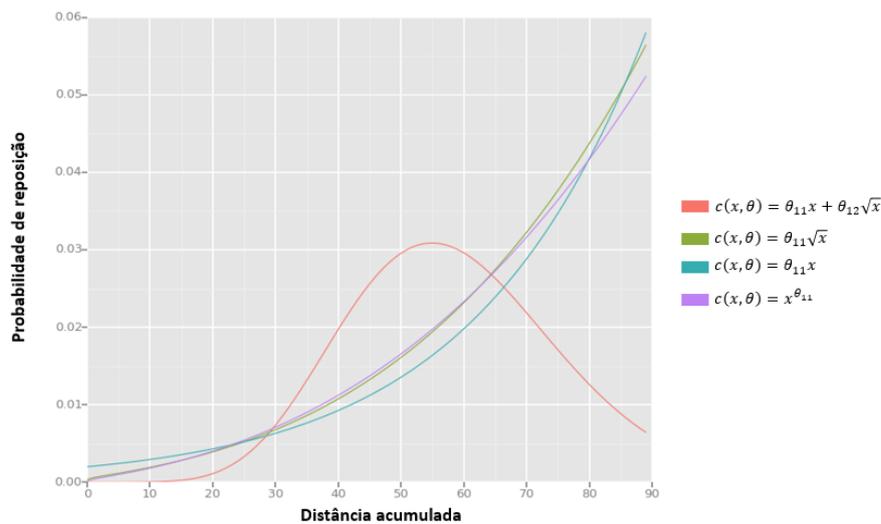


Figura 4.2: Probabilidade de reposição do motor do ônibus condicional à distância acumulada do veículo, discretizada em 90 intervalos e tomando  $\beta = 0,1$ .

Uma vez estimados os parâmetros estruturais, pode-se calcular as probabilidades condi-

onais de o agente trocar o motor do ônibus, bem como as probabilidades de transição entre os estados. Nessas condições, pode-se estimar a demanda anual por bens de capital do agente.

Sob a hipótese de que a transição dos estados pode ser descrita por um processo markoviano regenerativo, a probabilidade de o agente fazer a reposição no tempo  $t$  em um único ônibus é denotada por  $\gamma_t$ , é dada por:

$$\gamma_t = \sum_{j=1}^t (T_0^j \cdot P)^T \cdot \mathbf{1}/\mathbf{90} \quad (4.1)$$

Aparentemente, a Eq. (4.1) pode parecer um pouco confusa, mas não é de fato. Recorde que  $T_0$  é a matriz estocástica do processo, que é quadrada e de ordem 90. Nessas condições, sabe-se não somente que a potenciação de  $T_0$  está bem definida como também  $T_0^t, t = 1, 2, \dots$  é uma matriz de transição cujo elemento  $a_{ij}$  denota a probabilidade de, no tempo  $t$ , o agente estar no estado  $j$ , dado o estado inicial  $i$ . Mais ainda, defina  $P \in \mathbb{R}^{90}$  como:

$$P := \begin{bmatrix} P(i_t = 1|x_t = 0) \\ P(i_t = 1|x_t = 1) \\ \vdots \\ P(i_t = 1|x_t = 89) \end{bmatrix} \quad (4.2)$$

e tome  $\mathbf{1}/\mathbf{90} \in \mathbb{R}^{90}$  como sendo o vetor cujas entradas são todas  $\frac{1}{90}$ , que define probabilidades iguais para o estado inicial. Assim, tem-se que o produto interno na Eq. (4.1) está bem definido também e que expressa a probabilidade de haver reposição no tempo  $t$ , quando o estado inicial pode ser qualquer um dos possíveis com chance igual. Note que, em virtude da característica regenerativa do processo markoviano, a evolução entre dois pontos de regeneração depende apenas do estado em que houve regeneração e não da história do processo antes de o processo ter sido regenerado.

Segue, por conseguinte, que a demanda anual por motor do agente para um único ônibus,  $D_{12}(RC)$ , é dada por:

$$D_{12}(RC) = RC \sum_{t=1}^{12} \gamma_t = RC \sum_{t=1}^{12} \sum_{j=1}^t (T_0^j \cdot P)^T \cdot \mathbf{1}/\mathbf{90} \quad (4.3)$$

Vale destacar o efeito de  $RC$  sobre  $D_{12}(RC)$  não é linear, como pode parecer pela Eq. (4.3). Recorde que as probabilidades de reposição que compõe  $P$  também são calculadas a partir do valor de  $RC$ . Assim, pode ser traçado como que varia o gasto esperado anual com reposição de motores de ônibus à medida que varia o custo unitário desta reposição.

Agora, tomaremos dois casos como exemplo para estimar a demanda por bens de capital: função de custo  $c_2$  com  $\beta = 0,1$  e com  $\beta = 0,9$ . Nas Fig. (4.3) e (4.4) é analisada a variação da demanda anual esperada por motor de ônibus para um único veículo com respeito ao subsídio

ou imposto estabelecido pelo governo sobre o custo de reposição. Vale destacar que o ponto  $x = 0$  representa o caso em que o governo subsidia 100% do custo de reposição e o ponto final é quando incide um imposto de 50%.

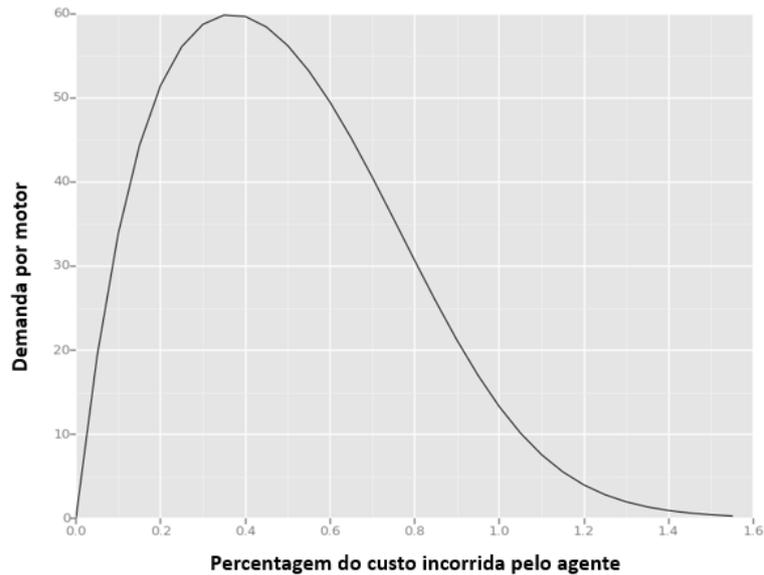


Figura 4.3: Demanda do agente por motor à medida em que varia a participação do governo no custo de reposição ( $\beta = 0,9$ ).

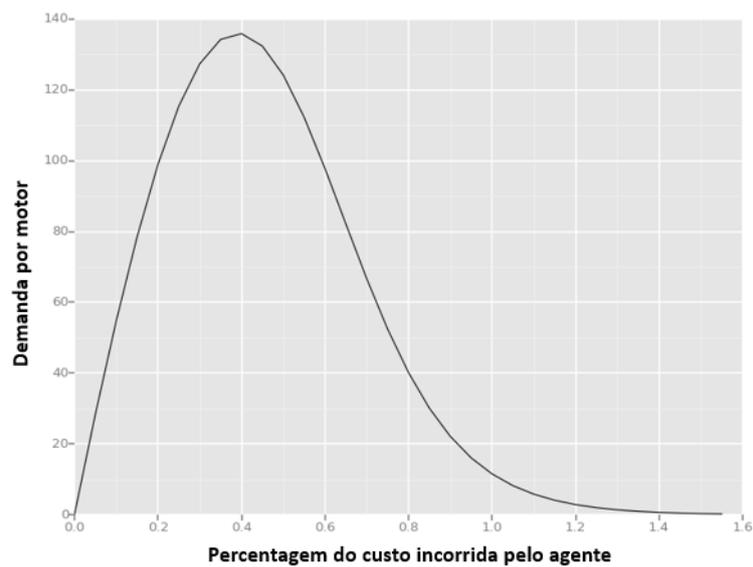


Figura 4.4: Demanda do agente por motor à medida em que varia a participação do governo no custo de reposição ( $\beta = 0,1$ ).

Este é um exemplo de como, após estimado o modelo, pode-se simular as consequências econômicas de uma intervenção externa sobre o ambiente no qual o agente está inserido, dadas as suas preferências, que caracterizam sua função de utilidade.



# Capítulo 5

## Conclusão

Este trabalho tem o intuito de introduzir de forma compreensiva os fundamentos, as ferramentas e os métodos básicos empregados na modelagem estrutural na Economia para novos praticantes ou interessados.

Dessa maneira, buscou-se, inicialmente, situar o leitor no surgimento da modelagem estrutural, no que foi feito até agora na área e no que tem sido feito atualmente, destacando a relevância dessa temática para Economia, suas principais vantagens, como possibilidade de conduzir políticas contrafactuais, e os principais entraves para a incorporação de estrutura aos modelos empíricos, ligados, mormente, às dificuldades computacionais.

O trabalho também discute uma forma clássica de modelagem estrutural baseada em modelos DDC e implementa detalhadamente as soluções do modelo em Python. Uma vez que o código é apresentado, motiva-se o desenvolvimento de algoritmos semelhantes numa linguagem moderna e descomplicada, de tal forma que o pesquisador possa modificar os parâmetros de acordo com seu problema. Nestas condições, podem ser alterados facilmente o espaço no qual a variável de estado é discretizada, a distribuição trinomial sobre os estados ou até mesmo a forma de estimação (paramétrica ou não-paramétrica) da distribuição sobre as transições de estado.

De modo geral, foi visto que a modelagem estrutural capta aspectos relacionados às preferências dos agentes e é mais indicada para captar comportamentos de longo prazo. É, sobretudo, versátil, do ponto de vista que permite ao pesquisador captar mudanças de variáveis sobre outras de acordo com as primitivas do modelo, que são especificações sobre o comportamento esperado dos indivíduos, como por exemplo valorização dos *payoffs* futuros e a forma funcional da função *payoff*.

Trabalhos futuros podem utilizar modelos estruturais para outros problemas relevantes, como na modelagem da escolha de consumidores por determinados bens, incluindo maior número de agentes, de variáveis de decisão ou uma variável de decisão não binária, por exemplo.



# Apêndice A

## Demonstração do Teorema do ponto fixo de Banach

Será provada, primeiramente, a existência do ponto fixo. Tome  $x_0 \in S$  e uma sequência  $\{x_n\}_1^\infty$  em que  $x_n = T(x_{n-1})$ . Queremos mostrar que a sequência  $\{x_n\}_1^\infty$  é de Cauchy. Como  $T$  é uma contração:

$$d(x_1, x_2) = d(T(x_0), T(x_1)) \leq \alpha d(x_0, x_1) \quad (\text{A.1})$$

$$d(x_2, x_3) = d(T(x_1), T(x_2)) \leq \alpha d(x_1, x_2) \leq \alpha^2 d(x_0, x_1) \quad (\text{A.2})$$

Indutivamente, temos que:

$$d(x_n, x_{n+1}) \leq \alpha^n d(x_0, x_1) \quad (\text{A.3})$$

Para  $p$  qualquer, pela desigualdade triangular, tem-se que:

$$d(x_n, x_{n+p}) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \dots + d(x_{n+p-1}, x_{n+p}) \quad (\text{A.4})$$

Note que, por (A.3), escreve-se:

$$\begin{aligned} d(x_n, x_{n+1}) &\leq \alpha^n d(x_0, x_1) \\ d(x_{n+1}, x_{n+2}) &\leq \alpha^{n+1} d(x_0, x_1) \\ &\vdots \\ d(x_{n+p-1}, x_{n+p}) &\leq \alpha^{n+p-1} d(x_0, x_1) \end{aligned} \quad (\text{A.5})$$

Daí, segue, portanto, que:

$$d(x_n, x_{n+p}) \leq (\alpha^n + \alpha^{n+1} + \dots + \alpha^{n+p-1}) d(x_0, x_1) \quad (\text{A.6})$$

Aplicando  $\lim_{n \rightarrow \infty}(\cdot)$  sobre (A.6), é obtido que:

$$\lim_{n \rightarrow \infty} d(x_n, x_{n+p}) \leq d(x_0, x_1) \lim_{n \rightarrow \infty} \sum_{j=n}^{n+p-1} \alpha^j \quad (\text{A.7})$$

Defina  $s_{p,n}$  como a soma parcial numa sequência dos  $p$  elementos seguintes a  $x_n, n \in \mathbb{N}$ , de tal forma que  $s_{p,n} = \alpha^n + \alpha^{n+1} + \dots + \alpha^{n+p-1}$ . Faz-se, por conseguinte:

$$\begin{aligned} \alpha s_{p,n} &= \alpha^{n+1} + \alpha^{n+2} + \dots + \alpha^{n+p} \\ s_{p,n} - \alpha s_{p,n} &= \alpha^n - \alpha^{n+p} \\ s_{p,n}(1 - \alpha) &= \alpha^n(1 - \alpha^p) \\ s_{p,n} &= \frac{\alpha^n(1 - \alpha^p)}{(1 - \alpha)} \end{aligned} \quad (\text{A.8})$$

Note que  $\lim_{n \rightarrow \infty}(s_{p,n}) = 0$ , uma vez que  $\alpha \in [0, 1)$ . Isso implica que pode-se escrever (A.7) como:

$$\lim_{n \rightarrow \infty} d(x_n, x_{n+p}) \leq 0 \quad (\text{A.9})$$

Como a métrica  $d$  é uma função positiva, segue que  $\lim_{n \rightarrow \infty} d(x_n, x_{n+p}) = 0$ . Isso significa que  $\{x_n\}_1^\infty$  é uma sequência de Cauchy. Como  $S$  é um espaço completo, a sequência converge em  $S$ , i.e  $\lim_{n \rightarrow \infty} x_n = c$ . Finalmente, aplicando o limite sobre  $x_n = T(x_{n-1})$  e usando a continuidade de  $T$ , temos:

$$c = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} T(x_{n-1}) = T\left(\lim_{n \rightarrow \infty} x_n\right) = T(c) \Rightarrow T(c) = c \quad (\text{A.10})$$

Provada a existência do ponto fixo, verificar que ele é único é direto. Sejam  $z$  e  $b$  dois pontos fixos de  $T$ . Segue que,

$$d(z, b) = d(T(z), T(b)) \leq \alpha d(z, b) \Rightarrow (1 - \alpha)d(z, b) \leq 0 \quad (\text{A.11})$$

Como  $0 \leq \alpha < 1$ , tem-se que  $(1 - \alpha) > 0$ . Mais ainda, se  $d$  é uma aplicação positiva, segue, de (A.11), que  $d(z, b) = 0 \Rightarrow z = b$ .

## Apêndice B

# Derivação da função logit dinâmica

Nesta Seção será mostrado como obter a Eq. (3.9) a partir da Eq. (3.8). Recorde que é assumido que  $\varepsilon_i$  provém de uma distribuição de valor extremo do tipo I ou Gumbel, em que  $i = 0, 1$  denota a decisão tomada pelo agente no tempo  $t$  (omitido para simplificar a notação). Assim, tem-se que a função densidade de  $\varepsilon_i$  é do tipo:

$$f(\varepsilon_i) = e^{-\varepsilon_i} \cdot e^{-e^{-\varepsilon_i}} \quad (\text{B.1})$$

e que a função de distribuição acumulada é dada por:

$$F(\varepsilon_i) = e^{-e^{-\varepsilon_i}} \quad (\text{B.2})$$

Nessas condições, podemos reescrever a Eq. (3.8) em termos da distribuição acumulada de  $\varepsilon_i$  até o ponto  $\varepsilon_1 + u_1 - u_0$ . Uma vez que é assumido que os  $\varepsilon_i$  são independentes, a probabilidade descrita na Eq. (3.8) para um dado  $\varepsilon_1$  pode ser escrita como sendo simplesmente a função de distribuição acumulada  $F(\varepsilon_0)$ . Mas para reescrever Eq. (3.8), precisamos considerar que  $\varepsilon_1$  não é dado, de tal forma que precisa-se incluir a integral sobre sua função de densidade. Nessas condições:

$$P(i_t | x_t, \theta) = \int_{-\infty}^{\infty} e^{-e^{-(\varepsilon_1 + u_1 - u_0)}} \cdot e^{-\varepsilon_1} \cdot e^{-e^{-\varepsilon_1}} d\varepsilon_1 \quad (\text{B.3})$$

Fazendo a multiplicação e a divisão do integrando da Eq. (B.3) por  $e^{-e^{-(\varepsilon_1 + u_1 - u_0)}}$ , pelas propriedades da função exponencial e agregando os expoentes, obtemos que:

$$P(i_t | x_t, \theta) = \int_{-\infty}^{\infty} e^{-\sum_{j=0}^1 e^{-(\varepsilon_1 + u_1 - u_j)}} \cdot e^{-\varepsilon_1} d\varepsilon_1 \quad (\text{B.4})$$

Com o artifício da substituição de variáveis, defina  $k := e^{-\varepsilon_1}$ , de tal forma que  $dk = -e^{-\varepsilon} d\varepsilon$ . Note que quando  $\varepsilon_1$  tende ao infinito positivo,  $k$  tende a 0 e que quando  $\varepsilon_1$  tende ao infinito negativo,  $k$  tende ao infinito positivo. Assim, reescrevemos a Eq. (B.4) como sendo:

$$P(i_t|x_t, \theta) = - \int_{\infty}^0 e^{-k \sum_{j=0}^1 e^{-(u_1-u_j)}} dk = \int_0^{\infty} e^{-k \sum_{j=0}^1 e^{-(u_1-u_j)}} dk \quad (\text{B.5})$$

Para simplificar, defina  $C := \sum_{j=0}^1 e^{-(u_1-u_j)}$ , de tal forma que  $C$  é constante para todo valor de  $k$ . Com esse artifício, pode-se resolver a integral na Eq. (B.5) de forma simples. Teremos então, que:

$$\begin{aligned} P(i_t|x_t, \theta) &= \int_0^{\infty} e^{-Ck} dk \\ &= - \frac{e^{Ck}}{C} \Big|_0^{\infty} \\ &= \frac{1}{C} = \frac{1}{\sum_{j=0}^1 e^{-(u_1-u_j)}} \\ &= \frac{1}{e^{-u_1} \sum_{j=0}^1 e^{u_j}} \\ &= \frac{e^{u_1}}{\sum_{j=0}^1 e^{u_j}} \end{aligned} \quad (\text{B.6})$$

# Referências Bibliográficas

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19 (6), p. 716–723, Dec 1974.
- ARCIDIACONO, P.; MILLER, R. A. *Identifying dynamic discrete choice models off short panels*. 2013.
- BALKE, A.; PEARL, J. *Counterfactuals and policy analysis in structural models*. 2013.
- CHIONG, K. X.; GALICHON, A.; SHUM, M. Duality in dynamic discrete-choice models. *Quantitative Economics*, Wiley Online Library, v. 7, n. 1, p. 83–115, 2016.
- COHEN, A.; SIEGELMAN, P. *Testing for Adverse Selection in Insurance Markets*. [S.l.], 2009.
- FOSS, M. O.; HOST, A. *Hydroelectric Real Options: A Structural Estimation Approach*. Dissertação (Mestrado) — Norwegian University of Science and Technology, 2011.
- FRIEDMAN, M. *A Theory of the Consumption Function*. [S.l.]: Princeton University Press, 1957.
- FRISCH, R. "editorial". *Econometrica*, 1, p. 2, 1933.
- HAUSMAN, J. A. Specification and estimation of simultaneous equation models. In: \_\_\_\_\_. [S.l.]: Elsevier BV, 1983. (Handbook of Econometrics, v. 1), cap. 7, p. 391–448.
- JEZIORSKI, P.; KRASNOKUTSKAYA, E.; CECCARINI, O. Adverse selection and moral hazard in a dynamic model of auto insurance. In: . [S.l.: s.n.], 2017.
- JUDD, K.; SU, C.-L. A new optimization approach to maximum likelihood estimation of structural models. *Computing in Economics and Finance 2006*, 2006.
- KALOUPTSIDI, M.; SCOTT, P. T.; SOUZA-RODRIGUES, E. On the non-identification of counterfactuals in dynamic discrete games. *International Journal of Industrial Organization*, Elsevier BV, v. 50, p. 362–371, jan 2017.
- KEANE, M. P. Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, Elsevier BV, v. 156, n. 1, p. 3–20, May 2010.
- KEANE, M. P.; WOLPIN, K. I. The career decisions of young men. *Journal of political Economy*, The University of Chicago Press, v. 105, n. 3, p. 473–522, 1997.
- LJUNGQVIST, L.; SARGENT, T. J. *Recursive Macroeconomic Theory*. [S.l.]: MIT Press. Cambridge, MA, 2000.

- LUCAS, R. E. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, Elsevier BV, v. 1, p. 19–46, Jan 1976.
- MATHEWS, J. H.; FINK, K. K. *Numerical Methods Using Matlab*. 4. ed. [S.l.]: Pearson, 2004.
- NEVO, A.; WHINSTON, M. D. Taking the dogma out of econometrics: Structural modeling and credible inference. *Journal of Economic Perspectives*, American Economic Association, v. 24, n. 2, p. 69–82, May 2010.
- NORETS, A.; TANG, X. Semiparametric inference in dynamic binary choice models. *The Review of Economic Studies*, Oxford University Press (OUP), v. 81, n. 3, p. 1229–1262, Dec 2013.
- PEARL, J. Structural counterfactuals: A brief introduction. *Cognitive Science*, Wiley-Blackwell, v. 37, n. 6, p. 977–985, Aug 2013.
- REISS, P. C.; WOLAK, F. A. Chapter 64 Structural econometric modeling: Rationales and examples from industrial organization. In: *Handbook of Econometrics*. [S.l.]: Elsevier BV, 2007. p. 4277–4415.
- RUST, J. Optimal replacement of GMC bus engines: An empirical model of harold zurcher. *Econometrica*, JSTOR, v. 55, n. 5, p. 999, Sep 1987.
- RUST, J. Chapter 51 Structural estimation of markov decision processes. In: *Handbook of Econometrics*. [S.l.]: Elsevier BV, 1994. p. 3081–3143.
- RUST, J.; PHELAN, C. How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 781–831, 1997.
- SU, C.-L.; JUDD, K. L. Constrained optimization approaches to estimation of structural models. *Econometrica*, The Econometric Society, v. 80, n. 5, p. 2213–2230, 2012.
- SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning I: Introduction*. 1998.
- TRAIN, K. *Discrete Choice Methods with Simulation*. 1. ed. [S.l.]: Cambridge University Press, 2002.
- WOLPIN, K. I. *The Limits of Inference without Theory*. [S.l.]: MIT Press, 2013.
- YOUNG, A. *The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience*. [S.l.], 1994.