

UNIVERSIDADE FEDERAL DO PARANÁ

HECTOR DE MOURA LUZ

**TRÊS ENSAIOS SOBRE CONTRIBUIÇÕES ECONÔMICAS PARA A ESCOLHA
DE MODAL**

CURITIBA

2017

HECTOR DE MOURA LUZ

**TRÊS ENSAIOS SOBRE CONTRIBUIÇÕES ECONÔMICAS PARA A ESCOLHA
DE MODAL**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Desenvolvimento Econômico, no Programa de Pós-Graduação em Desenvolvimento Econômico, Setor de Ciências Sociais Aplicadas, da Universidade Federal do Paraná.

Orientadora: Prof^a. Dra. Adriana Sbicca Fernandes

CURITIBA

2017

UNIVERSIDADE FEDERAL DO PARANÁ. SISTEMA DE BIBLIOTECAS.
CATALOGAÇÃO NA FONTE

Luz, Hector de Moura

Três ensaios sobre contribuições econômicas para a escolha de modal /
Hector de Moura Luz. – 2017.

85 f.

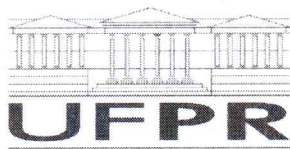
Orientadora: Adriana Sbicca Fernandes.

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de
Ciências Sociais Aplicadas, Programa de Pós-Graduação em
Desenvolvimento Econômico.

Defesa: Curitiba, 2017.

1. Transporte – Aspectos econômicos. 2. Transporte – Políticas públicas.
3. Mobilidade urbana. 4. Economia comportamental. I. Fernandes, Adriana
Sbicca, 1969- II. Universidade Federal do Paraná. Setor de Ciências Sociais
Aplicadas. Programa de Pós-Graduação em Desenvolvimento Econômico.
III. Título.

CDD 388.11



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em DESENVOLVIMENTO ECONÔMICO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **HECTOR DE MOURA LUZ** intitulada: **Contribuições Econômicas para a Escolha de Modal**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO.

Curitiba, 30 de Março de 2017.

ADRIANA SBICCA FERNANDES
Presidente da Banca Examinadora (UFPR)

ALEXANDRE ALVES PORSSE
Avaliador Interno (UFPR)

HERMES HIGACHI
Avaliador Externo (UEPG)

A minha mãe, Roselene,
que acompanhou
todo o processo deste trabalho.

AGRADECIMENTOS

Agradeço primeiramente a meus pais. A minha mãe, pelo exemplo de determinação e persistência que, aliados a sua constante serenidade, me impulsionaram no decorrer deste trabalho. A meu pai, por todo o apoio que desempenhou, facilitando muito a execução deste projeto.

É profundo meu agradecimento aos meus irmãos, Hugo e Santiago, pelas mais diversificadas e diferenciadas formas de apoio, demonstrando e reforçando nossa verdadeira irmandade. Gratidão esta que, evidentemente, se estende a Eliane e Juliana.

Dentre todos os amigos feitos no PPGDE, devo destacar Daniel e Ângelo, companheiros de orientação. Ao Daniel, sou grato pelas conversas, desabafos, discussões, almoços e essencial abrigo. Ao Ângelo, por tantas caronas, tantas conversas e tanta boa música.

Caronas também que me fazem agradecer ao Davi e ao Matheus. Aquele, que me poupou do frio matinal curitibano, e este, que tantas vezes me levou de volta a minha terra natal com tão boas conversas.

É impossível deixar de agradecer ao Mailliw, companheiro desta aventura curitibana e de tantas caminhadas. Que possamos seguir nessa frutífera troca de conhecimentos.

Agradeço ao Marcus e a Paula, pelo abrigo e atenção tão importantes ao longe desse mestrado.

À minha Professora e Orientadora Adriana Sbicca, minha profunda gratidão pela paciência e compreensão nesse período tão conturbado. A força e a orientação nos momentos precisos foram essenciais para a conclusão deste projeto.

Também agradeço ao Professor Basílio, que tantas ideias deu para o desdobramento deste projeto e possibilitou a arrancada inicial com os cursos de Economia e Complexidade.

Devo agradecer ainda ao Professor Alexandre Porsse, por tantas boas “pequenas” orientações, que aprimoraram o resultado final.

Não posso deixar de agradecer à Professora Raquel que, além da excelência acadêmica, auxilia os pós-graduandos com verdadeira e honesta atenção.

Agradeço, de modo geral a fim de não alongar tanto esta seção, a todos aqueles pertencentes ao PPGDE, tanto alunos quanto professores e funcionários. Sem a presença de cada um de vocês esse trabalho não seria o mesmo.

Por fim, é especial minha gratidão a Camila que, apesar de todas as dificuldades que nos são apresentadas, segue minha maravilhosa companheira.

“Eu quis prever o futuro, consertar o passado
Calculando os riscos
Bem devagar, ponderado
Perfeitamente equilibrado

Até que num dia qualquer
Eu vi que alguma coisa mudara
Trocaram os nomes das ruas
E as pessoas tinham outras caras”

Herbert Vianna, 1996.

RESUMO

A mobilidade urbana, assim como o trânsito a ela relacionado, apresentam grandes impactos na qualidade de vida da população, tanto diretamente através do gasto com tempo em deslocamento, como pela decorrente perda de produtividade. Devido a tais impactos, políticas públicas a fim de promover melhor fluidez ao trânsito e mobilidade aos habitantes aparecem como essenciais. Para melhor avaliação de tais políticas, a compreensão e previsão de escolha de modal de cada indivíduo são importantes. Os modelos de escolha discreta são as principais ferramentas utilizadas para tal fim, mas apresentam algumas limitações e/ou dificuldades. O presente trabalho busca contribuir com soluções para algumas delas. Primeiramente, explora-se a utilização do Google Maps como alternativa para estimação dos valores dos atributos das alternativas não escolhidas. Tal abordagem, proposta para enfrentar uma dificuldade encontrada na aplicação de modelos de escolha discreta a pesquisas de preferência revelada, é comparada à estimação através de regressão linear, opção recorrentemente utilizada na literatura brasileira. Esta comparação é realizada com a aplicação de modelos de escolha discreta utilizando a base de dados da Pesquisa Origem e Destino de 2007 da Região Metropolitana de São Paulo. Outra contribuição apresentada pelo presente trabalho refere-se diretamente à especificação dos modelos de escolha discreta. Sendo baseados em pressupostos de maximização de utilidade tradicional, a incorporação de proposições relacionadas à Economia Comportamental pode contribuir para a compreensão da escolha. Desta forma, o segundo artigo que compõe este trabalho realiza uma análise empírica de um modelo de escolha discreta por modal de transporte na Região Metropolitana de São Paulo em viagens a trabalho, considerando regras compensatórias e não compensatórias. Para tanto, utiliza-se da modelagem de funções de densidade de probabilidade mistas desenvolvida por Swait (2009), aplicando o Modelo 2-Mix com Condição de Rejeição a dados da Pesquisa Origem e Destino recentes (2007). As duas primeiras contribuições propostas apresentaram bons resultados, sugerindo a utilização e demonstrando potencial de aplicação conjunta. Por fim, o terceiro e último artigo despende enfoque especial no caráter estático dos modelos de escolha discreta, quando da avaliação de políticas de transporte. Para tanto, desenvolve modelos baseados em agentes simplificados, discutindo questões também relacionadas à formação de expectativas. Os resultados obtidos permitem a observação de dinâmica coevolutiva e propriedades emergentes. Conclui-se sobre a importância desta modelagem para explicação do fenômeno e elucidação de sua dinâmica, evidenciando as possibilidades de complementaridade entre os ferramentais.

Palavras-chave: Transporte. Modelos de escolha discreta. Políticas públicas. Pesquisa Origem e Destino. Economia Comportamental. Modelo Baseado em Agentes.

ABSTRACT

Urban mobility and its related traffic have major impacts on the quality of life of the population. They impact both directly, through the expense of time on the road, and indirectly, due to the loss of productivity. Due to such impacts, public policies are essential to promote better traffic flow and mobility to the inhabitants. The understanding and prediction of modal choice of each individual are important to make better evaluation of such policies. Discrete choice models are the main tools used for this purpose, but they present some limitations and/or difficulties. The present work seeks to propose solutions for some of them. Firstly, the use of Google Maps is explored as an alternative to estimate the values of the attributes of the non-chosen alternatives. This approach is proposed to face a difficulty found in the application of models of discrete choice to revealed preference surveys and it is compared to the estimation through linear regression, an option that is frequently used in the Brazilian literature. This comparison is carried out with the application of discrete choice models using the 2007 Origin and Destination Survey database of the Metropolitan Region of São Paulo. Another contribution presented by the present study refers directly to the specification of discrete choice models. Since these models are based on traditional assumptions of utility maximization, the incorporation of propositions related to the Behavioral Economics can contribute to the understanding of the choice. In this way, the second article that compose this work performs an empirical analysis of a model of discrete choice by mode of transportation in the Metropolitan Region of São Paulo in work trips, considering compensatory and non-compensatory rules. For this purpose, it is used the model of mixed probability density functions developed by Swait (2009). It is applied the 2-Mix Model with Rejection Condition to data from the recent Origin and Destination Survey (2007). The first two proposed contributions presented good results, suggesting their use and showing potential of joint application. Finally, the third and final article focuses on the static nature of discrete choice models when evaluating transport policies. To do so, it develops models based on simplified agents, discussing issues also related to the formation of expectations. The results obtained allow the observation of coevolutionary dynamics and emergent properties. It is concluded on the importance of this modeling to explain the phenomenon and elucidation of its dynamics, evidencing the possibilities of complementarity between the tools.

Key-words: Transport. Discrete choice models. Public policies. Origin and Destination survey. Behavioural Economics. Agent Based Models.

LISTA DE FIGURAS

FIGURA 1 – AMBIENTE E CONDIÇÃO ESPACIAL INICIAL.....	58
FIGURA 2 – EMERGÊNCIA DE CONGESTIONAMENTO.	60

LISTA DE GRÁFICOS

GRÁFICO 1 – EQUILÍBRIO ESTÁVEL	66
GRÁFICO 2 – EQUILÍBRIO INSTÁVEL	67
GRÁFICO 3 – EQUILÍBRIO DINÂMICO.....	70
GRÁFICO 4 – EQUILÍBRIO DINÂMICO: CICLOS SEMELHANTES A BOLHAS	72
GRÁFICO 5 – UTILIDADE MÉDIA DOS INDIVÍDUOS POR MODAL – CICLOS DE QUEDA BRUSCA DA UTILIDADE DOS USUÁRIOS DE CARRO.....	73
GRÁFICO 6 – EQUILÍBRIO DINÂMICO E POSSÍVEL SUPERESTIMAÇÃO	76

LISTA DE TABELAS

TABELA 1 – EXEMPLO ILUSTRATIVO DA DIFICULDADE DE COMPARAÇÃO POR ANÁLISE DE ERROS	23
TABELA 2 – APANHADO ILUSTRATIVO PARA COMPARAÇÃO ENTRE BASES ESTIMADAS.....	25
TABELA 3 - MÉDIAS DE DURAÇÃO POR SUBAMOSTRA DOS TRÊS MODAIS DE MAIOR UTILIZAÇÃO	27
TABELA 4 - COMPARAÇÃO DAS ESTIMAÇÕES LOGIT MULTINOMIAL.....	29
TABELA 5 - COMPARAÇÃO ENTRE ESTIMAÇÕES LOGIT MISTAS	33
TABELA 6 - COMPARAÇÃO ENTRE ESTIMAÇÕES DOS MODELOS 2-MIX E LOGIT MULTINOMIAL	49
TABELA 7 - PROBABILIDADE DE ESTAR NA CONDIÇÃO DE TRADEOFF POR MODAL DE TRANSPORTE	52
TABELA 8 – PARÂMETROS UTILIZADOS NOS EXPERIMENTOS DO MODELO BÁSICO.....	65

SUMÁRIO

1	INTRODUÇÃO	13
2	ARTIGO 1: USO DO GOOGLE MAPS PARA ESTIMAÇÃO DE VALORES DE ATRIBUTOS NÃO ESCOLHIDOS EM PESQUISAS DE PREFERÊNCIA REVELADA.....	17
2.1	INTRODUÇÃO.....	17
2.2	AUSÊNCIA DOS ATRIBUTOS DAS ALTERNATIVAS NÃO ESCOLHIDAS	19
2.3	GOOGLE MAPS COMO ALTERNATIVA	21
2.4	DIFICULDADES DE COMPARAÇÃO.....	22
2.5	ANÁLISE E COMPARAÇÃO DAS ESTIMAÇÕES.....	24
2.5.1	Comparação entre Bases de Dados Estimadas	25
2.5.2	Estimação dos Modelos de Escolha Discreta	27
2.5.3	Comparação de Resultados dos Modelos de Escolha Discreta	29
2.6	CONCLUSÃO	35
3	ARTIGO 2: MODELAGEM 2-MIX – UMA APLICAÇÃO COM NOÇÕES DA ECONOMIA COMPORTAMENTAL À REGIÃO METROPOLITANA DE SÃO PAULO	37
3.1	INTRODUÇÃO.....	37
3.2	REVISÃO DE LITERATURA	38
3.3	METODOLOGIA	44
3.4	RESULTADOS	48
3.5	CONSIDERAÇÕES FINAIS.....	53
4	ARTIGO 3: DINÂMICA COEVOLUTIVA NA ANÁLISE DE POLÍTICA DE TRANSPORTES	54
4.1	INTRODUÇÃO.....	54
4.2	MODELO SIMPLIFICADO: ADICIONANDO DINÂMICA AO LOGIT MULTINOMIAL.....	56
4.2.1	Logit Multinomial.....	56
4.2.2	Modelo Baseado em Agentes: Configuração Básica.....	57
4.2.3	Emergência	60
4.2.4	Modelo Baseado em Agentes: Escolha	61
4.2.5	Resultados.....	64

4.3	POSSÍVEIS ALTERAÇÕES	68
4.3.1	Alteração 1: incluindo aleatoriedade.....	69
4.3.2	Alteração 2: incluindo heterogeneidade.....	74
4.4	CONCLUSÃO	77
5	CONCLUSÃO.....	78
6	REFERÊNCIAS.....	80

1 INTRODUÇÃO

A mobilidade urbana, assim como o trânsito a ela relacionado, apresentam inegáveis impactos na qualidade de vida da população, principalmente urbana e de regiões metropolitanas. O impacto direto, em decorrência da perda de tempo no deslocamento de ida e volta de casa ao trabalho, é vultoso e ainda crescente em diversas regiões. A Tabela 1 apresenta informações com relação às 6 regiões metropolitanas brasileiras com maior e às 6 com menor tempo médio de deslocamento casa – trabalho – casa. Tal tabela evidencia a grande magnitude de tais impactos na vida dos habitantes de tais localidades. Chama a atenção, de início, o tempo médio enfrentado nas gigantescas regiões metropolitanas, como Rio de Janeiro, 2 horas e 21 minutos, e São Paulo, 2 horas e 12 minutos (FIRJAN, 2015). No entanto, é também relevante o elevado dispêndio de tempo em regiões menores, como Maringá, 1 hora e 45 minutos, e Carbonífera, 1 hora e 43 minutos (FIRJAN, 2015).

TABELA 1 – EVOLUÇÃO DO TEMPO MÉDIO DE DESLOCAMENTO CASA – TRABALHO – CASA EM ÁREAS METROPOLITANAS BRASILEIRAS E POPULAÇÃO OCUPADA AFETADA

Área Metropolitana	Tempo gasto no deslocamento (minutos)			Trabalhadores com deslocamento acima de 30 minutos		
	Minutos		Variação (12/11)	2012	2011	Variação % (12/11)
	2012	2011				
Rio de Janeiro - RJ	141	130	8,4%	2.794.364	2.838.340	-1,5
São Paulo – SP	132	131	1,1%	5.536.241	5.297.456	4,5
Salvador – BA	128	123	4,5%	806.905	794.787	1,5
Belo Horizonte - MG	125	123	1,5%	1.084.115	1.059.506	2,3
Recife – PE	122	115	6,0%	668.958	659.180	1,5
Curitiba - PR	122	120	1,1%	588.676	575.454	2,3
Maringá – PR	105	104	1,2%	80.007	80.223	-0,3
Vale do Itajaí – SC	103	103	0,9%	75.787	76.159	-0,5
Carbonífera – SC	103	102	1,3%	37.167	37.382	-0,3
Chapecó – SC	103	102	1,2%	22.415	22.492	-0,3
Foz do Rio Itajaí – SC	103	101	1,4%	35.147	35.249	-0,3
Vale do Aço - MG	102	101	1,1%	51.187	51.423	-0,5

FONTE: Adaptado de FIRJAN (2015)

O impacto direto, porém, não é o único impacto relevante decorrente das complicações de tráfego. Um importante aspecto indireto que influencia negativamente a vida dos habitantes da região refere-se à “produção sacrificada” – ou seja, àquilo que deixa de ser produzido na economia devido ao tempo perdido nos deslocamentos. Tal impacto indireto, dado pela perda de produtividade, é apresentado na Tabela 2. Como exemplo da magnitude de tal custo, a produção sacrificada apenas na Região Metropolitana de São Paulo no ano de 2012 ultrapassou 44 bilhões de reais. Mesmo para regiões menores, como o Vale do Itajaí, as cifras são relevantes, ultrapassando 300 milhões de reais. Tais valores, é preciso lembrar, ganham maior expressividade ao se levar em conta a difusão econômica e espacial de tal perda de produtividade: conforme apontam Haddad et al. (2013), o impacto do trânsito pode extrapolar região de origem e afetar o produto de todo o país.

TABELA 2 – EVOLUÇÃO DO CUSTO MÉDIO DO DESLOCAMENTO CASA – TRABALHO – CASA EM ÁREAS METROPOLITANAS BRASILEIRAS

Área Metropolitana	Custo do deslocamento acima de 30 minutos			Custo/PIB	
	R\$1.000		Variação (12/11)	2012	2011
	2012	2011			
São Paulo – SP	44.819.738	43.800.488	2,3%	5,7%	5,5%
Rio de Janeiro – RJ	19.048.148	17.425.491	9,3%	5,9%	5,5%
Belo Horizonte – MG	5.464.372	5.549.206	-1,5%	4,0%	4,1%
Recife – PE	3.366.565	2.936.875	14,6%	4,3%	4,2%
Salvador – BA	3.365.241	3.206.780	4,9%	4,6%	4,2%
Curitiba – PR	3.353.147	3.523.499	-4,8%	3,3%	3,3%
Vale do Itajaí – SC	326.718	328.850	-0,6%	1,5%	1,5%
Foz do Rio Itajaí – SC	252.212	254.367	-0,8%	0,9%	1,0%
Maringá – PR	250.181	245.086	2,1%	1,6%	1,6%
Vale do Aço – MG	189.045	204.074	-7,4%	1,8%	1,9%
Carbonífera – SC	121.378	116.284	4,4%	1,0%	1,0%
Chapecó – SC	84.722	87.528	-3,2%	0,9%	0,9%

FONTE: Adaptado de FIRJAN (2015).

Além das formas já referenciadas, o trânsito afeta a vida das pessoas de diversas formas. A poluição dele decorrente é mais uma faceta a ser elencada com fins ilustrativos da importância de tal problemática. Silva et al (2012) e Silva et al

(2010) apontam os impactos da poluição do ar à saúde cardiorrespiratória. Sørensen et al (2012) e Orban et al (2016), por sua vez, apontam os impactos da poluição sonora advinda do trânsito à saúde cardíaca e mental, respectivamente. Estas complicações de saúde afetam os habitantes, por sua vez, de diversas formas. Diretamente, incorre em todos os danos que uma doença causa à pessoa. Indiretamente, tais complicações impactam diversos habitantes, em decorrência da perda de produtividade pela morbidade ou mortalidade – e até mesmo pelos gastos em saúde pública dispendidos como consequência.

Tendo em vista a importância do trânsito para a economia, fica evidente como é essencial prezar pelo bom funcionamento de tal sistema. Para tanto, políticas públicas adequadas devem ser realizadas, a fim de promover a melhor fluidez ao trânsito e mobilidade aos habitantes. Avaliar uma política pública de transporte antes de sua realização, pois, é importante. Para melhor realização de tal análise, deve-se compreender o processo de escolha dos indivíduos, buscando antever as reações destes – suas novas escolhas de modal – às políticas implementadas.

Uma das principais ferramentas utilizadas para esta avaliação são os modelos de escolha discreta, como pode ser observado pela ampla utilização tanto nacional (LUCINDA; MEYER; LEDO, 2013; LUCINDA et al, 2015, PACHECO; CHAGAS, 2016) quanto internacionalmente (KOPPELMAN; BHAT, 2006; DISSANAYAKE; MORIKAWA, 2010).¹ Modelos de escolha discreta buscam descrever, explicar e até mesmo prever escolhas entre duas ou mais alternativas e, desta forma, figuram como essenciais para a previsão de resultados de políticas públicas de transporte. A utilização de tais modelos, no entanto, apresenta certas ressalvas. Como são ferramentas consolidadas e de potencial amplamente avaliado, suas limitações e dificuldades de aplicação também são exploradas (TRAIN, 2003).

Buscando melhorar as avaliações de políticas de transporte utilizando de modelos de escolha discreta, o presente trabalho propõe 3 diferentes abordagens. Cada uma delas é explorada em um capítulo que, desenvolvido em formato de artigo científico, apresenta estrutura independente.

¹ Ademais, o “desenvolvimento de teoria e métodos para análise de escolha discreta” foi a justificativa da entrega do Prêmio Nobel de Economia de 2000 para Daniel McFadden.

Desta forma, após a introdução, o segundo capítulo, que se refere ao primeiro artigo, propõe uma complementação à base de dados que usualmente embasa modelos de escolha discreta para transportes. Sendo esta geralmente uma base de dados formada a partir de pesquisas de preferência revelada, apresenta apenas os valores dos atributos das alternativas escolhidas pelos respondentes. Para a utilização de um modelo de escolha discreta é necessária, pois, a estimação dos valores dos atributos não escolhidos. Este artigo propõe a utilização da ferramenta Google Maps para tal fim e a compara com as recentes soluções utilizadas pela literatura brasileira. Para tanto, foram estimados modelos de escolha discreta para cada uma das alternativas de complementação, realizando uma aplicação sobre a base de dados da Pesquisa Origem e Destino de 2007 da Região Metropolitana de São Paulo.

Enquanto o segundo capítulo apresenta uma discussão acerca da base de dados sobre a qual ser aplicado o modelo de escolha discreta, o terceiro capítulo discute os pressupostos que embasam tal modelagem. Para tanto, incorpora pontuações da Economia Comportamental que discutem a presença de comportamentos não-maximizadores, como utilização de heurísticas. Tais pontuações são aplicadas à modelagem econométrica conforme apresentada por Swait (2009) e aplicadas à Pesquisa Origem e Destino de 2007 da Região Metropolitana de São Paulo. Os resultados, por fim, são comparados aos recentes trabalhos que utilizam da mesma base de dados.

O último artigo, apresentado no quarto capítulo, consistiu-se em um ensaio voltado à avaliação dos modelos de escolha discreta como ferramenta de previsão de política pública – em específico, de seu caráter estático. Para este fim, um modelo baseado em agentes é apresentado, explorando a dinâmica coevolutiva resultante da interação entre os agentes.

Por fim, encerrando o presente trabalho, uma conclusão geral é realizada, discutindo não apenas as contribuições e limitações de cada artigo, como discutindo as potenciais interações entre as propostas de cada um deles e seus possíveis desdobramentos.

2 ARTIGO 1: USO DO GOOGLE MAPS PARA ESTIMAÇÃO DE VALORES DE ATRIBUTOS NÃO ESCOLHIDOS EM PESQUISAS DE PREFERÊNCIA REVELADA

Resumo

Pesquisas de preferência revelada do tipo Origem e Destino formam bases de dados importantes para análise de políticas de transporte, geralmente embasando modelos de escolha discreta. Um de seus problemas, no entanto, é a ausência de atributos de alternativas não escolhidas pelos entrevistados. Para esse problema, a realização de pesquisa de preferência declarada e a estimação através de técnicas econométricas são as alternativas mais recomendadas, sendo a última a mais presente no Brasil. O presente artigo busca fornecer uma alternativa para obtenção destes valores para a pesquisa de 2007 da Região Metropolitana de São Paulo. Para tanto, construiu-se, com uso do Google Maps, uma base de dados análoga àquelas estimadas em outras pesquisas por Mínimos Quadrados Ordinários. Após discutir a dificuldade de comparação dos resultados através de análise de erros, o texto compara a significância e adequação dos resultados de escolha discreta, maiores com a utilização dos recursos Google.

2.1 INTRODUÇÃO

À medida que as grandes cidades tornam-se mais complexas, o trânsito ganha mais importância: os impactos deste sobre a vida dos habitantes do município, tais como o excessivo tempo despendido nos deslocamentos, extrapolam tal dimensão e passam a afetar o país como um todo, principalmente através da perda de produtividade (HADDAD et al, 2015).

Para o caso brasileiro, a Região Metropolitana de São Paulo, que quase alcança os 20 milhões de habitantes e detém cerca de 19% do PIB brasileiro

(EMPLASA, 2011), desperta especial atenção, uma vez que tais vultosos números refletem sua grande importância econômica.

Buscando melhorar a qualidade de vida da população e melhorar a produtividade da economia, diversas medidas vêm sendo tomadas na região, como a ampliação de trechos das Marginais Tietê e Pinheiros e a implantação de ciclofaixas e faixas exclusivas para ônibus.

No entanto, para aumentar a efetividade de tais medidas, é necessário compreender a escolha dos indivíduos sobre o modal de transporte. Entre as principais ferramentas utilizadas por economistas para tal fim estão os ditos modelos de escolha discreta, amplamente utilizados internacionalmente (KOPPELMAN; BHAT, 2006; DISSANAYAKE; MORIKAWA, 2010) e cuja crescente utilização no país é evidente nos últimos anos (LUCINDA; MEYER; LEDO, 2013; LUCINDA et al, 2015; PACHECO; CHAGAS, 2016).

Pesquisas de preferência revelada do tipo Origem e Destino aparecem como bases de dados importantes para a utilização de tais modelos em estudos sobre transporte. Tais pesquisas, entretanto, apresentam alguns desafios para sua manipulação. Dentre estes, chama a atenção a ausência dos atributos não escolhidas das alternativas não escolhidas pelos entrevistados, necessários para a aplicação da modelagem (HENSHER et al, 2005). O objetivo do presente artigo é apresentar uma alternativa às tradicionais formas de complementação de tais bases de dados: a utilização de informações obtidas com a ferramenta Google Maps. Tal apresentação é realizada com uma aplicação de tal alternativa, buscando complementar a Pesquisa Origem e Destino da Região Metropolitana de São Paulo de 2007.

Na seção seguinte, discorrer-se-á acerca da ausência dos atributos das alternativas não escolhidas e as principais formas utilizadas para suprir tal carência; a terceira seção apresentará o Google Maps como alternativa para este fim, assim como as características do banco de dados com ele formado; a quarta apresentará as possibilidades e dificuldades de comparação entre as alternativas, que será realizada na quinta seção. Por fim, conclui-se com uma discussão sobre as vantagens, dificuldades e outras possibilidades apresentadas pela alternativa proposta.

2.2 AUSÊNCIA DOS ATRIBUTOS DAS ALTERNATIVAS NÃO ESCOLHIDAS

Modelos de demanda por viagens historicamente têm se apoiado em pesquisas de preferência revelada, geralmente consideradas as melhores bases para tal modelagem (BEN-AKIVA; LERMAN, 1985). Dentre as pesquisas desta categoria, aquelas do tipo Origem e Destino apresentam especial relevância. Para as análises de municípios brasileiros, tal relevância é ainda acentuada pela sua ampla utilização há décadas e em diversas localidades (SWAIT, 1984; SWAIT; BEN-AKIVA, 1987; LUCINDA; MEYER; LEDO, 2013; LUCINDA et al, 2015, PACHECO; CHAGAS, 2016).

No entanto, tais bases de dados apresentam como característica a disponibilidade de informações apenas referentes à alternativa escolhida. Desta forma, um dos problemas para lidar com a Pesquisa Origem e Destino é a ausência dos atributos das alternativas não escolhidas pelos entrevistados, carência esta que impossibilita a utilização de modelos de escolha discreta conforme desenvolvidos até então. Quanto à forma de solucionar este problema, Hensher et al (2005) apresentam 4 principais soluções encontradas na literatura.

O primeiro método por eles apresentado envolve a utilização das médias ou medianas dos atributos das alternativas observadas, que são assumidas como os valores para os atributos das alternativas não escolhidas relativas àqueles que não as escolheram. A segunda forma envolve uma distribuição dos níveis dos atributos das alternativas observadas para aquelas não-observadas; tal distribuição pode ser realizada tanto aleatoriamente quanto seguindo estratégias de *matching*. A realização de uma pesquisa de preferência declarada, com questionários sobre situações hipotéticas de escolha, é a terceira forma apresentada pelos autores. Por fim, a quarta alternativa referenciada é descrita como síntese de dados. Como apontam Hensher et al (2005), a norma desta alternativa é “usar informação conhecida tal como distância de viagem ou outras características sociodemográficas e condicionar os dados sintetizados sobre esses”. Exemplificando o uso geral, pode-se citar a utilização de regressões lineares em que a duração de cada modal aparece como variável dependente em cada uma delas, usando dados socioeconômicos e específicos daquela viagem como variáveis independentes.

É importante ressaltar que, além das quatro formas referenciadas por Hensher et al (2005), outras duas vêm sendo apontadas mais recentemente na

literatura: a utilização de matrizes $skim^2$, com dados de duração e distância entre zonas, como apontado por Washington et al (2014) e Javanmardi et al (2015), e métodos de imputação múltipla Bayesiana, como realizada por Washington et al (2014).

Embora Hensher et al (2005) recomendem a utilização das duas últimas alternativas por eles apresentadas e as formas mais recentemente utilizadas (e acima referenciadas) venham apresentando bons resultados, as dificuldades relacionadas à obtenção destes tipos de dados favorece a ampla utilização da alternativa sintética na literatura brasileira. Com relação a esta, destacam-se Lucinda, Meyer e Ledo (2013), Barcellos (2014), Lucinda et al (2015) e Feres (2015), que utilizaram a alternativa sintética para a complementação da Pesquisa Origem e Destino 2007 da Região Metropolitana de São Paulo.

Antes de avançar à discussão de tal síntese, é importante destacar algumas das características de tal pesquisa. Aplicada pela Companhia do Metropolitano de São Paulo, o Metrô, traz 168.582 observações com informações de viagens realizadas na referida região metropolitana. Para tanto, cerca de 30 mil domicílios foram entrevistados e validados, apresentando características sociodemográficas também a nível individual, como renda e idade.

A fim de complementar esta pesquisa quanto à carência dos atributos das alternativas não escolhidas, os referidos autores brasileiros realizaram uma mesma síntese de dados utilizando Mínimos Quadrados Ordinários. Para este caso, foi necessário estimar o custo e o tempo de opções que, embora disponíveis, não tenham sido escolhidas pelo tomador de decisão.

Com exceção do custo de algumas alternativas que pôde ser determinado através de informações externas, como de ônibus, metrô e trem, o custo das demais alternativas e o tempo de todas elas foram estimados a partir de um modelo de regressão. Tal modelo usou as escolhas observadas para cada modal, sendo as variáveis dependentes os logaritmos do custo e da duração da viagem. As variáveis independentes utilizadas foram tanto variáveis *dummy* para a hora de partida e de chegada e para motivação da viagem, como a distância em quilômetros entre as

² Matrizes que possuem informações, como distância, pertinentes às zonas da área em questão.

zonas de origem e destino³. A partir de tal modelo, estimou-se coeficientes que, por sua vez, foram utilizados para a obtenção do tempo e do custo esperados para as alternativas não escolhidas.

É importante ressaltar que este tipo de complementação deve levar em conta a disponibilidade de cada modal para cada tomador de decisão, uma vez que a adequação do conjunto de escolha é essencial para a estimação (SWAIT, 1984). A fim de minimizar tal problema, os autores restringiram a disponibilidade de escolha de trem e metrô apenas para as zonas que tivessem algum respondente escolhendo um destes modais.

Por fim, tal síntese, por sua recente e recorrente aplicação na literatura brasileira, figura como referência para a formação de uma base de dados complementar à Pesquisa Origem e Destino, a ser comparada à alternativa proposta no presente artigo, apresentada na seção seguinte.

2.3 GOOGLE MAPS COMO ALTERNATIVA

Como opção às alternativas propostas por Hensher et al (2005) e em especial à síntese realizada por Lucinda, Meyer e Ledo (2013), Barcellos (2014), Lucinda et al (2015) e Feres (2015), propõe-se, no presente artigo, a utilização da ferramenta de pesquisa e visualização de mapas e rotas da empresa estadunidense Google, o Google Maps. Tal alternativa foi proposta apenas recentemente e em comparação à utilização de matrizes skim, pelo trabalho em desenvolvimento de Javanmardi et al (2015).

Dentre sua ampla variedade de recursos, importa à presente pesquisa o fato de tal ferramenta ser capaz de sugerir rotas entre origem e destino através de diversos modais. Sobre tais rotas, provê informações como custo e duração, ficando evidente sua possibilidade de utilização a fim de suprir a referida carência das Pesquisas Origem e Destino.

Em comparação com pesquisas de preferência declarada, a utilização de tal ferramenta apresenta um custo muito menor, além de poder apresentar informações mais precisas em relação aos trajetos que os indivíduos não costumam utilizar. Comparando-se à alternativa sintética, por sua vez, o potencial do Google Maps

³ A Pesquisa Origem e Destino da Região Metropolitana de São Paulo de 2007 divide a região em 460 zonas.

reside na utilização de maior quantidade e diversidade de informações, que podem ser ilustrados pelos milhões de usuários diários que a ferramenta possui, além da utilização de georreferenciamento e de informações sobre linhas de transporte público. Estas últimas informações, além de serem potencialmente importantes para a obtenção de estimativas mais precisas, também permite que se componha o conjunto de escolhas disponível para o tomador de decisão de forma mais adequada.

Com a finalidade de obter tais dados de forma sistemática, foram utilizados outros recursos da própria Google, sendo: Google Maps API, Google Apps Script e Google Sheets. Com a utilização destes recursos, os dados foram então obtidos a partir das coordenadas geográficas fornecidas pela Pesquisa Origem e Destino, de tal forma a obter a duração prevista para viagens a partir de transporte individual motorizado, a pé e das três melhores opções de transporte público de acordo com o ordenamento do próprio Google Maps. As alternativas de transporte público, quando iguais, foram agregadas na de menor duração.

É importante ressaltar que o Google Maps utiliza as definições das linhas de transporte mais recentes para o planejamento da rota. Dessa forma, programou-se para que os resultados que incluíssem linhas de transporte público adicionadas após a Pesquisa Origem e Destino 2007 fossem descartadas.

Enfim, com os dados coletados, foi montada uma base de dados complementar à Pesquisa Origem e Destino 2007 da RMSP de maneira análoga àquela sintetizada pelos referidos autores brasileiros. Devido às restrições quanto a quantidade de dados coletados diariamente, foi utilizado, para o presente artigo, um recorte da Pesquisa Origem e Destino 2007, composta de observações de 25995 indivíduos, referentes às viagens motivadas a trabalho, no horário de pico da manhã (das 6h às 9h59). Tal seleção é análoga à realizada por Swait e Ben-Akiva (1987) e também foi realizada sobre a amostra utilizando a formatação referência para a quarta alternativa proposta por Hensher et al (2005).

2.4 DIFICULDADES DE COMPARAÇÃO

A fim de comparar as duas metodologias - a sintética e a coleta via Google Maps -, três formas principais são apresentadas: uma teórica, uma através de análise de erros e, enfim, uma através de verificação de ajuste e significância.

A comparação teórica, no sentido de análise das propriedades dos estimadores, aparece como primeira opção, seguindo a prática comum na literatura econométrica. A dificuldade em se obter o método de estimação da duração utilizado pelo Google Maps - uma espécie de caixa preta - impede, no entanto, tal abordagem.

A segunda alternativa de comparação refere-se à análise de erros, comparando erros médios e erros absolutos médios entre as estimações. No entanto, para obter valores para tais erros, é necessário fazer a comparação da estimação com os dados reais - no caso, com aqueles disponíveis na Pesquisa Origem e Destino. Assim, pode-se comparar apenas os tempos reais e os tempos estimados da mesma observação e, portanto, de um mesmo modal. Em decorrência do método utilizado pelos autores de utilizar a hora de partida e chegada para a estimação, é razoável supor que os erros possíveis de serem avaliados sejam menores do que aqueles referentes às estimações das alternativas não escolhidas, uma vez que não é possível utilizar-se de uma informação tão precisa para este caso.

A fim de exemplificar tal situação, pode-se considerar dois vizinhos que compartilhem o mesmo local de trabalho. O primeiro vizinho se locomove ao trabalho utilizando ônibus e, para que chegue às 8h40, deve sair às 6h40. O segundo vizinho, no entanto, vai de carro e, para chegar no horário do trabalho, deverá sair às 8h10. Desta forma, ainda que a estimação preveja perfeitamente a duração do deslocamento de ônibus para o primeiro vizinho, o fará porque utiliza os "inputs" corretos. Caso realize a previsão a partir dos "inputs" da viagem por carro, provavelmente o fará de maneira errônea, uma vez que, ao invés de utilizar a hora de partida adequada (6h), utilizará aquela relativa ao carro (8h). Assim, é razoável supor que estime um tempo diferente e, com isso, um erro maior. Tal situação é ilustrada na Tabela 3.

TABELA 3 – EXEMPLO ILUSTRATIVO DA DIFICULDADE DE COMPARAÇÃO POR ANÁLISE DE ERROS

Vizinhos	Modal Escolhido	Hora de Partida	Hora de Chegada	Duração Modal	Duração Estimada
Vizinho 1	Ônibus	6h	8h	2h	2h
Vizinho 2	Carro	8h	8h	30min	? 2h

FONTE: elaboração própria

De fato, tal dificuldade não advém apenas da utilização de tais variáveis, mas é também decorrente de que tal estimação pode apresentar viés de seleção amostral. Conforme apontam Washington et al (2014, tradução nossa): “Imputar ou sintetizar valores de atributos de alternativas não escolhidas é um problema de dados faltantes. Em contraste aos Dados Faltantes Completamente Aleatórios e aos Dados Faltantes Aleatórios, atributos dos modais de transporte faltantes são uma função de seus estado não observado e não-ignorável”.

Desta forma, observa-se que, sendo um caso de Dados Faltantes Não Aleatórios, utilizar-se das informações sobre as viagens através de um modal fornecidas apenas por aqueles que o utilizaram, poderá implicar em viés de seleção quando da estimação para aqueles que não utilizaram tal meio de transporte.

Tal viés de seleção será ilustrado na seção seguinte, que apresentará breve comparação entre as amostras complementares estimadas, assim como comparação entre as estimações dos modelos de escolha discreta utilizando-se das diferentes bases de dados.

2.5 ANÁLISE E COMPARAÇÃO DAS ESTIMAÇÕES

A fim de comparar as alternativas para complementação da base de dados, deve-se, pois, comparar também os resultados das estimações dos modelos de escolha discreta. Dentre os trabalhos referidos que optam por sintetizar os atributos das alternativas não escolhidas, Lucinda, Meyer e Ledo (2013) e Barcellos (2014) apresentam tais informações de forma clara e completa e, desta forma, são aqui utilizados para comparação com as estimações obtidas com uso do Google Maps.

O recorte imposto à base de dados complementada com utilização desta ferramenta a diferenciou das subamostras utilizadas nos referidos artigos. Assim, optou-se por reestimar os valores das variáveis não escolhidas utilizando das mesmas especificações por eles apresentadas - a diferença, final, refere-se apenas a quais foram observações utilizadas para a estimação do modelo de escolha discreta.

Para melhor compreensão dos resultados, esta seção dividir-se-á em três partes, referindo-se primeiramente a uma breve comparação das bases de dados

estimadas, seguindo à estimação realizada pelos autores citados e então à dos resultados dos modelos de escolha discreta.

2.5.1 Comparação entre Bases de Dados Estimadas

A fim de ilustrar indícios de viés de seleção resultante da estimação, através de regressão linear, dos atributos das alternativas não escolhidas (no caso, as durações através de tais modais) a partir dos atributos das alternativas escolhidas, duas tabelas são aqui apresentadas.

A primeira delas, Tabela 4, apresenta um apanhado de 3 indivíduos da subamostra utilizada para a estimação dos modelos discretos, com suas respectivas alternativas possíveis, escolhas de modal e durações de viagem, em minutos, tanto da estimação realizada através de regressão, quanto daquela utilizando-se do Google Maps.

TABELA 4 – APANHADO ILUSTRATIVO PARA COMPARAÇÃO ENTRE BASES ESTIMADAS

ID do Respondente	Alternativa	Escolha	Duração Regressão	Duração Google Maps
10	Carro	Não	32	18
10	Ônibus	Sim	90	90
10	Outros	Não	22	89
11	Carro	Não	36	23
11	Ônibus	Sim	90	90
11	Outros	Não	25	115
12	Carro	Não	40	25
12	Ônibus	Sim	105	105
12	Outros	Não	27	141

FONTE: elaboração própria.

Embora outras alternativas de modais estejam disponíveis, foram apresentadas às de maior utilização pela população: carro, ônibus e outros. A partir da observação desses indivíduos, nota-se que as durações obtidas através do Google Maps para viagens de carro são recorrentemente inferiores àquelas estimadas. Isto pode decorrer do fato de que muitos dos indivíduos que não utilizem

de automóvel, deixem de usá-lo porque, dada a curta distância do trajeto, optem pelo baixo custo de ir a pé ou de bicicleta (e de outras vantagens que tais modais podem apresentar), apesar de o tempo a ser despendido seja maior do que aquele utilizando-se de carro.

Ainda é importante notar que, ao se analisar os valores estimados para viagens utilizando-se dos modais agrupados em “Outros”, observa-se que os indivíduos apresentam, de acordo com a regressão, viagens com durações menores inclusive do que as viagens de carro. Além disso, apesar de a duração ser tão menor do que a duração de ônibus e não incorrer em custos, a opção dos indivíduos é a mesma por tal transporte coletivo.

Tais características são observadas também através da Tabela 5, que traz as médias de duração em minutos para cada subamostra de interesse. Observa-se que a regressão estima uma média de tempo utilizando-se de “Outros” inferior àquela estimada para a utilização de automóveis. Ainda que esteja sintonizada com as médias encontradas na Pesquisa Origem e Destino, deve-se lembrar que esta subamostra corresponde apenas às viagens daqueles que optaram por viajar através destes modais. O viés de seleção, neste caso, parece subestimar a duração das viagens.

Enquanto as médias de duração através de automóveis reforçam a interpretação explorada anteriormente, as médias referentes às viagens de ônibus podem também indicar viés de seleção, porém em um sentido contrário - de afastamento entre os valores obtidos entre a subamostra com dados apenas da Pesquisa Origem e Destino e aquela complementada através do uso de regressão. Tal fato pode se dar em decorrência de as pessoas que se utilizam de ônibus o fazerem também pela disponibilidade de infraestrutura, sendo esta possivelmente menor para aqueles que não a utilizam. Desta forma, a regressão poderia obter parâmetros que subestimam o tempo de tal modal.

Por fim, a comparação entre a ordenação e magnitude dos valores imputados através da regressão e do Google Maps aponta que os resultados do último corroboram com o esperado pela intuição: as viagens de carro são as de menor duração, seguidas com alguma margem pelas de ônibus, enquanto as viagens utilizando-se do modal “Outros”, como viagens a pé ou de bicicleta apresentam um tempo médio razoavelmente superior aos demais.

TABELA 5 - MÉDIAS DE DURAÇÃO POR SUBAMOSTRA DOS TRÊS MODAIS DE MAIOR UTILIZAÇÃO

Subamostra	Ônibus	Carro	Outros
Origem e Destino	64.6	36.3	20.4
Regressão	52.1	37.0	27.5
Google Maps	62.5	26.9	102.6

FONTE: elaboração própria.

2.5.2 Estimação dos Modelos de Escolha Discreta

Tanto o artigo de Barcellos (2014) quanto de Lucinda et al (2013) utilizam-se dos mesmos modelos de escolha discreta: o Logit Multinomial e o Misto. Antes de prosseguir aos procedimentos práticos concernentes à estimação, é relevante uma breve apresentação de tais modelagens.

Estes modelos compreendem que existem fatores que coletivamente determinam a escolha dos agentes, mas que nem todos são observados pelo pesquisador. O processo de escolha, dessa forma, é expresso por uma função do tipo $y = h(x, \varepsilon)$, em que x refere-se aos fatores observados e ε aos fatores não observados. Contendo termos não observados, pois, não é possível prever y com exatidão. Ao invés disso, uma probabilidade de resultado é então derivada a partir da consideração de que os termos não observados são aleatórios com densidade $f(\varepsilon)$. Assim, a probabilidade de um tal resultado, dados os fatores observados, pode ser expressa por: $P(y|x) = Prob(\varepsilon \text{ s. a. } h(x, \varepsilon) = y)$.

Utilizando-se do arcabouço de maximização de utilidade, pode-se reexpressar a modelagem acima, assumindo que a *utilidade representativa*, V_{nj} , é determinada pelos atributos das J alternativas encontradas pelos N agentes, x_{nj} , assim como pelas características destes agentes, s_n : $V_{nj} = (x_{nj}, s_n) \forall j$. A utilidade representativa, no entanto, difere da utilidade real, U_{nj} , que possui um componente aleatório: $U_{nj} = V_{nj} + \varepsilon_{nj}$. Desta forma, a probabilidade do indivíduo n escolher a alternativa i em detrimento às outras alternativas é a probabilidade de que a utilidade obtida por tal indivíduo proveniente da escolha de i seja maior do que as demais, tal como $P_{ni} = Prob(U_{ni} > U_{nj}, \forall j \neq i)$ (TRAIN, 2003).

O modelo Logit Multinomial, o mais simples e amplamente utilizado para este tipo de modelagem, assume que cada ε_{nj} é um valor extremo independente e

identicamente distribuído de acordo com uma distribuição Gumbel. Com esta distribuição de probabilidade, McFadden (1974) obteve a função de probabilidade de escolha Logit Multinomial, $P_{ni} = Prob(U_{ni} > U_{nj}, \forall j \neq i) = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$.

Uma importante limitação deste modelo Logit decorre justamente das hipóteses sobre as características de probabilidades de escolha das quais originalmente Luce (1959) derivou sua formulação. Tais hipóteses constituem a propriedade de independência das alternativas irrelevantes (*independence from irrelevant alternatives* - IIA), segundo a qual, para duas alternativas i e k , a razão de probabilidade Logit não depende de outras alternativas que não i e k . Tal propriedade não apenas restringe os padrões de substituição possíveis do modelo, como também implica em erro de previsão sempre que a razão de probabilidades entre duas alternativas mudar com a introdução ou alteração de uma outra alternativa (TRAIN, 2003).

Este Logit básico apresenta duas outras importantes limitações, além desta decorrente da propriedade IIA. Embora possa representar variação de preferência sistemática, é incapaz de fazê-lo para variações aleatórias. Tal modelagem, ainda, não pode ser usada com dados em painel quando fatores observados são correlacionados ao longo do tempo para cada tomador de decisão. (TRAIN, 2003)

Superando tais limitações, foram desenvolvidos os modelos Logit Mistos que, sendo definidos sobre a base da forma funcional de suas probabilidades de escolha, permitem diferentes especificações comportamentais. A fim de ilustrar tal modelagem, assume-se aqui que $U_{nj} = \beta'_n x_{nj} + \epsilon_{nj}$, onde x_{nj} são as variáveis observadas que se relacionam aos tomadores de decisão e às alternativas, β_n , é um vetor de coeficientes dessas variáveis para a pessoa n , e ϵ_{nj} é um termo aleatório na população com densidade $f = (\beta)$. Como observado, apresenta, pois, a mesma especificação do Logit Multinomial, com exceção da variabilidade de beta entre os indivíduos. Desta forma, para este modelo, as probabilidades são as integrais de probabilidades Logit Padrão multiplicadas por uma função de densidade dos parâmetros: $P_{ni} = \int \left(\frac{e^{\beta'_n x_{ni}}}{\sum_j e^{\beta'_n x_{nj}}} \right) f(\beta) d\beta$.

Realizada esta breve apresentação, pode-se prosseguir, então, aos procedimentos diretamente relacionados à estimação. Primeiramente, é importante especificar a forma como os autores agregaram os modais para o fim da pesquisa -

o fizeram constituindo como alternativas as seguintes categorias: automóvel, ônibus, metrô e trem, motocicleta, táxi e outros.

Por fim, quanto às variáveis explicativas, custo e tempo foram utilizadas em todos os modelos como específicas das alternativas. As variáveis dos indivíduos consideradas, por sua vez, foram:

- idade;
- renda individual (em R\$1000);
- tamanho da família (medido em número de moradores no domicílio);
- *dummy* para emprego formal;
- *dummy* para gênero feminino;
- *dummy* para estudante.

Enquanto o modelo apresentado por Lucinda et al (2013) utilizou as variáveis apontadas, aquele desenvolvido por Barcellos (2014) incluiu também iterações de custo e tempo com a renda individual. A fim de estabelecer comparações, foi estimado um Logit Multinomial na mesma configuração de Barcellos (2014), e Logit Mistos nas configurações de ambos os trabalhos, utilizando-se da base de dados obtida com o Google Maps. Tais resultados são apresentados a seguir.

2.5.3 Comparação de Resultados dos Modelos de Escolha Discreta

A primeira comparação é estabelecida entre o modelo Logit Multinomial estimado como Barcellos (2014) e aquele análogo estimado com uso do Google Maps. Tal comparação é apresentada na Tabela 6 e a referência à autora é mantida para fins de facilitar a compreensão, apesar de toda a estimação ter sido replicada no presente trabalho. É importante ressaltar que todas as estimações foram realizadas utilizando a alternativa “Outros” como referência e, portanto, todos os coeficientes apresentados o são em relação a este modal.

TABELA 6 - COMPARAÇÃO DAS ESTIMAÇÕES LOGIT MULTINOMIAL

	Google Maps	Barcellos (2014)
Variáveis Específicas das Alternativas		
Custo	-0.1262 ***	-0.1287 ***
Tempo	-0.0153 ***	0.0337 ***

Custo X Renda	0.0098 ***	0.0143 ***
Tempo X Renda	0.0002	-0.0051 ***
Ônibus		
Renda (R\$1000)	-0.1360 ***	0.0663 ***
Idade	-0.0032 ***	-0.0120 ***
Mulher	0.2298 ***	-0.0041
Estudante	0.2691 ***	0.1957 ***
Tamanho Família	-0.0219 ***	-0.0634 ***
Emprego Formal	0.7831 ***	0.4343 ***
Trem e Metrô		
Renda (R\$1000)	-0.0255	0.2819 ***
Idade	-0.0092 ***	-0.0261 ***
Mulher	0.1457 ***	-0.2894 ***
Estudante	0.2816 ***	0.0785 *
Tamanho Família	-0.0077	-0.1201 ***
Emprego Formal	1.1626 ***	0.3112 ***
Carro		
Renda (R\$1000)	0.3273 ***	0.3822 ***
Idade	0.0090 ***	0.0104 ***
Mulher	-1.0953 ***	-1.1802 ***
Estudante	-0.2708 ***	0.2521 ***
Tamanho Família	-0.2169 ***	-0.1656 ***
Emprego Formal	0.0895 **	0.4286 ***
Moto		
Renda (R\$1000)	0.2116 ***	0.2386 ***
Idade	-0.0610 ***	-0.0532 ***
Mulher	-2.5687 ***	-2.6767 ***
Estudante	-0.2798 ***	-0.1965 ***
Tamanho Família	-0.1390 ***	-0.0861 ***

Emprego Formal	0.2006 **	0.7708 ***
Táxi		
Renda (R\$1000)	0.2773 ***	0.2160 ***
Idade	-0.0134	-0.0092
Mulher	-0.7819 *	-0.3800
Estudante	-0.9482	-0.1456
Tamanho Família	-0.8544 ***	-1.1675 ***
Emprego Formal	-0.8846	-0.2484
Log-Verossimilhança (Convergência)	-26778	-33630
R ² de McFadden	0.29044	0.10889

Fonte: elaboração própria.

p < 0.1: *; p < 0.05: **; p < 0.01 ***

Inicialmente, observa-se um maior valor para o log-verossimilhança de convergência para o modelo imputado utilizando-se do Google Maps. O mesmo ocorre para o R² de McFadden que, análogo ao R² tradicional, avalia a significância do modelo estimado. Tal consideração indica uma maior capacidade de previsão e ajustamento do modelo utilizando a base de dados proposta, e vai no mesmo sentido dos resultados obtidos por Javanmarti et al (2015) em comparação à utilização de matrizes *skim*.

Além de tal resultado, é relevante notar que a estimação da forma de Barcellos (2014) gerou um coeficiente com sinal positivo para a variável tempo, enquanto a alternativa proposta apresentou o sinal negativo. O sinal da última corrobora com o esperado intuitivamente: a utilidade dos indivíduos diminui conforme a duração da viagem aumenta. O coeficiente imputado para a interação entre tempo e renda, no entanto, diminui o problema apresentado no modelo estimado com uso de regressão: com o sinal negativo e estatisticamente significativo, indica que, conforme a renda aumenta, a duração passa a ser menos apreciada, até que gere desutilidade.

Este resultado pode ser proveniente da subestimação da duração para o modal “Outros” quando da utilização da regressão linear. Como discutido anteriormente, tal subestimação pode, por sua vez, advir do viés de seleção

amostral por utilizar-se apenas das alternativas escolhidas para a estimação dos atributos das alternativas não escolhidas. Corroborando com o apresentado em 5.1, a base de dados imputada com uso do Google Maps parece sofrer menos de tal dificuldade.

Quanto às variáveis relacionadas aos indivíduos, é relevante observar que os coeficientes relacionados aos gênero feminino apresentaram alterações tanto para ônibus, quanto para trem e metrô e táxi. Tais coeficientes passaram a ser estatisticamente significantes com uso da base de dados do Google Maps para os modais ônibus e táxi. O sinal dos coeficientes, no entanto, passaram a ser positivos para ônibus e trem e metrô, e negativos para táxi. Ainda que possam ser contraintuitivos, tal fato pode ser explicado pela tradicional estrutura da família brasileira que, colocando o homem como “chefe de família”, favorece a utilização do homem pelo carro (ou o táxi), restando à mulher os modais alternativos. Diversos estudos, como em Swait (1987), utilizam da variável “chefe de família” para melhor especificar seus modelos. Ainda que esta pudesse ser uma alternativa a ser explorada, optou-se por ser o mais fidedigno possível aos estudos de Barcellos (2014) e Lucinda et al (2013).

Por fim, é importante ressaltar que os coeficientes de renda e tamanho da família relacionados ao modal Trem e Metrô deixaram de ser estatisticamente significantes no modelo que utiliza da base de dados proposta. Por outro lado, os coeficientes de renda da alternativa ônibus e da *dummy* de estudante para a alternativa carro passaram a ter o sinal esperado: negativo.

Finalizando assim a comparação entre os modelos Logit Multinomial, resta analisar os modelos Logit Mistos. Assim, apresenta-se na Tabela 7 a comparação entre as estimações utilizando tal modelagem: a de Barcellos (2014), a de Lucinda et al (2013) e duas estimações utilizando a base de dados do Google Maps, cada uma utilizando uma das especificações de variáveis de cada uma das estimações referidas. Estas estimações estão na coluna central, sendo que se comparam aos modelos apresentados na coluna mais próxima. Para exemplificar, vemos que o coeficiente estimado para renda usando a base de dados e especificação como Barcellos (2014) foi de 0.0164, não sendo estatisticamente significativo em níveis abaixo de 10%. Vemos também que, utilizando da base complementar estimada com auxílio do Google Maps, o coeficiente para renda, no modelo de especificação análogo ao de Barcellos (2014) foi de -0.1214, enquanto na especificação análoga à

de Lucinda et al (2013) foi de -0.1212 - sendo ambos estatisticamente significantes a 1%. Por fim, o coeficiente para renda com base de dados e especificação como Lucinda et al (2013) foi de -0.0271 e não pode ser considerado estatisticamente significativo a níveis menores do que 10%.

TABELA 7 - COMPARAÇÃO ENTRE ESTIMAÇÕES LOGIT MISTAS

	Barcellos (2014)	Google Maps	Lucinda et al (2013)
Fixos			
Ônibus			
Renda (R\$1000)	0.0164	-0.1214 *** / -0.1212 ***	-0.0271
Idade	-0.0190 ***	-0.0051 *** / -0.0039 ***	-0.0188 ***
Mulher	0.0292	0.2750 *** / 0.2741 ***	0.0319
Estudante	0.2152 ***	0.3080 *** / 0.3108 ***	0.2178 ***
Tamanho Família	-0.0994 ***	-0.0249 *** / -0.0204 ***	-0.0978 ***
Emprego Formal	0.2032 ***	0.8609 *** / 0.8883 ***	0.2143 ***
Trem e Metrô			
Renda (R\$1000)	0.2484 ***	-0.0282 / 0.0278	0.2006 ***
Idade	-0.0346 ***	-0.0095 *** / -0.0093 ***	-0.0343 ***
Mulher	-0.2660 ***	0.1926 *** / 0.1918 ***	-0.2622
Estudante	0.0614	0.2949 *** / 0.3138 ***	0.0648
Tamanho Família	-0.1653 ***	0.0035 / 0.0059	-0.1636 ***
Emprego Formal	0.0816	1.3515 *** / 1.3424 ***	0.0944 *
Carro			
Renda (R\$1000)	0.3593 ***	0.3362 *** / 0.3551 ***	0.3566 ***
Idade	0.0100 ***	0.0108 *** / 0.0109 ***	0.0101 ***
Mulher	-1.1891 ***	-1.2124 *** / -1.231 ***	-1.1915 ***
Estudante	-0.2906 ***	-0.3346 *** / -0.3405 ***	-0.2923 ***
Tamanho Família	-0.1786 ***	-0.2405 *** / -0.2403 ***	-0.0942 ***
Emprego Formal	0.3369 ***	0.0532 / 0.0544	0.3292 ***

Moto			
Renda (R\$1000)	0.2318 ***	0.2702 *** / 0.2239 ***	0.2342 ***
Idade	-0.0528 ***	-0.0733 *** / - 0.0725 ***	-0.0528 ***
Mulher	-2.6822 ***	-2.9481 *** / -3.033 ***	-2.6812 ***
Estudante	-0.2024 ***	-0.5610 *** / -0.4892 ***	-0.2027 ***
Tamanho Família	-0.0939 ***	-0.1311 *** / -0.1335 ***	-0.0942 ***
Emprego Formal	0.7611 ***	0.3016 *** / 0.2789 ***	0.7596 ***
Táxi			
Renda (R\$1000)	0.3279 ***	0.9402 *** / 0.9945 ***	0.5066 ***
Idade	-0.0063	0.0482 *** / 0.0506 ***	-0.0087
Mulher	-0.7040	-0.6523 / -0.1605	-0.7594 *
Estudante	-1.3724	-17.951 *** / -21.475 ***	-1.4547
Tamanho Família	-1.1672 ***	-3.4182 *** / -3.7853 ***	-1.1782 ***
Emprego Formal	-1.3238 **	-13.492 *** / -11.728 ***	-1.4133 ***
Desvios-padrões			
Custo X Renda	0.0084 ***	0.0072 * / -	-
Tempo X Renda	-0.0034 ***	-0.0010 *** / -	-
Custo	-0.1926 ***	-0.1122 *** / -0.1933 ***	-0.0629 ***
Tempo	-0.0173 ***	-0.0116 *** / -0.0189 ***	0.0554 ***
Custo	0.0963 ***	0.0453 *** / 0.0825 ***	0.0008
Tempo	0.0023	0.0068 *** / 0.0039 ***	0.0445 ***
Log-Verossimilhança (Convergência)	-32082	-27377 / -27403	-32096
R ² de McFadden	0.14988	0.27457 / 0.27389	0.14951

Fonte: elaboração própria.

p < 0.1: *; p < 0.05: **; p < 0.01 ***

Assim como observado para os modelos anteriores, observa-se de imediato um maior valor para log-verossimilhança de convergência para os dois modelos que utilizaram de bases de dados complementadas com a utilização do Google Maps.

Quanto ao coeficiente de tempo, observa-se que o Logit Misto na formatação de Barcellos (2014) logrou superar a dificuldade anterior, apresentando agora sinal negativo. Aquele que seguiu Lucinda et al (2013), no entanto, manteve o problema. É visto, ainda, que o modelo estimado com a base de dados alternativa e a formatação de Barcellos (2014) passou a apresentar coeficientes estatisticamente significativos e com o sinal esperado para as variáveis iteradas.

Apesar da melhoria apresentada, alguns coeficientes deixaram de ser significantes, como estudante em relação a trem e metrô para ambos os modelos que utilizaram base de dados estimadas com regressão linear. Também deixaram de ser significantes o coeficiente de emprego formal para trem e metrô na estimação como Barcellos (2014) e para carro nas estimações com uso do Google Maps. Ainda, o coeficiente relacionado às mulheres com relação a táxi manteve sua significância apenas no modelo análogo ao de Lucinda et al (2013).

Por fim, é relevante destacar as alterações observadas no modelo de base de dados alternativa quanto ao táxi. Para este modal, os coeficientes de idade e estudante passaram a ser estatisticamente significantes, além de apresentarem sinais razoáveis: quanto maior a idade, maior a utilidade de se utilizar de tal transporte, enquanto ser estudante diminui, e muito, a probabilidade de escolhê-lo

2.6 CONCLUSÃO

O presente artigo propôs uma alternativa às práticas da literatura quanto à dificuldade de lidar com a ausência dos atributos das alternativas não escolhidas em pesquisas de preferência revelada do tipo Origem e Destino. Ainda que dificuldades tenham sido encontradas para a comparação entre a alternativa proposta, a utilização do Google Maps, e a síntese desenvolvida por Lucinda et al (2013) e Barcellos (2014), a estimação dos modelos de escolha discreta apresentou resultados com melhores propriedades estatísticas com uso da alternativa aqui proposta. Ressalta-se novamente que os achados corroboram com as semelhantes vantagens percebidas por Javanmardi et al (2015).

É razoável crer que a contribuição da alternativa proposta oscile, a depender das características das informações possuídas pelo Google Maps para cada localidade. Em cidades com menor utilização de ferramentas Google ou relacionadas, como as mais afastadas cidades do interior, a quantidade e qualidade da informação provavelmente serão menos interessantes do que aquelas obtidas em grandes cidades, como São Paulo e Curitiba. No entanto, as poucas barreiras quanto à construção de tal base de dados complementar permite que seja avaliada sua qualidade aplicado-a em novos estudos.

É importante destacar que tal banco de dados possui outras vantagens para além dos aspectos apontados da complementação. Como observado na análise dos resultados, a agregação de metrô e trem em uma mesma alternativa pode representar um problema, pois é razoável crer que as características dos usuários de tais modais sejam bastante diferentes. Tal agregação é justificada pela pequena amostra de usuários de trem, que pode ser resolvida com a amplitude de informações obtidas sobre este modal com o Google Maps - mesmo para aqueles que não o escolhem. Outra vantagem potencial da alternativa proposta reside na melhor especificação da cesta de escolhas disponível para cada indivíduo, decorrente do uso de informações georreferenciadas sobre as linhas de transporte público. Tais vantagens apontadas ilustram, mas não encerram o potencial por ela apresentado.

Deve-se ressaltar, por fim, que outras especificações de modelos poderiam ter sido exploradas, mas a critério de comparação, optou-se por manter adesão às estimações realizadas por Barcellos (2014) e Lucinda et al (2013). Ademais, a base de dados pode ser explorada de outras formas além da estimação direta para modelos de escolha discreta. Seguindo os trabalhos de Washington et al (2014), é razoável a utilização de tal base como distribuição a priori a realizar a imputação Bayesiana pelos autores desenvolvida.

3 ARTIGO 2: MODELAGEM 2-MIX – UMA APLICAÇÃO COM NOÇÕES DA ECONOMIA COMPORTAMENTAL À REGIÃO METROPOLITANA DE SÃO PAULO

Resumo

Modelos de escolha discreta, como os modelos Logit, são ferramentas importantes para avaliação de políticas públicas. Sendo baseados em pressupostos de maximização de utilidade tradicional, a incorporação de proposições relacionadas à Economia Comportamental pode contribuir para a compreensão da escolha. Este artigo realiza uma análise empírica de um modelo de escolha discreta por modal de transporte na Região Metropolitana de São Paulo em viagens a trabalho, considerando regras compensatórias e não compensatórias. Para tanto, utiliza-se da modelagem de funções de densidade de probabilidade mistas desenvolvida por Swait (2009), aplicando o Modelo 2-Mix com Condição de Rejeição a dados da Pesquisa Origem e Destino recentes (2007). Os resultados apontam uma maior significância estatística para o modelo proposto, sugerindo a utilização também de regras não compensatórias.

3.1 INTRODUÇÃO

À medida que as grandes cidades tornam-se mais complexas, o trânsito ganha mais importância: seus impactos, principalmente através da perda de produtividade, podem passar a extrapolar a vida dos munícipes e afetar o país como um todo (HADDAD et al, 2015).

Os grandes números, tanto econômicos quanto demográficos, da Região Metropolitana de São Paulo são ilustrativos da importância e do impacto que o trânsito de tal região pode exercer na economia brasileira: ela quase alcança os 20 milhões de habitantes e detém cerca de 19% do PIB do país (EMPLASA, 2011).

Buscando melhorar a qualidade de vida da população e melhorar a produtividade da economia, diversas medidas vêm sendo tomadas na região, como

a ampliação de trechos das Marginais Tietê e Pinheiros e a implantação de ciclofaixas e faixas exclusivas para ônibus.

No entanto, para aumentar a efetividade de tais medidas, é necessário compreender a escolha dos indivíduos sobre o modal de transporte, evitando que grandes investimentos sejam desperdiçados por incapacidade de compreender como os indivíduos reagirão quanto a suas escolhas em decorrência dessas políticas. Nesse sentido, o presente trabalho busca aprofundar trabalhos anteriores (BARCELLOS, 2014; LUCINDA et al, 2013) incorporando contribuições da Economia Comportamental (SWAIT, 2009). Mais especificamente, o presente trabalho aplicou a modelagem desenvolvida por Swait (2009) - a utilização de funções de densidade de probabilidades mistas em modelos de escolha discreta, a fim de representar e capturar comportamentos não maximizadores - aos dados mais recentes da Pesquisa Origem e Destino, como em Barcellos (2014) e Lucinda et al (2013).

Após a introdução, a segunda seção faz uma revisão da literatura, buscando elucidar o desenvolvimento dos modelos de escolha discreta e as contribuições da Economia Comportamental à discussão. A seção seguinte discorre brevemente sobre a obtenção de dados. A quarta seção busca elucidar questões relacionadas à metodologia, principalmente às alterações propostas por Swait (2009) e as implicações para a estimação. Os resultados são discutidos no penúltimo tópico. Por fim, uma última seção é reservada às conclusões, com as considerações finais e potenciais aprimoramentos do trabalho.

3.2 REVISÃO DE LITERATURA

Uma das principais ferramentas estatísticas para a compreensão do processo comportamental que leva à escolha do agente é a modelagem de escolha discreta. Tais modelos geralmente partem da hipótese de que o tomador de decisão busca maximizar sua utilidade (TRAIN, 2003), seguindo o modelo padrão da Teoria Econômica, com preferências inatas e estáveis⁴ (MCFADDEN, 2001).

Os principais modelos desenvolvidos para tal compreensão partiram do artigo seminal de Thurstone (1927). Sua formulação sobre a percepção de estímulos

⁴ Tal abordagem é referida no presente artigo com os adjetivos “padrão” e “clássico”, seguindo a nomenclatura utilizada por McFadden (2001).

foi adaptada por Marschak (1960) para um modelo de escolha econômica que, por incluir elementos aleatórios, passou a ser chamado de modelo de Maximização de Utilidade Aleatória (*Random Utility Maximization* - RUM).

Tal modelagem compreende que existem fatores que coletivamente determinam a escolha dos agentes, mas que nem todos são observados pelo pesquisador. Aqueles que são, são nomeados x , enquanto os fatores não observados são nomeados ε . O processo comportamental, dessa forma, é expresso por uma função do tipo $y = h(x, \varepsilon)$ que, contendo termos não observados, não pode prever y com exatidão. Ao invés disso, uma probabilidade de resultado é então derivada a partir da consideração de que os termos não observados são aleatórios com densidade $f(\varepsilon)$. Assim, a probabilidade de um tal resultado, dados os fatores observados, pode ser expressa por: $P(y | x) = \text{Prob}(\varepsilon \text{ s. a. } h(x, \varepsilon) = y)$.

Utilizando-se do arcabouço de maximização de utilidade, pode-se reexpressar a modelagem acima, assumindo que a utilidade representativa, V_{nj} , é determinada pelos atributos das J alternativas encontradas pelos N agentes, x_{nj} , assim como pelas características destes agentes, s_n : $V_{nj} = (x_{nj}, s_n) \forall j$. A utilidade representativa, no entanto, difere da utilidade real, U_{nj} , que possui um componente aleatório: $U_{nj} = V_{nj} + \varepsilon_{nj}$. Desta forma, a probabilidade do indivíduo n escolher a alternativa i em detrimento às outras alternativas é a probabilidade de que a utilidade obtida por tal indivíduo proveniente da escolha de i seja maior do que as demais, tal como $P_{ni} = \text{Prob}(U_{ni} > U_{nj}, \forall j \neq i)$ (TRAIN, 2003).

Swait (2009) aponta que os modelos de escolha discreta são geralmente obtidos a partir de diferentes hipóteses quanto à função de densidade contínua para o vetor de fatores não observados ε . O modelo Logit, o mais amplamente utilizado para este tipo de modelagem, assume que cada ε_{nj} é um valor extremo independente e identicamente distribuído de acordo com uma distribuição Gumbel, $f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} \cdot e^{-e^{-\varepsilon_{nj}}}$, de modo que a função de densidade cumulativa pode ser expressa por $F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$. Com esta distribuição de probabilidade, McFadden (1974) obteve a função de probabilidade de escolha Logit, $P_{ni} = \text{Prob}(U_{ni} > U_{nj} \forall j \neq i) = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$.

Uma importante limitação do modelo Logit decorre justamente das hipóteses sobre as características de probabilidades de escolha das quais originalmente Luce

(1959) derivou sua formulação. Tais hipóteses constituem a propriedade de independência das alternativas irrelevantes (*independence from irrelevant alternatives* - IIA), segundo a qual, para duas alternativas i e k , a razão de probabilidade Logit não depende de outras alternativas que não i e k . Tal propriedade não apenas restringe os padrões de substituição possíveis do modelo, como também implica em erro de previsão sempre que a razão de probabilidades entre duas alternativas mudar com a introdução ou alteração de uma outra alternativa (TRAIN, 2003).

O Logit apresenta duas outras importantes limitações, além desta decorrente da propriedade IIA. Embora possa representar variação de preferência sistemática, é incapaz de fazê-lo para variações aleatórias. Tal modelagem, ainda, não pode ser usada com dados em painel quando fatores observados são correlacionados ao longo do tempo para cada tomador de decisão. (TRAIN, 2003)

Buscando superar tais limitações, diversos modelos foram desenvolvidos, assumindo diferentes funções de densidade contínua para o vetor de fatores não observados ε . A primeira limitação, decorrente da propriedade IIA, é superada com a utilização de modelos de valor extremo generalizado (*General Extreme Value* - GEV). Essa distribuição permite correlações entre alternativas e, como seu nome indica, é uma generalização da distribuição de valor extremo univariado que é utilizado para modelos Logit padrão. O principal modelo da família GEV é o Logit Aninhado, que agrupa as alternativas de modo a necessitar que a propriedade IIA seja válida apenas dentro destes "ninhos", flexibilizando os padrões de substituição. Outros modelos GEV também foram derivados, como o Logit Combinatorial Pareado, o Logit Aninhado Generalizado e o Logit Heterocedástico (TRAIN, 2003).

Ainda que superem uma das limitações do Logit, os modelos GEV ainda são incapazes de representar variações aleatórias de preferência e de lidar com fatores observados correlacionados no tempo. Os modelos Probit são capazes de superar as três principais limitações apresentadas, sendo sua única limitação a necessidade de distribuições normais para todos os componentes não observados da utilidade.

Superando também esta última limitação, foram desenvolvidos os modelos Logit Mistos que, sendo definidos sobre a base da forma funcional de suas probabilidades de escolha, permitem diferentes especificações comportamentais. As probabilidades, para este modelo, são as integrais de probabilidades Logit Padrão multiplicadas por uma função de densidade dos parâmetros.

Toda a modelagem acima descrita, no entanto, é desenvolvida sob as hipóteses da economia clássica de maximização de utilidade, como já pontuado. McFadden (2001), no entanto, aponta que a evidência experimental dá suporte à visão de que outros fatores, como regras heurísticas, seriam os guias aproximados da maior parte do comportamento humano. O autor cita o psicólogo Prelec (1991, tradução nossa) para elucidar sobre a distinção entre as diferentes visões acerca da tomada de decisão:

Análise de decisão, que codifica o modelo racional, vê escolha como um problema fundamentalmente técnico de escolher o curso de ação que maximiza um critério unidimensional, a utilidade. A atividade mental primária é a redução de múltiplos atributos ou dimensões a apenas um, através da especificação de valores de tradeoffs. Para a ação governada por regras, o problema fundamental da decisão é aquele quasi-legal de construir uma interpretação satisfatória da situação de escolha. A atividade mental primária envolvida nesse processo é a exploração de analogias e distinções em que uma regra ou princípio únicos se aplicam de forma não ambígua. (...) A proposta das regras deve ser derivada de alguma fraqueza de nosso sistema de contabilidade custo-benefício, e pode-se esperar encontrar regras se proliferando exatamente nesses domínios de escolha onde o utilitarianismo natural não produz resultados satisfatórios.

Simon (1955) corrobora com a interpretação de que outras formas que não a racionalidade global⁵ do homem econômico sejam utilizadas durante o processo decisório. Segundo o autor (tradução nossa), "devido aos limites psicológicos do organismo (particularmente com respeito a habilidade computacional e preditivos), o esforço de racionalidade humano pode, no máximo, ser uma aproximação extremamente crua e simplificada para o tipo de racionalidade global que é implicada, por exemplo, pelos modelos de teoria dos jogos" (SIMON, 1955). Assim, aponta a necessidade de substituir tal racionalidade global "por um tipo de comportamento racional que seja compatível com o acesso à informação e às capacidades computacionais que são de fato possuídas pelos organismos, incluindo o homem, nos tipos de ambientes em que tais organismos existem" (SIMON, 1955).

⁵ Em se tratando do homem econômico e da racionalidade global, o presente artigo faz referência à pontuação de Simon (1955): "A teoria econômica tradicional postula um "homem econômico" que, no sentido de ser "econômico", é também racional. A esse homem é assumido ter conhecimento dos aspectos relevantes de seu ambiente que, se não absolutamente completo, é ao menos impressionantemente claro e volumoso. Ele é assumido também ter um sistema de preferências bem-organizado e estável, e uma habilidade em computação que o permite calcular, para os cursos alternativos de ação que a ele estão disponíveis, quais desses o permitirão alcançar o ponto atingível mais alto em sua escala de preferência."

Ponto central deste artigo de Simon (1955) é a apresentação da heurística de *satisficing*, estratégia em que o agente, ao invés de avaliar todo o conjunto de possibilidades, considera uma alternativa de cada vez, escolhendo a primeira alternativa que apresentar valores que superem um valor de corte predefinido. Se nenhuma alternativa passar em todos os cortes, eles podem ser relaxados e o processo repetido, ou uma alternativa pode ser aleatoriamente selecionada (PAYNE et al, 1993).

Payne et al. (1993, tradução nossa) aponta que, além da heurística de *satisficing*, "o mesmo indivíduo usará muitos tipos diferentes de estratégias em tomada de decisão, contingente sobre alguns fatores, como a forma em que informação é apresentada, a natureza da resposta, e a complexidade do problema". Assim, elenca algumas outras heurísticas relevantes, como a lexicográfica e a de eliminação-por-aspectos. O autor ainda faz uma distinção entre heurísticas compensatórias e não-compensatórias - as primeiras fazendo tradeoffs, enquanto as últimas evitam tais "conflitos".

Tal distinção de Payne et al. (1993) faz paralelo a outra crítica à abordagem clássica de tomada de decisão, realizada por Katona (1953). Este autor, buscando trazer possíveis contribuições da psicologia à análise da racionalidade, sugere formular as diferenças entre formação de hábitos e solução de problemas:

O princípio básico da primeira forma de comportamento é repetição. Aqui o argumento de Guthrie vale: "A mais certa e dependente informação com relação ao que um homem fará em qualquer situação é a informação sobre o que ele fez naquela situação em sua última ocorrência". Essa forma de comportamento depende da frequência de repetição assim como da proximidade temporal e do sucesso das performances anteriores. (...) Hábitos então formados são em alguma extensão automáticos e inflexíveis. Em contraste, comportamento solucionador de problemas tem sido caracterizado pelo aparecimento de um problema ou questão, pela deliberação que envolve reorganização e "direção", pela compreensão dos requerimentos da situação, pela ponderação de alternativa e por levar em consideração e, finalmente, por realizar a escolha entre cursos alternativos de ação. (KATONA, 1993, tradução nossa)

Adamowicz et al. (2008, tradução nossa), no entanto, são bastante diretos em sua oposição ao modelo clássico de maximização:

Nós não precisamos de estudos acadêmicos para saber que processos de escolha variam entre pessoas e situações. Enquanto um consumidor pode fazer escolhas comparando cuidadosamente os prós e contras de opções sobre atributos diferentes (uma regra compensatória), outro pode tomar a mesma decisão escolhendo aquele que é melhor pelo atributo mais

importante (uma regra não compensatória). (...) Finalmente, gostos e processos decisórios podem ser condicionados por fatores situacionais tais como tamanho do conjunto de escolha e variação de atributo.

Tendo em vista as críticas apresentadas ao arcabouço econômico clássico de maximização de utilidade, observa-se a possibilidade de refinamento dos modelos de escolha discreta, incorporando às modelagens contribuições da Psicologia e da Economia Comportamental. Adamowicz et al. (2008) apontam que, neste sentido, três visões sobre a forma como os indivíduos fazem escolhas estão subjacentes na pesquisa contemporânea de modelagem de escolha. Além da tradicional visão "econômica", que toma a perspectiva de que os consumidores fazem escolhas de forma consistente com a maximização de utilidade aleatória, outras duas são apontadas. Quanto a estas últimas visões, apontam que uma "visão é mais comportamental e psicológica e argumenta que aqueles processos de escolha real podem carregar pouca semelhança com os processos racionais que economistas assumem" (ADAMOWICZ et al, 2008). Nessa visão, se preferências sequer existem, elas são grumosas e imprecisas; além disso, as escolhas resultariam de regras heurísticas únicas associadas com o aparecimento externo de opções em conjuntos de escolha. Uma última visão, por sua vez, foca primariamente em formas estatísticas de modelos de resultados discretos (nesse caso, escolhas). Os adeptos desta visão, segundo os autores, "agem como se eles fossem ideologicamente neutros com relação aos processos de preferência e escolha. Ou seja, eles tendem a ver escolhas simplesmente como "dados"" (ADAMOWICZ et al, 2008).

Neste sentido, diversas modelagens foram desenvolvidas a fim de incluir nos modelos de escolha discreta tradicionais as contribuições referidas. Com relação a uma delas, relacionada ao conjunto de escolha, Swait (1984) aponta a importância da consideração adequada deste para evitar vieses na estimação dos parâmetros. Ainda que Simon (1955) considere esta adequação como pertencente ao arcabouço clássico, pode-se relacioná-la a heurísticas de formação de conjuntos de escolha, como aquelas de *screening* conjuntivo ou disjuntivo⁶. Outro exemplo de tentativa de aperfeiçoamento refere-se à utilização de combinação de dados de preferência

⁶ *Screening* conjuntivo refere-se à situação em que o indivíduo estabelece sua cesta de escolha com elementos que satisfaçam um conjunto de requisitos. *Screening* disjuntivo, por sua vez, refere-se àquela em que a cesta é formada por elementos que satisfaçam algum desses requisitos.

revelada, comum para análises de escolha discreta, e dados de preferência declarada (DISSANAYAKE E MORIKAWA (2010); MORIKAWA (1989)).

Swait (2009) propõe um modelo que, ao invés de alterar a função de densidade de probabilidade contínua a descrever os termos de erro, considera para estes uma função de densidade de probabilidade mista, incluindo uma parcela contínua e outra discreta. Tal modelagem, nomeada *k-Mix*, incorpora importantes avanços. Primeiramente, formaliza uma forma de modelar heterogeneidade de regra de decisão em uma população, enfrentando o problema de que os componentes geralmente incluídos em modelos de escolha podem ser viesados pelas omissões estruturais de especificações particulares. Além disto, permite que o modelo capte tanto comportamentos compensatórios, maximizadores de utilidade, quanto não-compensatórios, como heurísticas lexicográficas e de *satisficing*. O modelo, ainda, estende o conceito de formação de conjunto de escolha. É importante ressaltar que a abordagem de Swait não busca responder a uma demanda específica da Economia Comportamental, mas oferece um ferramental apto a lidar com formulações apresentadas por diversos autores e correntes.

Swait (2009) ainda exemplifica o potencial de sua modelagem realizando uma análise de escolha para modais de transporte na Região Metropolitana de São Paulo, aplicando-a para a base de dados da Pesquisa Origem e Destino de São Paulo de 1987. O autor compara esta aplicação à do Logit Multinomial, obtendo resultado significativamente melhores em favor da primeira. Utilizando-se das modelagens tradicionais que se baseiam nos pressupostos clássicos, Barcellos (2014) e Lucinda et al (2013) fazem análises de dados mais recentes para o caso paulistano, aplicando o Logit Multinomial e o Logit Misto - portanto, a modelagem tradicional. Com o exposto, segue-se à apresentação da metodologia referente à abordagem proposta: a aplicação da modelagem proposta por Swait (2009) aos dados mais recentes da Pesquisa Origem e Destino de 2007 da Região Metropolitana de São Paulo.

3.3 METODOLOGIA

Para a realização do presente trabalho, partiu-se do Banco de Dados da Pesquisa Origem e Destino de 2007 para a Região Metropolitana de São Paulo, realizada pela Companhia do Metropolitano de São Paulo (Metrô). Tal pesquisa,

composta por 168.582 observações, tem a característica de ser domiciliar e aleatória, coletando tanto dados socioeconômicos quanto referentes às viagens realizadas.

Foram examinadas, no presente artigo, as escolhas relacionadas às viagens a trabalho no horário de pico da manhã (das 6h às 9h59). Desta forma, tal recorte, análogo àquele realizado por Swait e Ben-Akiva (1987), apresentou 25.995 observações. Quanto às opções de escolha, foram agregadas nas seguintes alternativas com referência ao principal modal utilizado no trajeto, seguindo o estipulado por Barcellos (2014) e Lucinda et al (2013):

- (1) Carro,
- (2) Motocicleta,
- (3) Ônibus,
- (4) Metrô e Trem,
- (5) Táxi,
- (6) Outros.

No entanto, como pesquisa de preferência revelada, a Pesquisa Origem e Destino é carente das informações das alternativas não escolhidas. A fim de complementar a base de dados com tais informações, seguiu-se o realizado por Barcellos (2014) e Lucinda et al (2013): a partir de informações sobre as alternativas escolhidas e utilizando de uma regressão linear, foram estimados os dados faltantes. A exceção refere-se ao custo de algumas alternativas que pôde ser determinado através de informações externas, como de ônibus, metrô e trem.

O modelo de regressão utilizado para tal fim fez uso das escolhas observadas para cada modal, sendo as variáveis dependentes os logaritmos do custo e da duração da viagem. Já as variáveis independentes utilizadas foram a distância em quilômetros entre as zonas de origem e destino⁷ e variáveis *dummy* para a hora de partida e de chegada e para motivação da viagem. A partir de tal modelo, estimou-se coeficientes que, por sua vez, foram utilizados para a obtenção do tempo e do custo esperados para as alternativas não escolhidas.

É importante ressaltar que, a fim de minimizar problemas relativos à disponibilidade de cada modal para cada tomador de decisão, a disponibilidade de

⁷ A Pesquisa Origem e Destino da Região Metropolitana de São Paulo de 2007 divide a região em 460 zonas.

escolha de trem e metrô foi restrita apenas para as zonas que tivessem algum respondente escolhendo um destes modais. Tal procedimento foi o mesmo realizado por Barcellos (2014) e Lucinda et al (2013).

Uma vez especificada a base de dados, resta a apresentação do modelo de escolha discreta aqui utilizado. Embora Swait (2009) apresente uma modelagem aberta para modelos *k-Mix*, o presente trabalho abordará apenas o modelo *2-Mix*, com duas condições: a de *Tradeoff* e a de Rejeição. Desta forma, a utilidade do indivíduo deixa de ser formada apenas pela soma da utilidade sistemática com os fatores não-observados. Assumindo G como o conjunto universal de alternativas, a utilidade de cada alternativa $i \in G$ passa a ser expressa por:

$$U_i = V_i + \varepsilon_i, \text{ com probabilidade } q_i,$$

$$\text{ou } U_i = -\infty, \text{ com probabilidade } p_i.$$

Nesta notação todos os valores são determinados, com exceção das probabilidades p_i e q_i ($p_i + q_i = 1$), que referem-se, respectivamente, à Condição de *Tradeoff* e de Rejeição. Assim, a utilidade será, sob certa probabilidade, aquela derivada do processo compensatório e, com probabilidade complementar, aquela derivada do processo não-compensatório. Desta maneira, a função de densidade de probabilidade mista pode ser expressa por:

$$f_i(U_i) = q_i g_i(V_i, \varepsilon_i) + (1 - q_i) h_i(-\infty) \forall i \in G,$$

onde $g()$ é a função de densidade de probabilidade do termo de erro na Condição de *Tradeoff* e $h()$ é a função de massa de probabilidade para a Condição de Rejeição.

É importante ressaltar aqui uma limitação desta modelagem: a hipótese de que as probabilidades das condições são independentes das utilidades estocásticas ε . Com ela, assume-se que os mecanismos para adoção de modos de decisão diferentes são probabilisticamente independentes dos fatores não-observados na função de utilidade.

Com as referidas alterações na função utilidade, e conseqüentemente na função de densidade de probabilidade mista, a probabilidade de escolha também é alterada. Assim, sendo Γ_i o conjunto de subconjuntos de G que contêm a alternativa i , a expressão geral para a probabilidade de escolha de $i \in G$ é dada por

$$P_i = \frac{\prod_{k \in G} (1 - q_k)}{|G|} + \sum_{C \in \Gamma_i} Q(C) H_{(C)}(V_C, \varepsilon_C),$$

onde

$$Q(C) = \left(\prod_{j \in C} q_j \right) \left(\prod_{j \notin C} (1 - q_j) \right), C \in \Gamma_i$$

é a probabilidade de que todas as alternativas em C tenham utilidades que estejam na Condição de *Tradeoff* e que todas as alternativas excluídas de C tenham utilidades na Condição de Rejeição. $H_{(C)}(V_C, \varepsilon_C)$ é a probabilidade condicional de i para as alternativas em C , enquanto V_C e ε_C são vetores de utilidades sistemática e estocástica, respectivamente, das alternativas em C . Assumindo que os elementos de ε_C sejam independentes e identicamente distribuídos como Gumbel com uma escala unitária, como no modelo Logit padrão, a probabilidade de escolha aparece como:

$$P_i = \frac{\prod_{k \in G} (1 - q_k)}{|G|} + \sum_{C \in \Gamma_i} Q(C) \left(\frac{\exp(V_i)}{\sum_{j \in C} \exp(V_j)} \right).$$

O primeiro termo nas expressões para as probabilidades surge da condição em que todas as alternativas estejam na Condição de Rejeição, caso em que a escolha é aleatória e as probabilidades todas iguais a $|G|^{-1}$. O segundo termo pondera as probabilidades de escolha condicional da alternativa i sobre todos os possíveis subconjuntos contendo-o; cada subconjunto C representa um grupo de alternativas que estão na Condição de *Tradeoff*, sendo que todas as alternativas fora de C estão na Condição de Rejeição (SWAIT, 2009). É importante observar, através desta notação, que o Modelo Logit Multinomial está, portanto, contido nos Modelos k-Mix.

Para obter os estimadores de interesse, utiliza-se, conforme o aplicado ao Logit padrão por Train (2003), da função de máxima verossimilhança, da forma:

$$L(\beta, q_i) = \prod_{n=1}^N \prod_i (P_{ni})^{y_{ni}},$$

onde são vetores contendo os parâmetros do modelo. A função de log-verossimilhança pode ser expressa por

$$LL(\beta, q_i) = \sum_{n=1}^N \sum_i y_{ni} \ln P_{ni}$$

sendo os estimadores os valores de que maximizam essa função. No entanto, uma vez que foi alterada a probabilidade de escolha, as derivações

subsequentes para a maximização também são alteradas. Desta forma, tanto a derivada quanto a β quanto a q_i devem igualar a zero:

$$\frac{dLL(\beta, q_i)}{d\beta} = \frac{dLL(\beta, q_i)}{dq_i} = 0.$$

É importante ressaltar que a estimação foi realizada indiretamente para q_i , estimando-se δ_i como em: $q_i = [1 + (-\delta_i)]^{-1}$. Tal procedimento foi realizado para tornar as manipulações algébricas mais tratáveis.⁸

Para efeitos de comparação, um modelo Logit Multinomial é estimado com a mesma base de dados.^{9,10}

3.4 RESULTADOS

Os resultados de ambas as estimações são encontrados na Tabela 8 e utilizam “Outros” como alternativa de referência. Quanto aos parâmetros usuais, observa-se que, dentre aqueles relacionados às variáveis específicas das alternativas, os coeficientes de custo, da iteração entre custo e renda e entre tempo e renda apresentaram significância estatística e os sinais adequados à intuição em ambas as estimações. Chama a atenção, no entanto, o coeficiente relacionado ao tempo por apresentar sinal positivo e, assim, indicar que quanto maior o tempo despendido no trajeto, maior a utilidade obtida pelos indivíduos. Mantido nas duas estimações, tal imprecisão não advém do modelo de escolha discreta, mas potencialmente tem origem no modelo utilizado para a complementação da base de dados, uma vez que tal abordagem carrega certo viés de seleção amostral.

⁸ A partir de tais valores, é possível estimar a probabilidade de que cada modal esteja na zona de tradeoff, tal como será apontado na Tabela 9.

⁹ Os resultados das derivações, não sendo triviais, não serão apresentados neste artigo. O método de estimação utilizado para a estimação foi o BHHH (Berndt–Hall–Hall–Hausman), uma vez que Newton-Raphson demandaria excessivos processamento e tempo em decorrência da utilização da matriz Hessiana como parte do processo. O método BHHH, por sua vez, estimando a partir do produto externo dos escores, evita o cálculo de todas as segundas derivadas (TRAIN, 2003).

¹⁰ Toda a modelagem foi aplicada utilizando o software R, partindo do pacote mlogit (CROISSANT, 2012).

Quanto às variáveis sociodemográficas relacionadas aos modais, observa-se uma consonância geral entre os dois modelos. Algumas diferenças, porém, chamam a atenção. Primeiramente, observa-se que a variável renda relacionada ao ônibus, que apresenta sinal positivo no modelo Logit Multinomial, passa a apresentar sinal negativo no modelo 2-Mix. O resultado do último parece concordar com a intuição: conforme aumenta a renda, menor é a probabilidade de o indivíduo escolher este modal. Ainda quanto ao ônibus, é relevante a diferença de sinais apresentada pelo coeficiente relacionado ao emprego formal. Neste caso, o valor, positivo, encontrado pelo Logit é mais razoável que aquele obtido com uso do 2-Mix.

Outra diferença a ser notada é a perda de significância estatística da variável emprego formal tanto para trem e metrô, quanto para carro, no modelo 2-Mix. Com relação ao carro, porém, observa-se um sinal mais adequado para a *dummy* estudante - enquanto apresenta sinal positivo no modelo Logit Multinomial, ele aparece negativo na modelagem aqui proposta.

Encerrando a apresentação das diferenças entre as variáveis sociodemográficas, observa-se a perda de significância estatística dos coeficientes de renda e tamanho de família para o modal táxi, no modelo 2-Mix. Em oposição, vê-se que a variável idade, com uso desta modelagem, passa a ser estatisticamente significativa e a apresentar sinal negativo. Quanto às variáveis relacionadas ao uso de motocicletas, apenas variações de magnitude são observadas.

TABELA 8 - COMPARAÇÃO ENTRE ESTIMAÇÕES DOS MODELOS 2-MIX E LOGIT MULTINOMIAL

	Modelo 2-Mix	Modelo Logit Multinomial
Variáveis Específicas das Alternativas		
Custo	-0.0898 ***	-0.1287 ***
Tempo	2.8100 ***	2.0229 ***
Custo X Renda	0.0619 ***	0.0143 ***
Tempo X Renda	-0.2834 ***	-0.3082 ***
Ônibus		
Renda (R\$1000)	-0.0474 *	0.0663 ***

Idade	-0.0198 ***	-0.0120 ***
Mulher	-0.0399	-0.0041
Estudante	0.3305 ***	0.1957 ***
Tamanho Família	-0.1147 ***	-0.0634 ***
Emprego Formal	-0.2232 ***	0.4343 ***
Trem e Metrô		
Renda (R\$1000)	0.1831 ***	0.2819 ***
Idade	-0.0384 ***	-0.0261 ***
Mulher	-0.4223 ***	-0.2894 ***
Estudante	0.1325 **	0.0785 *
Tamanho Família	-0.1841 ***	-0.1201 ***
Emprego Formal	-0.0128	0.3112 ***
Carro		
Renda (R\$1000)	0.4317 ***	0.3822 ***
Idade	0.0151 ***	0.0104 ***
Mulher	-1.4795 ***	-1.1802 ***
Estudante	-0.2138 ***	0.2521 ***
Tamanho Família	-0.2510 ***	-0.1656 ***
Emprego Formal	-0.0600	0.4286 ***
Moto		
Renda (R\$1000)	0.2309 ***	0.2386 ***
Idade	-0.0810 ***	-0.0532 ***
Mulher	-2.6962 ***	-2.6767 ***
Estudante	-0.3242 ***	-0.1965 ***

Tamanho Família	-0.1966 ***	-0.0861 ***
Emprego Formal	1.8596 ***	0.7708 ***
Táxi		
Renda (R\$1000)	-2.7059	0.2160 ***
Idade	-1.7835 **	-0.0092
Mulher	0.4350	-0.3800
Estudante	1.7732	-0.1456
Tamanho Família	3.4084	-1.1675 ***
Emprego Formal	2.2587	-0.2484
δ_i: Parâmetros de Condição de <i>Tradeoff</i>		
Outros	0.5561 ***	-
Ônibus	1.8647 ***	-
Trem e Metrô	1.1105 ***	-
Carro	1.1107 ***	-
Moto	-0.0382	-
Táxi	-1.8665 ***	-
Log-Verossimilhança (Convergência)	-32919	-33630

FONTE: elaboração própria.

p < 0.1: *; p < 0.05: **; p < 0.01 ***

Realizada a avaliação sobre as variáveis usuais, pode-se seguir à análise dos parâmetros de Condição de *Tradeoff*. A fim de auxiliar na sua interpretação, foi construída a Tabela 9, que apresenta, partindo de tais parâmetros, a probabilidade de que cada modal esteja na condição de *tradeoff*. O modal mais presente nesta condição e, portanto, menos rejeitado, aparece como sendo o ônibus, incluso em 86.58% das cestas de escolha. Tal constatação parece corroborar com a intuição, uma vez que os ônibus são a forma de transporte público mais disseminada pela

Região Metropolitana de São Paulo, ao mesmo tempo que não apresenta durações de trajeto tão exorbitantes quanto seriam para deslocamentos a pé, por exemplo.

Os próximos modais com maior presença na condição de *tradeoff* são trem e metrô e carro, com aproximadamente 75%. É razoável que apresentem valores menores do que aqueles encontrados para ônibus, uma vez que são menos disponíveis à população em geral: enquanto trem e metrô apresentam poucas linhas pela Região Metropolitana, o alto custo dos carros impede seu amplo acesso. Tais modais são seguidos, no ordenamento de probabilidade de presença nesta zona de comparação, pelo agregado “Outros” que, embora seja disponível a todos, pode apresentar tempos de percurso muitas vezes não factíveis.

Por fim, os modais que apresentam maior rejeição são moto e táxi. Embora o parâmetro relacionado ao primeiro não apresente significância estatística, é razoável acreditar que, mesmo não sendo um modal financeiramente inviável, os riscos de sua utilização na Região Metropolitana de São Paulo o torna pouco atrativo para estar presente na condição de *tradeoff*. Último da lista, o modal táxi aparece com uma probabilidade de apenas 13,39% de estar nesta região. Para o ano de 2007, tal constatação parece corroborar com a realidade: sendo uma alternativa cara, com algumas questões quanto à segurança e de acesso razoavelmente difícil, poucos de fato cogitavam, à época, a possibilidade de se deslocar utilizando-o.

TABELA 9 - PROBABILIDADE DE ESTAR NA CONDIÇÃO DE TRADEOFF POR MODAL DE TRANSPORTE

Modal de Transporte	Probabilidade de estar na Condição de <i>Tradeoff</i>
Outros	63.55%
Ônibus	86.58%
Trem e Metrô	75.22%
Carro	75.23%
Moto	49.05%
Táxi	13.39%

FONTE: elaboração própria.

A diferença mais relevante, no entanto, refere-se à capacidade explicativa dos modelos comparados. Enquanto o Logit Multinomial apresenta um valor para log-verossimilhança de convergência de -33.630, o modelo 2-Mix apresenta o valor de -32.919. Uma vez que o modelo Logit Multinomial encontra-se aninhado ao modelo 2-

Mix, é possível verificar se este é estatisticamente superior àquele utilizando um teste de qui-quadrado. Observando-se que a estatística teste assume o valor de 1422, com 6 graus de liberdade, verifica-se que este é muito superior que o valor crítico a 0.05. Desta forma, conclui-se que restringir os parâmetros de condição de *tradeoff* para se obter um modelo Logit Multinomial pode ser uma alternativa seguramente rejeitada - conclusão que corrobora com o observado por Swait (2009).

3.5 CONSIDERAÇÕES FINAIS

O presente artigo realiza a aplicação, à mais recente Pesquisa Origem e Destino para Região Metropolitana de São Paulo, realizada em 2007, de um modelo de escolha discreta com maior fundamentação nas contribuições da Economia Comportamental. Nesse sentido, é extremamente relevante o fato de tal modelo apresentar melhores resultados quanto às estatística-teste, corroborando com os resultados encontrados por Joffre Swait (2009), que utilizou a base de dados da Pesquisa Origem e Destino da Região Metropolitana de São Paulo de 1987.

Esta melhor adequação e capacidade explicativa apresentada pelo modelo 2-Mix sugere a importância da consideração de comportamentos não-maximizadores na análise de escolha de modal e avaliação de políticas públicas de trânsito. Também reforça a importância da avaliação da cesta de escolhas, evitando que parâmetros sejam viesados por cestas que incorporam alternativas desprezadas *a priori*. Tais melhorias indicam que, utilizando da abordagem proposta por Swait (2009) – e, portanto, buscando embasamento na Economia Comportamental -, modelos com maior capacidade preditiva podem ser obtidos e, assim, políticas públicas mais ajustadas podem ser realizadas.

Por fim, com os resultados positivos obtidos, sugere-se também o aprofundamento da investigação relacionada a tais comportamentos, ampliando a modelagem com a incorporação da Condição de Dominância, como no modelo 3-Mix (SWAIT, 2009). A pesquisa também deve ser avançada buscando investigar os atributos que expliquem a inclusão ou não de um modal na Condição de *Tradeoff*, assim como explorar as regras heurísticas a fundamentar tal decisão.

4 ARTIGO 3: DINÂMICA COEVOLUTIVA NA ANÁLISE DE POLÍTICA DE TRANSPORTES

Resumo

Os modelos de escolha discreta são amplamente utilizados como ferramentas para análise de políticas de transporte. Tais modelos, no entanto, apresentam potenciais limitações para estas avaliações. Este artigo busca explorá-las, com especial enfoque em seu caráter estático. Para tanto, desenvolve modelos baseados em agentes simplificados, discutindo questões também relacionadas à formação de expectativas. Os resultados obtidos permitem a observação de dinâmica coevolutiva e propriedades emergentes. Conclui-se sobre a importância desta modelagem para explicação do fenômeno e elucidação de sua dinâmica, evidenciando as possibilidades de complementaridade entre os ferramentais.

4.1 INTRODUÇÃO

O trânsito afeta a vida de todos de forma direta. Como exemplo, pode-se citar que, em 2015, cada habitante da cidade de São Paulo gastou diariamente 2h38 em deslocamentos pela cidade, em média (IBOPE, 2015). Sendo vital para o funcionamento de toda a economia, os impactos indiretos do trânsito são ainda maiores, como se observa pelo fato de o PIB brasileiro perder, no longo prazo, cerca de R\$156 bilhões em decorrência da morosidade do trânsito apenas da referida cidade paulista (HADDAD, 2015).

Com tamanho problema, são extremamente necessárias políticas públicas eficientes, a fim de reduzir o tempo (e o dinheiro) perdido. Tomar as medidas adequadas, porém, não é trivial. Nagel e Rasmussen (1994) apontam que elas podem até mesmo ter consequências opostas às intenções - e citam como exemplo

o paradoxo de Braess, em que a adição de uma nova rua reduz a capacidade geral da malha viária, ao invés aumentá-la¹¹.

Para melhor análise de políticas públicas de transporte, é essencial a compreensão da escolha de modal. Entre as principais ferramentas utilizadas por economistas para tal fim estão os ditos modelos de escolha discreta, amplamente utilizados internacionalmente (KOPPELMAN; BHAT, 2006; DISSANAYAKE; MORIKAWA, 2010) e cuja crescente utilização no país é evidente nos últimos anos (LUCINDA; MEYER; LEDO, 2013; LUCINDA et al, 2015, PACHECO; CHAGAS, 2016). Dentre alguns dos problemas deste ferramental, é relevante a superestimação da utilização de um novo modal apresentada por alguns destes modelos - de acordo com Bhat (1995), aplicações de Logit Multinomial podem ser otimistas quanto à reação dos indivíduos às políticas propostas.

A principal limitação destes modelos de escolha discreta talvez seja seu caráter estático e incapaz de levar em conta as propriedades emergentes resultantes da interação entre os agentes e a dinâmica coevolutiva entre tais níveis de análise. Ainda que seja capaz de prever a redução inicial do número de usuários de automóveis em decorrência de alguma política pública, é incapaz, por exemplo, de considerar o impacto de tal redução sobre o tempo de trânsito destes. Tal incapacidade é mantida para avaliações subsequentes, tais como um novo aumento de usuários de automóveis em decorrência da queda do tempo de deslocamento através de tal modal. Esta abordagem não é capaz, pois, de avaliar dinâmicas em que estejam incluídas mudanças estruturais.

Assim, o presente artigo busca utilizar de modelagem baseada em agentes para explorar tal limitação, assim como avaliar sua capacidade de lidar com ela. Busca, pois, a partir de uma proposta generativista¹², explorar certos conjuntos de

¹¹ Cohen e Kelly (1990, tradução livre) apontam que tal paradoxo advém do fato de que, em um equilíbrio de Nash que não é Pareto-eficiente, “os indivíduos auto-interessados são incapazes de se abster de usar a capacidade adicional, mesmo que a sua utilização conduza à deterioração do tempo médio de trânsito”.

¹² De forma mais precisa, Epstein e Axtell (1997, tradução nossa) determinam o aspecto generativista através de seu objetivo: “prover as microespecificações iniciais (agentes iniciais, ambiente e regras) que sejam *suficientes para gerar* as macroestruturas de interesse.” Ao que continuam: “Nós consideramos que uma macroestrutura é “explicada” por uma dada microespecificação quando a suficiência generativa da última foi estabelecida. (...) nós interpretamos a questão, “você pode explicar isso?” como “você pode cultivar isso?”. De fato, *estamos propondo um programa generativista para as ciências sociais e vemos a sociedade artificial como seu principal instrumento científico.*”.

microespecificações que sejam suficientes para gerar tais macrofenômenos de interesse (EPSTEIN; AXTELL, 1996). Em outras palavras, o trabalho apresenta modelos construídos a partir de certos pressupostos microeconômicos que, com uso de simulação computacional, permitem observar e analisar resultados decorrentes da interação entre os agentes e sua dinâmica coevolutiva.

Após a introdução, a segunda seção explora algumas limitações do Logit Multinomial, aponta conceitos importantes para a análise e apresenta o modelo baseado em agentes mais básico a incluir dinâmica ao modelo de escolha discreta - permite, pois, que os indivíduos interajam no deslocamento e revejam suas escolhas de forma iterativa. A terceira seção divide-se em duas partes, que representam duas alterações com relação ao primeiro modelo baseado em agentes proposto: a inclusão de certa aleatoriedade na escolha, e de heterogeneidade nas funções de utilidade. Por fim, conclui-se com uma análise da importância da dinâmica coevolutiva para análise de políticas públicas de transporte, evidenciando a importância da modelagem proposta neste contexto.

4.2 MODELO SIMPLIFICADO: ADICIONANDO DINÂMICA AO LOGIT MULTINOMIAL

Dentre os modelos de escolha discreta para mais de duas alternativas, a mais simples especificação é aquela desenvolvida por McFadden (1974), o Logit Multinomial. Ainda que a mais simples delas, apresenta já uma limitação recorrente nos demais modelos: seu caráter estático. A presente seção busca apresentar brevemente esta modelagem e explicitar tal limitação. A forma proposta de explorar esta dificuldade é com um modelo baseado em agentes com pequenas variações com relação ao Logit Multinomial no que se refere à avaliação de políticas públicas. Esta seção tratará, então, de apresentar a configuração básica de tal modelo, assim como discutir conceitos necessários à discussão. Por fim, é concluída com a apresentação dos resultados e comparação entre as avaliações.

4.2.1 Logit Multinomial

O modelo Logit Multinomial é uma ferramenta econométrica desenvolvida por McFadden (1974) que, quando aplicado para escolhas de modal em transporte,

busca estimar, a partir de uma base de dados, a probabilidade de um indivíduo escolher cada uma das alternativas possíveis.

De forma resumida, utilizando-se do arcabouço de maximização de utilidade, o Logit Multinomial assume que a *utilidade representativa*, V_{nj} , é determinada pelos atributos das J alternativas encontradas pelos N agentes, x_{nj} , assim como pelas características destes agentes, s_n : $V_{nj} = V(x_{nj}, s_n) \forall j$. A utilidade representativa, no entanto, difere da utilidade real, U_{nj} , que possui um componente aleatório: $U_{nj} = V_{nj} + \varepsilon_{nj}$. Desta forma, a probabilidade do indivíduo n escolher a alternativa i em detrimento às outras alternativas é a probabilidade de que a utilidade obtida por tal indivíduo proveniente da escolha de i seja maior do que as demais, tal como $P_i = Prob(U_{ni} > U_{nj}, \forall j \neq i) = Prob(\varepsilon_{nj} - \varepsilon_{ni} < V_{nj} - V_{ni}, \forall j \neq i)$. (TRAIN, 2003).

Mais importante para a comparação em questão, no entanto, é a compreensão da forma como o Logit Multinomial é utilizado para a avaliação de políticas públicas. Primeiramente, os coeficientes das variáveis explicativas, como de custo, são estimados. Assim, pode-se obter um coeficiente a medir quanto um choque em uma variável afeta a utilidade do indivíduo e, por conseguinte, a probabilidade de escolha da alternativa. Desta forma, a análise da política pública é realizada alterando o valor da variável em questão e avaliando quais alternativas apresentarão a maior utilidade para cada indivíduo - sendo, portanto, as novas escolhas estimadas.

Tal ferramental, porém, é capaz de prever apenas a primeira reação, imediata e de curtíssimo prazo, à política pública. É capaz, por exemplo, de prever uma redução do número de usuários de carros após a implementação de um pedágio urbano. Mas é incapaz, por sua vez, de lidar com a redução da duração das viagens destes em decorrência do menor adensamento de veículos, assim como de lidar com toda causalidade subsequente. Sendo um modelo estático, carece, para melhor avaliação de política pública, da adição de um aspecto dinâmico.

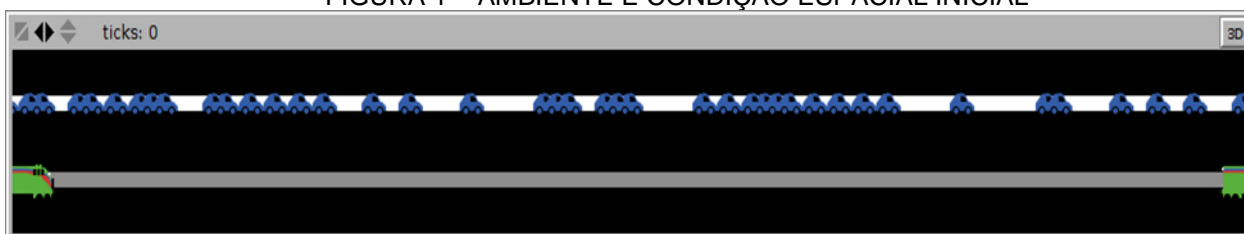
4.2.2 Modelo Baseado em Agentes: Configuração Básica

Com o objetivo de explorar a dinâmica coevolutiva não abarcada pelos modelos de escolha discreta, foi desenvolvido um modelo baseado em agentes simplificado, utilizando o software NetLogo. O modelo se resume em uma simulação

de trânsito em que os indivíduos devem optar, para a realização do necessário deslocamento, por um dos dois modais: carro ou metrô.

Primeiramente, é importante especificar o ambiente ocupado pelos agentes. Neste modelo simplificado, ele é composto por duas vias: a primeira, superior, representa uma rua com apenas uma faixa para circulação exclusiva de carros; a segunda, inferior, representa os trilhos pelos quais poderão circular um metrô. Ambas apresentam continuidade entre o fim e o começo, de modo que são representações de anéis. Cada via é representada, no modelo, por um espaço com dimensão na medida de 100 *patches*¹³ por 1 *patch*, como observado na Figura 1.

FIGURA 1 – AMBIENTE E CONDIÇÃO ESPACIAL INICIAL



Além da especificação do ambiente, é necessário especificar quais serão os agentes. Devido às características específicas deste modelo, no entanto, é importante antes apresentar como os agentes se relacionam com o ambiente - ou seja, como o metrô e os carros trafegam pelas vias.

O tráfego do metrô é bastante simples no modelo. Com início e fim de trajeto determinados, o metrô se desloca com velocidade constante e, assim, percorre o caminho sempre em um mesmo tempo.

A modelagem para o tráfego de carros, por sua vez, é menos trivial. Durante a década de 1990, uma série de estudos buscaram analisar a formação do congestionamento, investigando o tráfego como um fenômeno dinâmico de um sistema de muitas partículas (SUGIYAMA et al, 2008). Para tanto, modelos matemáticos e simulações (NAGEL; SCHRENKENBERG, 1992, NAGEL; RASMUSSEN, 1994, NAGEL, RASMUSSEN; BARRET, 1996, WILENSKY, 1997) foram desenvolvidos a fim de explorar as causas pelas quais “um sistema

¹³ *Patch* é como a medida básica de área dentro do ambiente NetLogo: cada quadrado “básico” do ambiente.

drasticamente altera seu aspecto macroscópico devido ao efeito do movimento coletivo de partículas que interagem” (SUGIYAMA, 2008).

A configuração aqui utilizada é inspirada diretamente no Modelo de Tráfego Básico de Wilensky (1997). Difere apenas na configuração inicial: no modelo de Wilensky, a velocidade inicial é aleatoriamente determinada entre 0.1 e 1, enquanto neste ela é igual a zero.

Inicialmente, os veículos são distribuídos aleatoriamente ao longo da via, com velocidade igual a zero, e aceleração e desaceleração exogenamente determinadas. Também é estabelecida de forma exógena uma velocidade máxima, sendo zero a velocidade mínima. O tráfego destes agentes pode ser expresso em 3 etapas, semelhante a Nagel e Schrenckenberg (1992)¹⁴:

- 1) Aceleração: se não houver outro carro a uma distância menor ou igual a 1 *patch* à frente do veículo, sua velocidade, v , passa a ser a velocidade anterior adicionada da aceleração exogenamente determinada, a , mas respeitando o limite de velocidade, $v_{máx}$.
- 2) Desaceleração: caso haja outro carro a uma distância menor ou igual a 1 *patch* à frente do veículo, a velocidade do carro, v , passa a ser a velocidade do carro à frente, v_{frente} , subtraída da desaceleração exogenamente determinada, d , respeitando o limite mínimo para que velocidade não seja inferior a zero, $v_{min} = 0$.
- 3) Movimento do carro: o veículo avança, neste passo, v *patches*, alterando sua posição, p .

O pseudocódigo para essa etapa pode ser expresso da seguinte forma:

Tráfego de carros:
1. SE (espaço entre carros $_{i,j} > 1$) ENTÃO $v_i = v_i + a$.
2. SE ($v_i > v_{máx}$) ENTÃO $v_i = v_{máx}$.
3. SE (espaço entre carros $_{i,j} \leq 1$) ENTÃO $v_i = v_j - d$.

¹⁴ Estes autores incluem uma etapa adicional, que representa, dentre outros fenômenos, uma redução de velocidade aleatória. A modelagem de Wilensky (1997) aqui utilizada, no entanto, abre mão de tal etapa, sem prejuízo para o modelo.

4. SE $(v_i < v_{min} = 0)$ ENTÃO $v_i = v_{min} = 0$.

5. $p_i = p_i + v_i$.

O resultado dos algoritmos simples apresentados para cada indivíduo que opta por utilizar carro como transporte é evidenciado na Figura 2. Tal figura representa uma situação ilustrativa do tráfego após algumas iterações. Os carros com a cor amarela apresentam baixa velocidade ($v < 0.25$), enquanto os de cor laranja apresentam média velocidade ($0.25 \leq v < 0.5$) e os vermelhos, alta ($v \geq 0.5$). Observa-se, portanto, um adensamento de veículos com menor velocidade: o congestionamento aparece como uma propriedade emergente do sistema.

FIGURA 2 – EMERGÊNCIA DE CONGESTIONAMENTO. OS CARROS AMARELOS, ADENSADOS, ESTÃO COM VELOCIDADE REDUZIDA, ($v < 0.25$); OS DE COR LARANJA, UM POUCO MAIS ESPAÇADOS, ESTÃO COM VELOCIDADE MÉDIA ($0.25 \leq v \leq 0.5$); OS VERMELHOS SÃO OS MAIS RÁPIDOS ($v > 0.5$)



Tal comportamento aqui obtido geralmente é assumido como dado em análises econômicas. Bruekner (2011), por exemplo, assume que a partir de uma quantidade determinada de carros, T^* , a entrada de um indivíduo adicional naquela via gera externalidades negativas para todos os demais indivíduos que a utilizam - aumento de tempo para todos em decorrência do congestionamento. Esta abordagem, no entanto, é razoável apenas em situações sem restrições a informação e/ou não se leve em conta cada situação particular de congestionamento encontrado.

4.2.3 Emergência

Além do congestionamento visualmente observável, tal fenômeno coletivo apresenta outros aspectos característicos, como transição de fase. Como apontam Epstein e Axtell (1996), nas ciências da complexidade, padrões agregados ou macroscópicos estáveis como estes podem ser chamados de estruturas

emergentes. No entanto, os autores pontuam que nenhuma teoria formal completamente satisfatória de emergência foi estabelecida, de modo que diversas discussões sobre a definição do termo ainda permanecem (WILENSKY; RAND, 2016).

Assim, embora alguns autores utilizem o termo “emergente” igualando-o a “surpreendente” ou “não esperado”, o termo é aqui utilizado da mesma forma que Epstein e Axtell (1996) utilizam, denotando padrões macroscópicos estáveis que decorrem de interações locais dos agentes. Tal definição é bastante semelhante à de Wilensky e Rand (2016), que a definem como o “aparecimento de estruturas, padrões e propriedades coerentes e novas através das interações de múltiplos elementos distribuídos”. Para estes autores, é importante ressaltar, estruturas emergentes não podem ser deduzidas exclusivamente das propriedades dos elementos, mas das interações entre eles.

Quanto ao tópico de interesse, Nagel e Rasmussen (1994) enfatizam que “todas as propriedades de performance que podemos nos interessar em um sistema de transporte (de fato, em qualquer sistema feito por humanos) são propriedades emergentes dos objetos que interagem no sistema. Elas não estão explicitamente representadas em nenhum lugar ao nível dos objetos que interagem. Elas são geradas através da dinâmica”. Como observado pela importância para a análise de congestionamentos, outras propriedades emergentes serão evidenciadas importantes, como o próprio padrão de escolha de modal.

Por fim, é importante ressaltar que as propriedades emergentes não são frutos apenas de simulações computacionais, mas refletem situações reais. SUGIYAMA et al (2008) realizaram experimentos reais de trânsito à forma dos modelos baseados em agentes e, referindo-se a eles, concluíram que “a emergência de um congestionamento sem gargalo provê uma prova do mecanismo essencial de formação de um congestionamento”, demonstrando este como um padrão estável decorrente das interações locais entre os motoristas.

4.2.4 Modelo Baseado em Agentes: Escolha

Uma vez especificados o ambiente e as regras de interação entre este e os agentes, é necessário especificar estes últimos. A fim de se assemelhar o máximo possível ao contexto de um Logit Multinomial, os agentes aqui especificados

apresentam uma função utilidade que, por simplificação, apresentam os mesmos parâmetros para todos os indivíduos - que se tornam, para este fim, homogêneos. A utilidade é definida negativamente em função do tempo do deslocamento, do custo e da lotação de cada modal¹⁵. Pode ser expressa da seguinte forma:

$$U(t, c, l) = -\alpha t - \beta c - \gamma l, \quad (1)$$

$$\text{sendo } \frac{\partial U}{\partial t} = -\alpha < 0; \frac{\partial U}{\partial c} = -\beta < 0; \frac{\partial U}{\partial l} = -\gamma < 0,$$

onde t representa o tempo gasto no trajeto, c o custo e l a lotação, e α , β e γ suas respectivas utilidades marginais.

Além da função utilidade, cada agente possui uma memória curta, guardando os atributos da última viagem realizada em cada modal. A referência a memória é bastante afastada das discussões relacionadas ao Logit Multinomial. Baseado na maximização de utilidade tradicional, tal modelagem inicialmente ignora questões relacionadas à disponibilidade de informação. Para a avaliação de políticas públicas a partir de modelos de escolha discreta, no entanto, é realizada de forma implícita uma hipótese quanto a memória. Ao analisar o efeito de um choque no custo, por exemplo, assume-se que o indivíduo realizará sua escolha compreendendo que todas as demais variáveis permanecerão constantes. Tal procedimento considera, portanto, que os indivíduos formam suas expectativas sobre o futuro igualando-as à última memória que retêm dos modais, ignorando quaisquer eventuais alterações futuras. Desta forma, ainda que se baseando em uma racionalidade indutiva¹⁶, os agentes na modelagem não-econométrica mantêm similaridade direta com aqueles das avaliações de políticas baseadas no Logit Multinomial, ao assumir que os indivíduos formam suas expectativas da referida forma.

Tendo em vista tal discussão sobre memória e expectativas na avaliação de políticas públicas, faz-se necessário esclarecer também diferenças

¹⁵ Observa-se, aqui, que a utilidade resultante será obrigatoriamente negativa. Isso se deve por referir-se à necessidade de se deslocar até outro local a fim de desempenhar, lá, uma atividade. Atividade esta que, espera-se, gere uma atividade positiva e superior à desutilidade do deslocamento – mas que não é assunto do presente trabalho.

¹⁶ Aqui, racionalidade indutiva é utilizada no sentido de Arthur (1994). O autor aponta que “em situações de complicação, agentes não podem se apoiar no fato de que os agentes com os quais eles estão lidando comportar-se-ão sob racionalidade perfeita, e então serão forçados a adivinhar seu comportamento”. Nesta situação, seriam necessárias deduções localizadas baseadas em suas próprias hipóteses, revendo as ações conforme os resultados. No caso em questão, os indivíduos fazem a hipótese de que os demais agentes não alterarão suas escolhas.

entre os modelos aqui discutidos - de escolha discreta e baseado em agentes - com relação ao modelo microeconômico geralmente referenciado em análises de congestionamento sob a ótica da Economia Urbana, como em Bruekner (2011). Enquanto a abordagem dos modelos de escolha discreta, quando tratando de avaliações de políticas públicas, faz a hipótese implícita de que os indivíduos igualam as expectativas à última experiência realizada, a abordagem da Economia Urbana assume informação perfeita dos indivíduos, que escolhem sabendo a escolha dos demais indivíduos, maximizando sua utilidade. Neste sentido, tendo em vista o caráter caótico do trânsito e seu grande volume de informações, o processo adaptativo de tentativa e erro é consonante à proposta de Simon (1955)¹⁷. Tendo em vista a importância da experiência de cada agente para sua escolha, observa-se a importância de avaliação de políticas públicas em que o congestionamento seja uma propriedade emergente da interação entre os próprios indivíduos.

Retornando à apresentação do modelo, segue-se à apresentação dos modais. Até então apenas brevemente apresentados, os dois modais em questão apresentam três atributos importantes para a análise: tempo, custo e lotação. Para o carro, o tempo é variável, dependendo do trânsito encontrado pelos agentes; a lotação, por sua vez, é sempre zero. Já para o metrô, a velocidade constante implica que o tempo também será fixo; no entanto, a lotação dependerá do número de usuários que optem por tal modal. Para ambos os modais, o custo é exogenamente determinado.

Por fim, resta explicitar do modelo o processo de escolha, que acaba se tornando bastante simples. Utilizando-se dos atributos guardados na memória, ou seja, da última experiência com cada modal, os indivíduos obtêm uma utilidade para cada. Escolhem, portanto, aquele que retornar a maior utilidade.

Todo o processo realizado pelo modelo pode ser sucintamente resumido no seguinte algoritmo:

¹⁷ Segundo o autor (tradução nossa), "devido aos limites psicológicos do organismo (particularmente com respeito a habilidade computacional e preditivos), o esforço de racionalidade humano pode, no máximo, ser uma aproximação extremamente crua e simplificada para o tipo de racionalidade global que é implicada, por exemplo, pelos modelos de teoria dos jogos". Assim, aponta a necessidade de substituir tal racionalidade global "por um tipo de comportamento racional que seja compatível com o acesso à informação e às capacidades computacionais que são de fato possuídas pelos organismos, incluindo o homem, nos tipos de ambientes em que tais organismos existem".

Algoritmo geral:

1. Agentes são distribuídos de forma exógena entre os modais.
2. Com tais escolhas, os indivíduos trafegam, armazenando os atributos desta viagem.
3. Avaliam suas funções utilidade, comparando os modais utilizando atributos da última viagem realizada através de cada um.
4. Escolhem individualmente o modal de maior utilidade.
5. Repetição dos itens 2 a 5.

4.2.5 Resultados

A modelagem baseada em agentes foi aplicada para diversas configurações de parâmetros. Foram executados experimentos realizando alterações tanto nos parâmetros da função de utilidade quanto nos preços¹⁸ iniciais e na distribuição inicial dos agentes entre os modais. A amplitude desta análise de sensibilidade está expressa na Tabela 10. É importante ainda ressaltar que os gráficos posteriormente apresentados são apenas até rodada¹⁹ suficiente para que se possa observar seu padrão de comportamento – evidentemente mantido em experimentos com maior duração.

Quanto aos parâmetros, é importante ressaltar ainda dois aspectos. Primeiramente, os experimentos realizados utilizaram-se de todas as possíveis combinações daqueles apresentados. Além disso, os parâmetros foram estipulados para que as dimensões da escolha - custo, tempo e lotação - fossem comparáveis, sendo a dominância de um sobre os demais refletida nas configurações em que estes fossem iguais a zero.

Como resultado de tais experimentos, dois resultados principais são obtidos: um equilíbrio estável e outro instável. É importante ressaltar que equilíbrio, no presente artigo, está sendo utilizado para referir-se a uma situação estável em que os indivíduos não apresentam incentivos para alterar sua escolha. As situações de equilíbrio são apontadas nos gráficos 1 e 2.

¹⁸ Referindo-se aos modais, preço e custo são usados como sinônimos ao longo do texto.

¹⁹ Cada rodada refere-se a um processo de escolha de modal e deslocamento, por todos os indivíduos.

TABELA 10 – PARÂMETROS UTILIZADOS NOS EXPERIMENTOS DO MODELO BÁSICO

Parâmetro	Valores utilizados
α : desutilidade marginal do tempo	0, 0,1 e 0,2
β : desutilidade marginal do custo	0, 10 e 20
γ : desutilidade marginal da lotação	0, 0,8 e 1,6
Preço do carro	6, 7 e 8
Preço do metrô	3, 4 e 5
Agentes inicialmente usando carros ²⁰	40, 50 e 60

O Gráfico 1 expressa o equilíbrio estável obtido. Nestas situações, todos os indivíduos escolhem o mesmo modal, no caso, metrô. Isso decorre da homogeneidade dos agentes que, quando encarando situações semelhantes, reagem da mesma forma, ou seja, utilizando um mesmo modal. No gráfico, a experiência daqueles que se utilizaram de carro foi extremamente ruim, de modo que, após alterarem a escolha para o metrô, não ousam retornar a utilizar o carro. Tal situação, em consequência da utilidade altamente negativa obtida de metrô, permanece apesar de choques relevantes no custo.

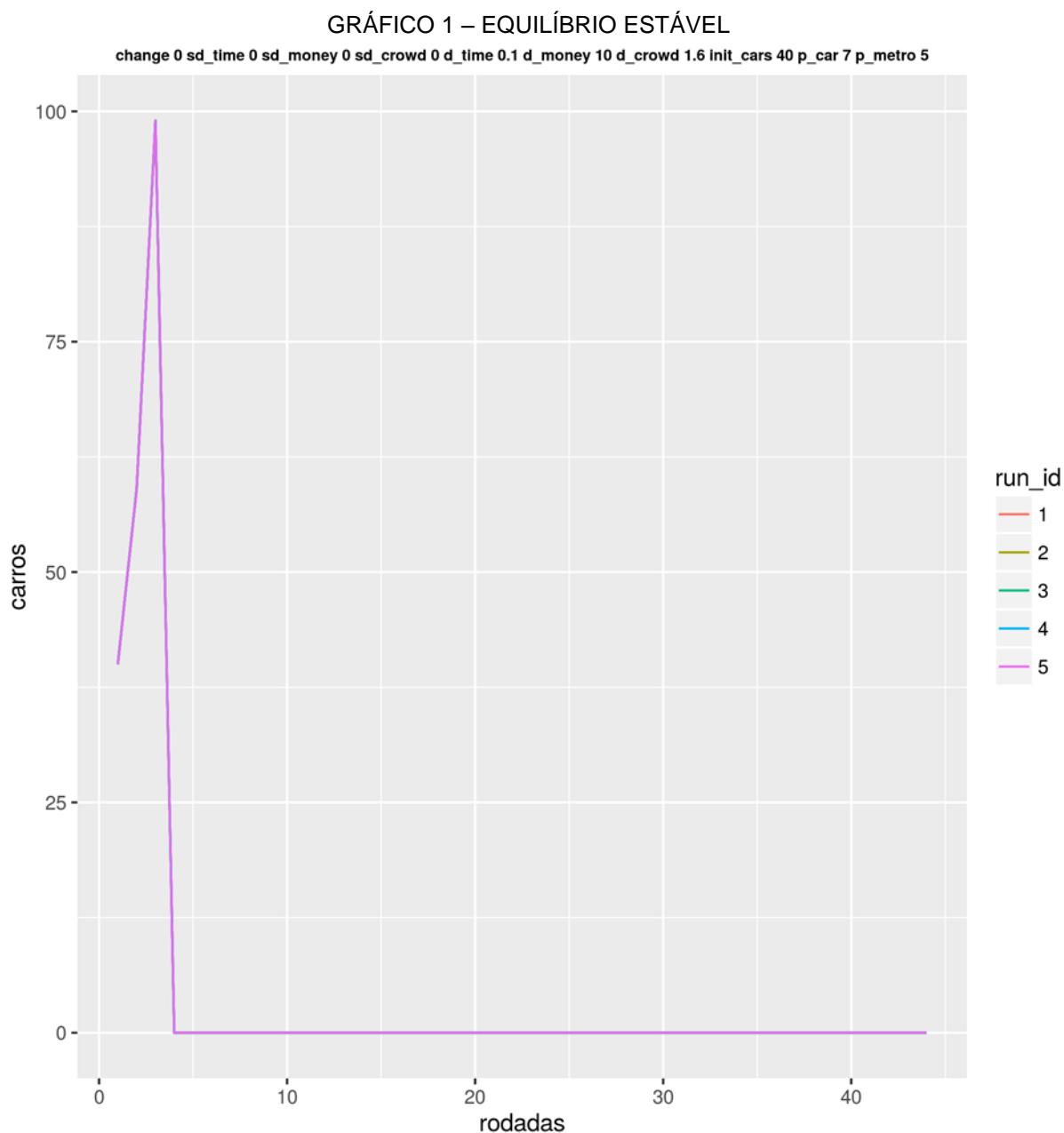
Antes de seguir ao próximo gráfico, é importante notar que todos os experimentos²¹ na especificação dos parâmetros utilizados na elaboração do Gráfico 1 apresentaram o mesmo resultado, indicando que a aleatoriedade inicial na disposição dos carros na via não afeta o resultado final. Esta observação também é válida com relação ao Gráfico 2.

O resultado mais interessante para a análise, enfim, é aquele de equilíbrio instável. Tal equilíbrio é estabelecido com uma proporção de indivíduos usando cada modal, de forma constante após algumas iterações. A avaliação de uma política pública nesse cenário gera resultados interessantes. A política pública em questão refere-se a uma redução dos custos de utilização do carro, promovendo um maior uso do modal. Tal política poderia visar o incentivo à indústria automobilística e, para

²⁰ O número total de agentes equivale à capacidade total da via para carros menos um: 99. Desta forma, a quantidade inicial de agentes que utilizam metrô é obtida como complemento àqueles que utilizaram carro.

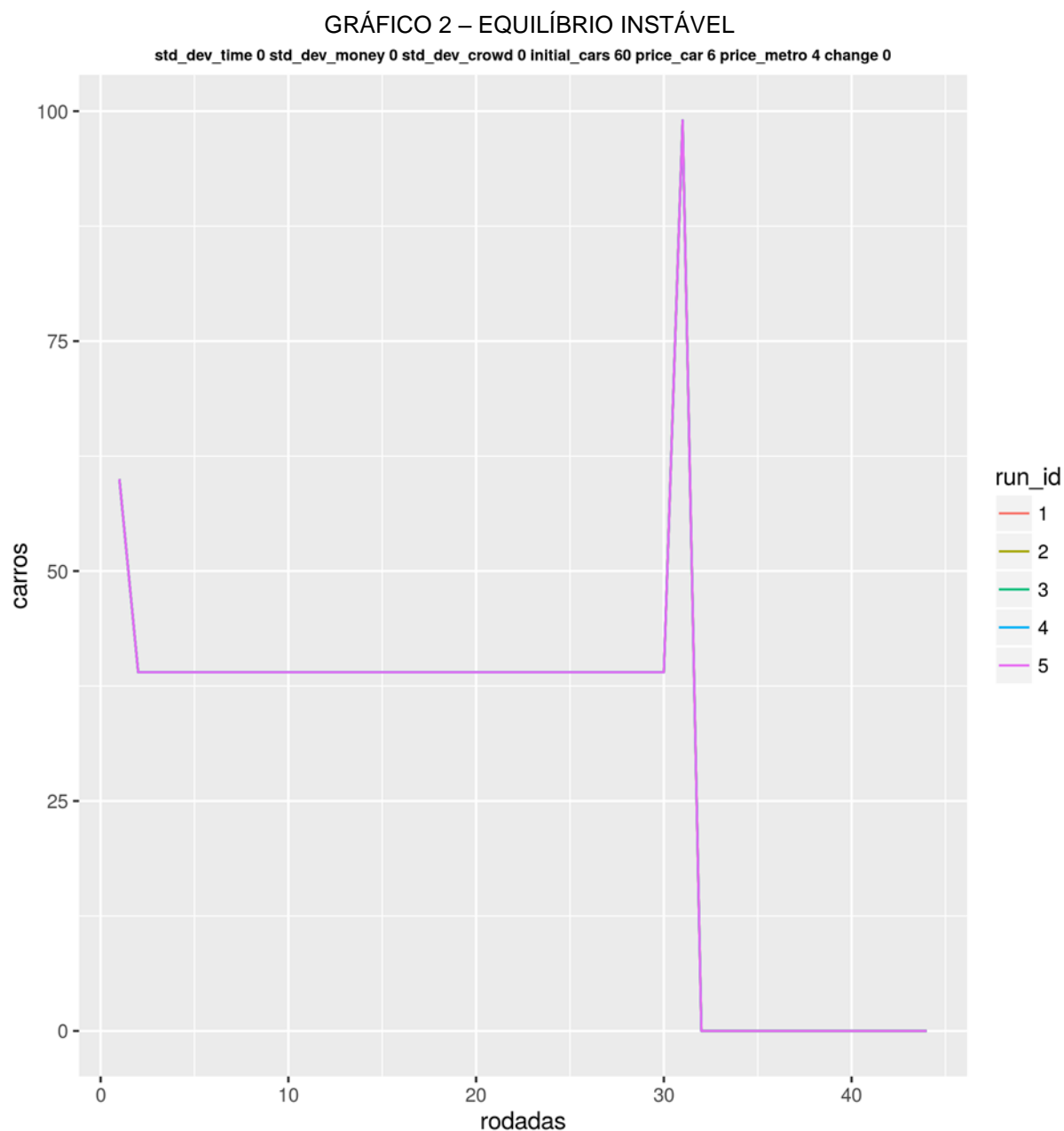
²¹ Os experimentos são referenciados, no gráfico, como “runs”.

tanto, ser realizada através de uma redução de impostos para combustíveis fósseis ou mesmo de impostos diretos sobre este setor, como realizado em 2008 no Brasil.



O efeito inicial de tal política é observado no Gráfico 2, com a elevação substancial da utilização de carros pelos indivíduos. É importante ressaltar que tal previsão é a mesma que seria realizada por um modelo Logit Multinomial. No entanto, em decorrência do elevado número de usuários de carro, forma-se um grande congestionamento. Tal congestionamento, por sua vez, eleva enormemente o tempo de deslocamento através deste modal. Desta forma, os indivíduos experimentam uma utilidade extremamente negativa e resolvem, na próxima

iteração, alterar sua escolha, passando a escolher metrô. Como a experiência de carro foi muito ruim, eles optam por permanecer utilizando o outro modal.



Tal equilíbrio instável, por sua vez, ilustra dois outros resultados importantes. Primeiramente, demonstra um potencial equívoco em uma análise de política pública de transporte utilizando-se de um modelo de escolha discreta. Além disso, demonstra uma situação que poderia explicar as causas pelas quais se obteria um resultado contrário àquele esperado de uma política pública, como apontam Nagel e Rasmussen (1994). De uma perspectiva bastante generativista, tal resultado seria uma resposta plausível à pergunta: “Como poderiam as interações locais de agentes

heterogêneos limitadamente racionais gerar a dada regularidade?” (EPSTEIN, 2005).

Como apontam Wilensky e Rand (2016), este resultado dá relevância a uma característica das propriedades emergentes crucial para a análise em questão: tais estruturas ao mesmo tempo em que são propriedades do sistema, geralmente retroalimentam cada elemento individual de que são compostas. A dinâmica coevolutiva entre as propriedades emergentes e os agentes que a compõem aparece, enfim, como figura essencial para a avaliação de políticas públicas de transporte.²²

4.3 POSSÍVEIS ALTERAÇÕES

A modelagem acima desenvolvida é suficiente para a observação de certa dinâmica coevolutiva, além de representar explicações para certos fenômenos de interesse. Outros padrões agregados estáveis, no entanto, são relevantes quando da análise de políticas de transporte, como equilíbrios dinâmicos e sub-reação a choques de preços. Neste sentido, algumas alterações do modelo básico foram também implementadas, a fim de obter conjuntos de microespecificações suficientes para gerar tais macrofenômenos, captando, assim, outros comportamentos plausível com relação à observação empírica.

O modelo básico aqui apresentado é bastante semelhante àquele desenvolvido por Nagel, Rasmussen e Barret (1996). Este, no entanto, busca analisar a distribuição de usuários não entre modais, mas entre duas vias. Para tanto, utiliza-se de dois artifícios. Primeiramente, faz uso de uma “tarifa dinâmica” com relação ao congestionamento, buscando analisar a eficiência desta na distribuição dos usuários. Além disto, o modelo permite uma certa aleatoriedade na escolha - tornando esta independente, pois, da função utilidade. Tal aleatoriedade, importante neste modelo, é a primeira alteração proposta ao modelo básico.

Por fim, como a referida pergunta generativista evidencia, a heterogeneidade é um fator importante para a análise. Desta forma, buscando também assemelhar-se

²² Por coevolução compreende-se, aqui, um processo amplo como apontado por Arthur (1994): “Assim como as espécies, para sobreviver e reproduzir, devem ser postas à prova pela competição e serem adaptadas dentro de um ambiente criado por outras espécies”, nesse mundo escolhas, para que sejam precisas e então realizadas, devem ser postas à prova pela competição e serem adaptadas dentro de um ambiente criado pelas escolhas dos outros agentes.

à análise de políticas públicas com uma modelagem Mixed Logit ou com um Logit Multinomial mais refinado, a inclusão de certa heterogeneidade é a segunda alteração implementada ao modelo inicial.

4.3.1 Alteração 1: incluindo aleatoriedade

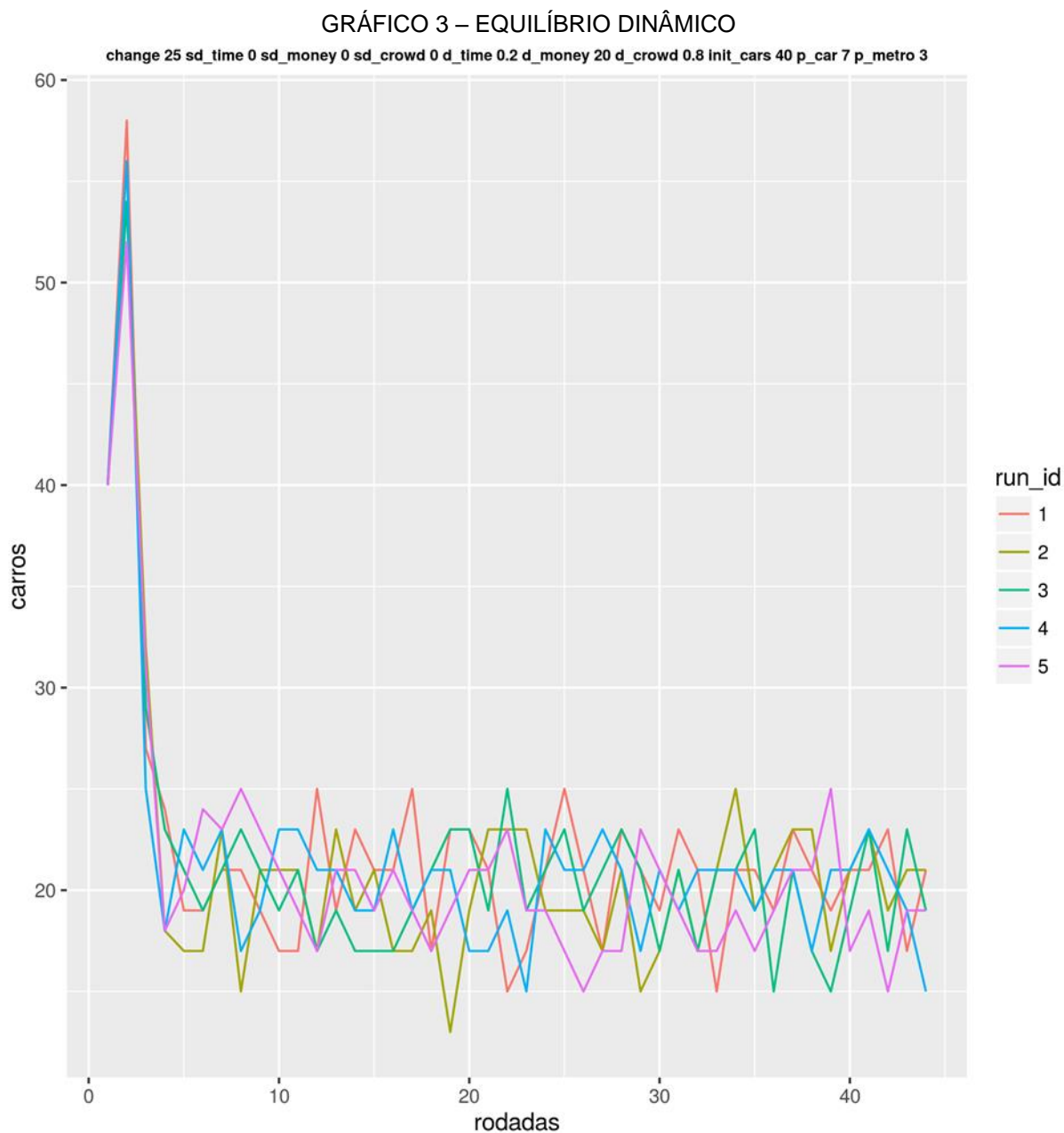
A modelagem proposta por Nagel, Rasmussen e Barret (1996) permite que os indivíduos, embora buscando escolher a rota mais barata, errem a escolha com uma probabilidade predefinida - no caso, 5%. Tal aleatoriedade, apontam os autores, “dá a cada motorista uma chance de tempos em tempos de atualizar sua informação sobre a outra possibilidade”.

Tal possibilidade foi também implementada no modelo básico de escolha de modal. Neste caso, no entanto, apesar da utilidade de uma alternativa ser maior, o indivíduo pode escolher a outra, com uma probabilidade também exogenamente determinada. Este “erro” poderia ser causado, por exemplo, por alguma dificuldade não antecipada pelo agente, como um problema mecânico no carro que o obriga a ir de metrô por um dia. O algoritmo geral que inclui tal alteração pode ser expresso como:

Algoritmo geral com aleatoriedade:

1. Agentes são distribuídos de forma exógena entre os modais.
2. Com tais escolhas, os indivíduos trafegam, armazenando os atributos desta viagem.
3. Avaliam suas funções utilidade, comparando os modais utilizando atributos da última viagem realizada através de cada um.
4. Escolhem individualmente o modal de maior utilidade.
5. Cada indivíduo pode errar a escolha, com uma probabilidade exogenamente determinada de inverter a escolha determinada em 4.
6. Repetição dos itens 2 a 6.

A aleatoriedade foi incluída também em diferentes magnitudes: 1%, 5%, 25% e 50%. O primeiro resultado evidente é a ausência de um equilíbrio como antes. Se anteriormente se obtinha um resultado constante, a inclusão de aleatoriedade impede que o número de usuários de cada modal seja constante. Observa-se, ao invés do equilíbrio estático, equilíbrios de alguma forma dinâmicos.



Assim como no modelo anterior, duas formas de equilíbrio resultantes chamam a atenção. A primeira, apresentada no Gráfico 3 através de 5 experimentos com mesma configuração de parâmetros, representa um equilíbrio dinâmico sem grandes alterações e/ou ciclos. Tal situação se aproxima bastante da realidade: embora existam pequenas oscilações na escolha do modal, as parcelas de utilização de cada um deles apresentam certa constância. Este resultado, porém, é obtido apenas em configurações mais extremas do modelo: com alta probabilidade de erro (25% e 50%) e parâmetros mais díspares em favor do metrô.

O segundo resultado de interesse é representado pelos Gráficos 4 e 5, em que dois experimentos com mesma configuração de parâmetros são apresentados

em cada gráfico. Tal situação, mais frequente nas diversas configurações de parâmetros, parece apresentar uma regularidade na oscilação entre a utilização de carro e metrô. Há, nesta especificação do modelo, a formação de ciclos semelhantes a bolhas. Após um processo inicial de ajustamento, todos ou quase todos os indivíduos passam a utilizar metrô. Aos poucos, alguns indivíduos trocam de modal, passando a utilizar carro, o que aumenta a utilidade tanto destes - que passam a pertencer a uma minoria a utilizar carro -, como daqueles que continuam a utilizar metrô - pois diminui a lotação dos vagões. À medida que os agentes vão se ajustando e se aproximando a um maior nível de utilidade, porém, o congestionamento torna-se um empecilho cada vez maior até que, quando o número de usuários de carro ultrapassa um determinado *threshold*²³, ele se torna insuportável para os indivíduos. Nesse momento, observa-se no Gráfico 5, há uma queda brusca da utilidade dos indivíduos que utilizam carro, e uma migração em massa do carro para o metrô, evidente no Gráfico 4.

²³ É importante ressaltar que tal *threshold* não é uma determinado *a priori* – é, na verdade, uma propriedade que emerge da interação entre os agentes.

GRÁFICO 4 – EQUILÍBRIO DINÂMICO: CICLOS SEMELHANTES A BOLHAS

std_dev_time 0 std_dev_money 0 std_dev_crowd 0 initial_cars 40 price_car 7 price_metro change 1

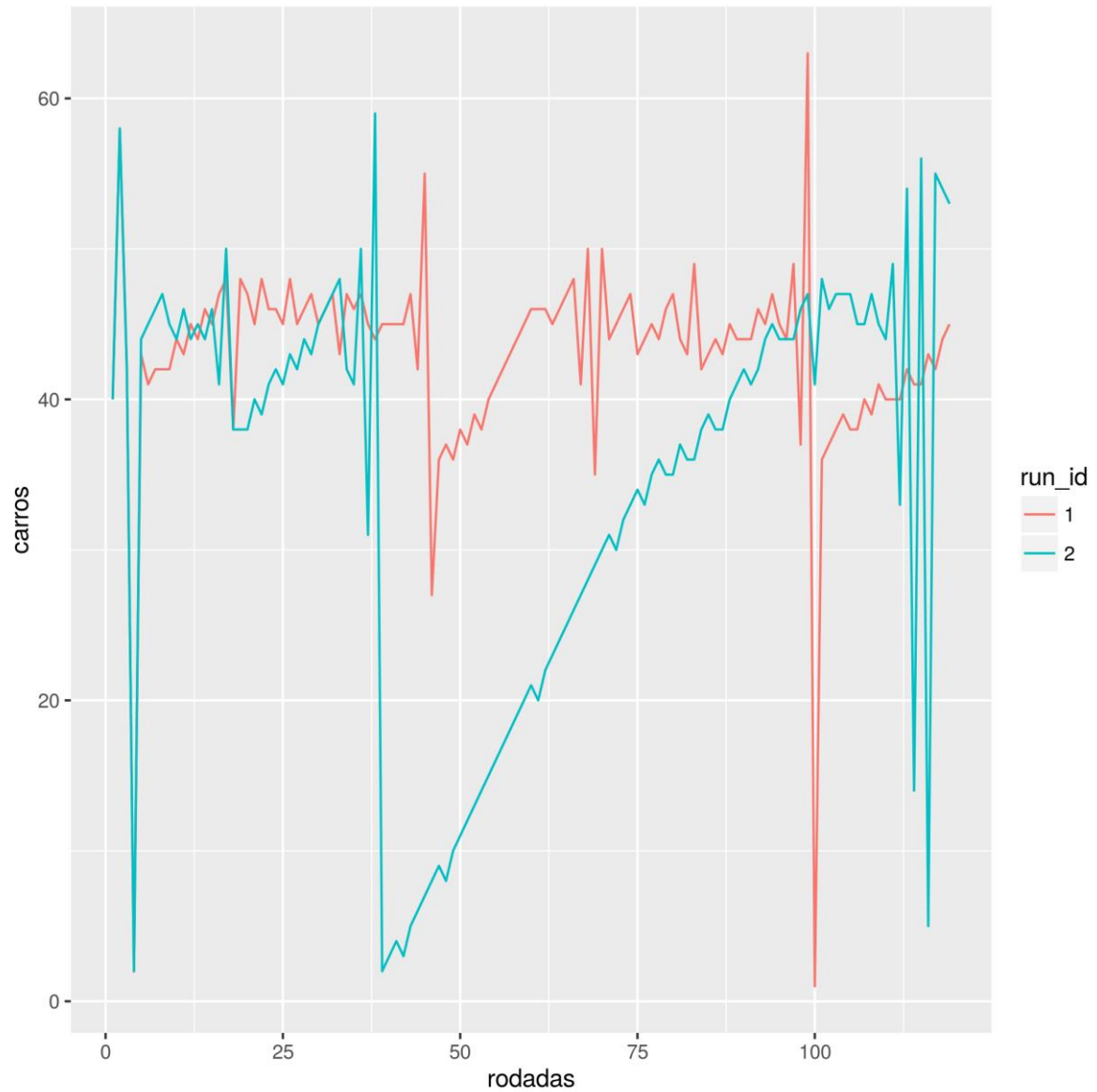


GRÁFICO 5 – UTILIDADE MÉDIA DOS INDIVÍDUOS POR MODAL – CICLOS DE QUEDA BRUSCA DA UTILIDADE DOS USUÁRIOS DE CARRO



Tal dinâmica, semelhante à formação de bolhas, ainda que pouco conectada com a observação cotidiana, parece conectada com as conclusões de Nagel, Rasmussen e Barret (1996): à medida que o trânsito se aproxima de uma situação mais eficiente, ele pode, também, apresentar maiores flutuações e, como consequência, menor previsibilidade, uma vez que o sistema está mais próximo à capacidade e então à *criticalidade*. Desta forma, corrobora também com as observações referidas por Batten (2000): “Uma descoberta importante de estudos de simulação de tráfego é esse regime crítico próximo à capacidade máxima, que para todos os intentos e propostas se parecem muito com um estado crítico auto-

organizado. Outros analistas apelidaram de “a fronteira do caos”. Nessa região, sistemas de transporte são muito sensíveis a pequenas perturbações.”

Esta dinâmica aparenta ser causada pela homogeneidade dos agentes: uma vez que o *threshold* é atingido, ele resulta na queda brusca da utilidade de todos os agentes que estavam utilizando carro. Como todos apresentam os mesmos parâmetros, todos optam pela utilização do metrô, estourando a bolha. Assim, a avaliação da heterogeneidade aparece como passo seguinte na expansão do modelo.

4.3.2 Alteração 2: incluindo heterogeneidade

Além do aspecto mencionado acima e da importância generativista, incluir heterogeneidade é uma forma importante de estabelecer um diálogo entre a modelagem baseada em agentes aqui proposta e a modelagem Mixed Logit, talvez o modelo mais refinado dentre aqueles de escolha discreta mais utilizados (LUCINDA; MEYER; LEDO, 2013; LUCINDA et al, 2015, PACHECO; CHAGAS, 2016).

Seguindo a concepção do Mixed Logit, portanto, atribui-se uma distribuição aos parâmetros entre os indivíduos: α, β e γ , como apontados na equação (1), seguem agora uma distribuição normal com média e desvio-padrão exogenamente determinados - no caso, a média utilizada foram iguais aos valores anteriormente estabelecidos, enquanto o desvio-padrão foi de 35% desta. É interessante ressaltar que tal definição pode representar, além do Mixed Logit, uma especificação de Logit Multinomial que inclua interações entre as variáveis, como tempo e renda, em uma população com alguma heterogeneidade na distribuição de renda.

Com esta alteração, o algoritmo geral passa a poder ser expresso como:

Algoritmo geral com aleatoriedade e heterogeneidade:

1. A cada agente são determinados os parâmetros α, β e γ que respeitam uma distribuição normal com média e desvio-padrão exogenamente determinados.
2. Agentes são distribuídos de forma exógena entre os modais.
3. Com tais escolhas, os indivíduos trafegam, armazenando os atributos desta viagem.
4. Avaliam suas funções utilidade, comparando os modais utilizando atributos da última viagem realizada através de cada um.
5. Escolhem individualmente o modal de maior utilidade.
6. Cada indivíduo pode errar a escolha, com uma probabilidade exogenamente determinada de inverter a escolha em 5.
7. Repetição dos itens 3 a 7.

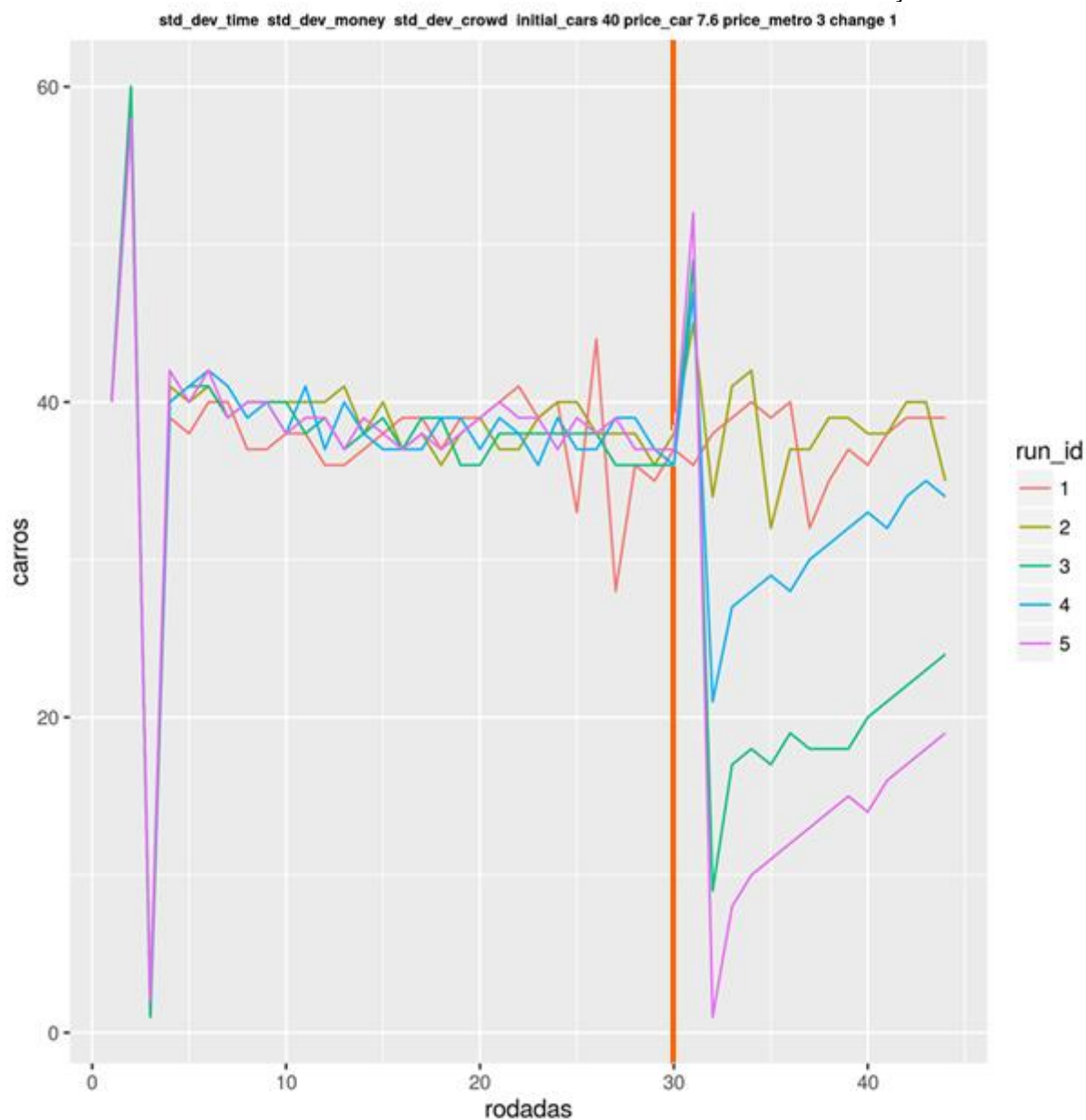
A inclusão da heterogeneidade, como esperado, elimina os ciclos semelhantes a bolhas antes observados. Assim, obtém-se como resultado mais comum uma espécie de equilíbrio dinâmico, de maneira bastante semelhante àquela obtida por Arthur (1994), tratando do problema do bar “El Farol”: um resultado dinâmico estável é obtido, em que há uma oscilação de usuários em cada modal, mantendo-se em torno de uma certa média. Tal equilíbrio, como expresso anteriormente, parece bastante alinhado com a realidade. Este padrão pode ser observado na primeira parte do Gráfico 6, mais precisamente até a rodada 30, marcada no gráfico pela linha alaranjada.

É relevante ressaltar, ainda, que tal resultado é também semelhante ao obtido por Nagel, Rasmussen e Barret (1996) com relação à escolha de indivíduos por rotas até o trabalho. Como apontam, tal abordagem diferencia-se daquela realizada por Arthur à medida que faz uso de apenas uma estratégia para obter o resultado estável. Esta diferença existe também entre o presente trabalho e o dos autores: enquanto eles fazem uso de uma “tarifa dinâmica” para tanto, aqui a estabilidade advém da referida heterogeneidade entre os agentes²⁴.

Outro importante resultado desta modelagem refere-se também à análise de políticas públicas. No gráfico referido, uma redução do custo de carro é simulada, de modo que se observa como resultado imediato uma grande migração de usuários do metrô para carro. Tal migração, é necessário notar, equivale à previsão de um modelo do tipo Mixed Logit. Percebe-se que, para além do curtíssimo prazo, tal previsão é superestimada, uma vez que, após o pico inicial relativo ao choque, os indivíduos se adaptam, reduzindo a utilização de metrô. O nível de médio e longo prazo pode até assumir valores superiores àqueles antes da política, mas evidentemente inferiores àqueles que seriam previsto pelo modelo de escolha discreta.

²⁴ Nagel, Rasmussen e Barret (1996) realizam uma modelagem baseada em agentes para simulação de transporte quanto à decisão sobre a utilização de uma ou outra via. Como apontado, partem de premissas semelhantes àquelas utilizadas neste artigo. No entanto, analisando o comportamento entre vias, estabelecem uma “tarifa dinâmica” para a utilização de uma das rodovias de modo a estabilizar as bolhas antes formadas.

GRÁFICO 6 – EQUILÍBRIO DINÂMICO E POSSÍVEL SUPERESTIMAÇÃO



É importante ressaltar novamente que tal comportamento é resultado da retroalimentação realizada pelas estruturas emergentes - como propriedades do sistema -, com relação a cada elemento individual que a compõe: uma dinâmica coevolutiva. Num primeiro momento, o preço mais baixo do metrô aparece como grande atrativo para os indivíduos, que migram em massa. A lotação não esperada, uma propriedade emergente, reverte a expectativa de alguns destes que, por sua vez, voltam a utilizar carros, até que um novo equilíbrio dinâmico - outra propriedade emergente - seja formado.

4.4 CONCLUSÃO

A modelagem baseada em agentes aqui desenvolvida é bastante simplificada. Como tal, apresenta várias limitações, chamando atenção por não lidar com questões como distribuição de renda. É provável, no entanto, que a incapacidade de fazer previsões seja uma das características de maior relevo.

Tal carência de previsão, porém, não deve obnubilar as demais contribuições apresentadas pelos modelos desenvolvidos. Como pontua Epstein (2008), tal modelagem pode ser justificada por diversas características.

Inicialmente, evidencia-se a capacidade de gerar explicações, conjuntos de microespecificações suficientes para gerar os macrofenômenos de interesse. Adicionalmente, e para a realização de tal explicação, tais modelos elucidam a dinâmica central do processo de escolha, possibilitando a avaliação das propriedades emergentes e dos processos coevolutivos a ele relacionados. Através destes, por fim, deixa claras as dificuldades das referidas modelagens econométricas em realizar previsões para além do curtíssimo prazo. Mais do que isso, esclarece as dificuldades relacionadas à utilização de tal modelagem para a elaboração de políticas públicas de transporte para mais longo prazo.

É importante ressaltar, ainda, que dinâmica possibilitada pela bordagem proposta permite explorar outras características além do congestionamento, como “as fronteiras do caos” tradicionais ao trânsito e a importância da heterogeneidade para manutenção de maior constância.

As limitações apresentadas, porém, levam a outra qualidade apresentada por Epstein (2008): o estabelecimento de novas perguntas. Neste sentido, refinamentos e expansões nos modelos aqui apresentados podem elucidar questões importantes, como a importância da distribuição de renda e/ou geográfica. Ao mesmo tempo, apresenta a possibilidade de complementaridade entre as duas metodologias, utilizando-se do ferramental econométrico para a calibração dos parâmetros e de modelos baseados em agentes mais refinados (como MATSim, ALBATROSS e TRANSIMS) para avaliação dos desdobramentos das políticas de transporte.

5 CONCLUSÃO

O presente trabalho objetivou apresentar e explorar potenciais contribuições para a escolha de modal e, portanto, para a melhoria da avaliação e elaboração políticas públicas de transporte. Neste sentido, todas as contribuições propostas se apresentaram como melhorias em relação às alternativas atualmente utilizadas.

Quanto ao primeiro artigo, a utilização da ferramenta Google Maps para a estimação dos valores das alternativas não escolhidas apresentou resultados com maior significância estatística e poder explicativo com relação ao método sintético utilizado de forma corrente na literatura brasileira. Tais resultados corroboram com aqueles encontrados por Javanmardi et al (2015), que apresentaram maior adequação em relação à utilização de matrizes *skim*.

Uma vez que tal abordagem foi aplicada apenas a uma subamostra da Pesquisa Origem e Destino de 2007, a utilização de toda a amostra, assim como de diferentes recortes desta, aparece como potencial para obtenção de novos resultados relevantes. É importante citar, ainda, que a base de dados formada com a utilização do Google Maps permite a construção de novos agregados de alternativas, possibilitando também a exploração de diferentes especificações do modelo.

Ainda quanto ao primeiro artigo, ressalta-se que a defasagem temporal entre a realização da Pesquisa Origem e Destino, realizada em 2007, e a obtenção de dados através do Google Maps, em 2016, implicou em alguma perda de informação em decorrência da inadequação da infraestrutura entre os dois períodos. Desta forma, como a elaboração da Pesquisa Origem e Destino da Região Metropolitana de São Paulo ocorre a cada 10 anos, existe uma grande possibilidade de obtenção de novos e melhores resultados ao estimar os valores das alternativas não escolhidas com maior proximidade temporal.

De maneira semelhante ao primeiro, o segundo artigo também apresentou maior poder explicativo, corroborando com os resultados obtidos por Swait (2009), aplicados agora à Pesquisa Origem e Destino de 2007. Estes resultados apresentam potencial de melhora com a utilização de diferentes especificações – como aquelas mais semelhantes às utilizadas por Swait (1984) e Swait (2009). Para o artigo em questão, optou-se por manter as especificações de Barcellos (2014) a fim de reforçar a comparabilidade entre os modelos.

Os resultados do segundo artigo também indicam a utilização de comportamentos não maximizadores pelos agentes. Dessa forma, novos estudos podem ser realizados para explorar também o modelo 3-Mix elaborado por Swait (2009), com Zona de Dominância. Outro desdobramento possível desta avaliação refere-se à possibilidade de analisar quais variáveis poderiam explicar o pertencimento de uma alternativa a uma destas determinadas zonas, assim como calcular os coeficientes a elas relacionados.

O terceiro artigo, por sua vez, aponta a importância da dinâmica coevolutiva e interação entre agentes no que se refere à avaliação de políticas públicas de transporte. O artigo sugere que, ao ignorar tais aspectos, pode-se subestimar ou superestimar os resultados das políticas. Além disso, aponta que comportamentos, como o aspecto caótico encontrado no trânsito não podem ser captados a partir de apenas de modelos de escolha discreta. Tais conclusões sugerem a utilização de modelos baseados em agentes para avaliação de políticas públicas de transporte (como MATSim, TRANSIMS ou ALBATROSS).

Além da contribuição individual de cada artigo, a utilização conjunta dos artigos apresenta potencial importância. A solução apresentada no primeiro artigo traz, para aquela explorada no segundo, novas possibilidades de especificação do modelo, permitindo inclusive agregações diferenciadas, dada a maior disponibilidade de informação. Traz, também, a necessidade de investigação quanto à manutenção dos melhores resultados obtidos com a utilização do modelo 2-Mix – isto, em decorrência do fato de a base de dados formada com o uso do Google Maps incluir refinamentos prévios com relação ao conjunto de escolha.

Já a combinação do modelo proposto no segundo artigo com aquele sugerido no terceiro traz novas possibilidades de avaliação para a dinâmica coevolutiva a serem exploradas. Assim como a inclusão de heterogeneidade e aleatoriedade, a utilização de comportamentos não-maximizadores pode elucidar dinâmicas não avaliadas em processos tradicionalmente maximizadores.

Por fim, o resultado os bons resultados individuais sugerem espaço para uma relevante melhoria da utilização conjunta das abordagens propostas. A utilização de uma base de dados mais robusta, atrelada a um modelo que explicita o processo de escolha de maneira mais precisa, além da inclusão de dinâmica coevolutiva ao modelo, apresenta relevante potencial para promover avaliações de políticas públicas com mais precisas e adequadas.

6 REFERÊNCIAS

ADAMOWICZ, W., BUNCH, D., CAMERON, T. A., DELLAERT, B. G. C., HANNEMAN, M., KEANE, M., LOUVIERE, J., MEYER, R., STEENBURGH, T., SWAIT, J. Behavioral frontiers in choice modeling. **Marketing Letters**, Springer, vol. 19(3), pages 215-228, dezembro, 2008.

ARTHUR, W. B. *Inductive Reasoning and Bounded Rationality: the El Farol problem*. **American Economic Review**, 84, 406-411, 1994.

BARCELLOS, T. M. *Não são só 20 centavos: efeitos sobre o tráfego da Região Metropolitana de São Paulo devido a redução na tarifa de ônibus financiada pelo aumento da CIDE nos combustíveis da cidade de São Paulo*. Dissertação (Mestrado) - Departamento de Economia, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, 2014.

BATTEN, D. F. **Discovering Artificial Economics: How Agents Learn and Economy Evolves**. Westview Pres, 2000.

BEN-AKIVA, M., LERMAN, S. R. **Discrete Choice Analysis: Theory and Application to Travel Demand**. MIT Press, Cambridge, Massachusetts, 1985.

BHAT, C. R. *A heteroscedastic extreme value model of intercity travel mode choice*. **Transportation Research Part B**, 29(6), 471-483, 1995.

BRUEKNER, J. K. **Lectures on Urban Economics**. Cambridge: The MIT Press, 2011.

COHEN, J., KELLY, F. P. *A Paradox of Congestion in a Queuing Network*. **Journal of Applied Probability**, v. 27, p. 730-734, 1990.

CROISSANT, Y. **Estimation of Multinomial Logit Model in R: The Package mlogit**. R package version 0.2-3. <http://CRAN.R-project.org/package=mlogit>. 2012.

DISSANAYAKE, D., MORIKAWA, T. *Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference Nested Logit model: case study in Bangkok Metropolitan Region*. **Journal of Transport Geography**, v. 10, p. 402-410, 2010.

EMPLASA, **Por Dentro da Região Metropolitana de São Paulo – RMSP**. São Paulo, São Paulo, Junho/2011.

EPSTEIN, J. M.; AXTELL, R. **Growing Artificial Societies: Social Science From The Bottom Up**. New York: Brookings Institute Press, 1996.

EPSTEIN, J. M. *Remarks on the Foundations of Agent-Based Generative Social Science*. SFI Working Paper, Santa Fe, 2005.

EPSTEIN, J. M. *Why Model?* *Journal of Artificial Societies and Social Simulation*, vol 11, nº 4 12, out 2008.

ESTER, O., MCDONALD, K., SUTCLIFFE, R., HOFFMANN, B., FUKS, K. B., DRAGANO, N., VIEHMANN, A., ERBEL, R., JÖCKEL, K-H, PUNDT, N., MOEBUS, S. *Residential Road Traffic Noise and High Depressive Symptoms after Five Years of Follow-up: Results from the Heinz Nixdorf Recall Study*. **Environmental Health Perspectives**. v. 124, 2016.

FERES, L. J. *Diferencial de tarifa entre pico e vale como ferramenta de suavização da demanda no sistema de transporte público da cidade de São Paulo*. Dissertação (Mestrado Profissional) - Programa de Mestrado Profissional em Economia, Insper Instituto de Ensino e Pesquisa, São Paulo, 2015

HADDAD, E. A.; HEWINGS, G. J. D.; PORSSE, A. A.; VAN LEUWEN, E.; VIEIRA, R. S. The Underground Economy: Tracking the Higher-order Economic Impacts of the São Paulo Subway System. **Transportation Research Part A: Policy and Practice**, v.73, p 18-30, 2015.

HADDAD, E. A.; VIEIRA, R. S. *Mobilidade, Acessibilidade e Produtividade: Nota sobre valoração econômica do tempo de viagem na Região Metropolitana de São Paulo*. TD-NEREUS, ago 2015.

HENSHER, D. A., ROSE, J. M., GREENE, W. H. **Applied Choice Analysis: A Primer**. Cambridge University Press, New York, 2005.

IBOPE. 9ª Pesquisa sobre Mobilidade Urbana - Rede Nossa São Paulo. *Instituto Brasileiro de Opinião e Estatística, Opinião Pública*. São Paulo, set. de 2015. Disponível em: <<http://www.nossasaopaulo.org.br/pesquisas/pesquisaibope2015completa.pdf>>. Acessado em: ago. 2016.

JAVANMARDI, M.; LANGERUDI, M. F.; ANBARANI, R. S.; MOHAMMADIAN, A. K. *Mode Choice Modelling Using Personalized Travel Time and Cost Data*. Artigo apresentado em 14th International Conference on Travel Behavior Research, England, July 19-23, 2015.

KATONA, G. Rational behavior and economic behavior. **Psychological Review**, Vol 60(5), pp. 307-318. Set. 1953

KOPPELMAN, F. S., BHAT, C. *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*. Preparado para U.S. Department of Transportation, Federal Transit Administration, 2006.

LUCINDA, C. R.; MEYER, L. G.; LEDO, B. A. *Urban Road Tax in a Large Emerging Market: Some Brazilian Evidence*. Artigo apresentado em 35º Encontro Brasileiro de Econometria. Foz do Iguaçu, 12 dez. 2013. Disponível em: <https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=sbe35&paper_id=110>. Acesso em: ago 2016.

LUCINDA, C. R.; MOITA, R. M. S; MEYER, L. G.; LEDO, B. A. *The Economics of Sub-optimal Policies for Traffic Congestion*. Working Paper 83, Rede de Economia Aplicada, set. 2015.

MARSCHAK, J. Binary Choice Constraints of Random Utility Indicators. **Stanford Symposium on Mathematical Methods in The Social Sciences**, Stanford Univ. Press: Stanford, 1960.

MCFADDEN, D. Conditional Logit Analysis of Qualitative Choice Behavior. **Frontiers in Econometrics**, Academic Press: New York, pp. 105-142, 1974.

MCFADDEN, D. Economic Choices, Nobel Lecture, December 2000. **American Economic Review**, Vol. 91, No. 3, 351-378, June 2001 e em Les Prix Nobel, 2001.

METRÔ. **Pesquisa Origem e Destino 2007**. [s.l: s.n.].

MORIKAWA, T. **Incorporating Stated Preference Data in Travel Demand Analysis**. Tese (Doutorado). Departamento de Engenharia Civil, MIT. 1989.

NAGEL, K.; RASMUSSEN, S. Traffic at the Edge of Chaos. In Brooks, R., ed., **Artificial Life IV**. Cambridge, MA: MIT Press, 1994.

NAGEL, K.; RASMUSSEN, S.; BARRETT, C. L. Network Traffic as a Self-Organized Critical Phenomenon. In: SCHWEITZER, F., ed. **Self-Organization of Complex Structures: From Individual to Collective Dynamics**. Gordon and Breach, pp. 579-592, 1997.

NAGEL, K.; SCHRENCKENBERG, M. *A cellular automaton model for freeway traffic*. **Journal de Physique**, v. 2, pp. 2221-2229, 1992.

PACHECO, T. S.; CHAGAS, A. L. S. *Demanda por Transporte na Região Metropolitana de São Paulo e Política de Pedágio Urbano para Redução de Congestionamento*. TD NEREUS, São Paulo, abr. 2016.

PAYNE, J.W., BETTMAN, J.R., JOHNSON, E.J. **The Adaptive Decision Maker**. Cambridge University Press, New York, 1993

PRELEC, D. Values and Principles: Some limitations on traditional economic analysis, **Perspectives on Socioeconomics**. London: M. E. Sharpe, 1991.

SILVA, C. B. P., SALDIVA, P. H. N., AMATO-LOURENÇO, L. F., RODRIGUES-SILVA, F., MIRAGLIA, S. G. E. K. *Evaluation of the air quality benefits of the subway system in São Paulo, Brazil*. **Journal of Environmental Management**, 101: 191, 2012

SILVA, F. R., SANTOS, U. P., SALDIVA, P. H. N., AMATO-LOURENÇO, L. F., MIRAGLIA, S. G. E. K. *Valuation of the absenteeism due to air pollution in Sao Paulo, Brazil*. **ISEE Conference 2010**. Oldenburg - Bremen 2010.

SIMON, H.A.. Behavioral model of rational choice. **Quarterly Journal of Economics**, 69 (1), pp. 99–118, 1955

SISTEMA FIRJAN. O custo dos deslocamentos nas principais áreas urbanas do Brasil. 2015. Disponível em: <http://www.firjan.com.br/lumis/portal/file/fileDownload.jsp?fileId=2C908A8F4F8A7DD3014FB26C8F3D26FE&inline=1>, acesso em março de 2017.

SORENSEN, M., ANDERSEN, Z. J., NORDSBORG, R. B., JENSEN, S. S., LILLELUND, K. G., BEELEN, R. SCHMIDT, E. B., TJONNELAND, A., OVERVAD, K., RAASCHOU-NIELSEN, O. *Road Traffic Noise and Incident Myocardial Infarction: A Prospective Cohort Study*. **PLOS One**. 2012.

SUGIYAMA, Y.; FUKUI, M.; KIKUCHI, M.; HASEBE, K.; NAKAYAMA, A.; NISHINARI, K.; TADAKI, S. *Traffic jams without bottlenecks - experimental evidence for the physical mechanism of the formation of a jam*. **New Journal of Physics**. v. 10, pp 7, mar. 2008.

SWAIT, J. **Probabilistic choice set formation in transportation demand models**. Tese (Doutorado) - Departamento de Engenharia Civil, MIT, 1984.

SWAIT, J., BEN-AKIVA, M. *Empirical Test of a Constrained Choice Discrete Model: Mode Choice in Sao Paulo, Brazil*. **Transportation Research Part B: Methodological**. v.21B, p. 103-115, 1987.

THURSTONE, L. L. A Law of Comparative Judgment. **Psychological Review**, 34, pp. 273-286, 1927.

TRAIN, K. **Discrete choice methods with simulation**. 2. ed. New York: Cambridge University Press, 2009.

WASHINGTON, S.; RAVULAPARTHY, S.; ROSE, J. M.; HENSHER, D.; PENDYALA, R. *Bayesian Imputation of Non-Chosen Attribute Values in Revealed Preference Surveys*. **Journal of Advanced Transportation**. 48:48-65. 2014.

WILENSKY, U. *NetLogo Traffic Basic Model*. 1997.