

Tese apresentada à Pró-Reitoria de Pós-Graduação e Pesquisa da Universidade de Brasília, como parte dos requisitos para obtenção do título de Doutor em Economia Aplicada no Departamento de Economia, Área de Economia Aplicada.

Patrick Franco Alves

**UMA ANÁLISE EMPÍRICA DO MERCADO DE
REVENDA DE CARROS USANDO RASPAGEM
DE DADOS DA INTERNET**

Tese aprovada em sua versão final pelos abaixo assinados:

Prof. Dr. Daniel Oliveira Cajueiro

Orientador

Campus Universitário Darcy Ribeiro
Brasília, DF - Brasil
2016

Dados Internacionais de Catalogação na Publicação (CIP)
Divisão Biblioteca Central da UnB

Alves, Patrick Franco
Uma Análise Empírica do Mercado de Revenda de Carros Usando Raspagem de Dados da Internet / Patrick Franco Alves.
Brasília, 2016.
161f.

Tese de Doutorado – Curso de Departamento de Economia. Área de Economia Aplicada – Universidade de Brasília, 2016. Orientador: Prof. Dr. Daniel Oliveira Cajueiro. .

1. Big Data. 2. Segmentação do Mercado Secundário. 3. Comunalidade. 4. Modelo Hierárquico Hedônico. 5. Depreciação de Carros. 6. Modelos Não-Lineares. I. Departamento de Economia. Universidade de Brasília. Divisão de Economia.

REFERÊNCIA BIBLIOGRÁFICA

ALVES, Patrick Franco. **Uma Análise Empírica do Mercado de Revenda de Carros Usando Raspagem de Dados da Internet**. 2016. 161f. Tese de Doutorado – Universidade de Brasília, Brasília.

CESSÃO DE DIREITOS

NOME DO AUTOR: Patrick Franco Alves

TÍTULO DO TRABALHO: Uma Análise Empírica do Mercado de Revenda de Carros Usando Raspagem de Dados da Internet.

TIPO DO TRABALHO/ANO: Tese / 2016

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Tese e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta Tese pode ser reproduzida sem a autorização do autor.

Patrick Franco Alves
Campus Universitário Darcy Ribeiro - Prédio da FACE
CEP 70910-900 – Brasília-DF

UMA ANÁLISE EMPÍRICA DO MERCADO DE REVENDA DE CARROS USANDO RASPAGEM DE DADOS DA INTERNET

Patrick Franco Alves

Composição da Banca Examinadora:

Prof. Dr. Alexandre Messa Silva	Membro Externo	-	DISET/IPEA
Prof. Dr. Daniel Oliveira Cajueiro	Orientador	-	UnB
Prof. Dr. Victor Gomes e Silva	Membro Interno	-	ECO/UnB
Prof. Dr. Herbert Kimura	Membro Interno	-	ADM/UnB
Prof. Dr. Pedro Henrique Albuquerque	Membro Externo	-	ADM/UnB

Agradecimentos

Bendito sejas Tu, Eterno, nosso Deus, Rei do Universo, que conservaste em vida, nos amparaste e nos fizeste chegar a esta época. Agradeço primeiramente ao Eterno, Criador do Universo, por me guiar e cuidar de mim diante dos obstáculos.

Agradeço à minha amada esposa, Marília, minha eterna namorada. Obrigado pelo afeto, pela paciência e por estar ao meu lado. Obrigado por cuidar de mim. *Because you kept holding my hand, the angels could not take-me away.*

Agradeço também aos meus queridos pais, Robernilce e Nelice, pelos ensinamentos, carinho, amizade e atenção. Aos meus irmãos, Petercley, Priscila e Pollyana por todos os anos de amizade e janelas de descontração durante os anos de estudo. Agradeço à minha querida sogra, Marina, por fornecer o abrigo e sossego de sua casa nos finais de semana durante os anos de estudo e elaboração da tese.

Agradeço ao meu orientador, Daniel Cajueiro, pelo crédito, paciência, e pelas conversas enriquecedoras durante a elaboração desta tese. Por ter me influenciado positivamente a estudar Python, R, Latex e Big Data, entre outros assuntos.

Agradeço também àqueles que salvaram o dia. Durante o primeiro ano do doutorado, um sonho quase interrompido. Agradeço à Dra. Cláudia Ottaiano, ao Dr. João Nunes e a todos os dedicados médicos e enfermeiros do Cettro. Obrigado por cuidarem de mim. Sem a ajuda de vocês eu não estaria completando esta fase da minha vida.

Aos amigos de estudo e pesquisa, pelos intervalos de descontração durante o cafezinho da UnB e solidariedade nos semestres difíceis. A todos que compartilharam momentos difíceis nessa jornada: André Maranhão, Amon Kablan, Calebe Figueiredo, Claudiano Neto, Cristiano Guedes, Dea Guerra, Giovanni Beviláqua, Gustavo Costa, Gustavo Varela, Hélio Fonseca, João Vítor, Márcia Paixão, Oliveira Alves, Roberto Batista, entre outros.

Agradeço também às secretárias do departamento de Economia da UnB, sempre muito prestativas, Carina Lima e Joana D'ark.

Agradeço à CAPES e ao CNPq pelo suporte financeiro.

"With Big Data comes big responsibility, 2014."
— HARVARD BUSINESS REVIEW STAFF -

Resumo

Estudamos o mercado secundário de automóveis utilizando anúncios de venda de carros na internet. O mercado secundário de carros é importante por introduzir uma série de substitutos parciais para os carros novos. Apesar de sua relevância, existem relativamente poucos estudos sobre o mercado de carros usados. Contribuímos para a literatura ilustrando como as técnicas de *Big Data* podem contribuir para pesquisa econômica. Inspirados pelo paradigma do *Big Data*, nós construímos uma base de dados utilizando *Python* e *Web Scraping*. A segmentação de carros usados, utilizando um algoritmo não-supervisionado, revela clusters que possuem baixa correlação com os segmentos de carros zero quilômetro. Sabe-se que as montadoras projetam carros utilizando comunalidade para obter economias de custos. Usando modelos de comunalidade, preços hedônicos e modelos hierárquicos, verificamos que os efeitos da comunalidade sobrevivem no mercado secundário de carros. Por último, caracterizamos a perda de preço dos veículos utilizando modelos de crescimento exponencial, onde o ponto de inflexão determina o momento de desaceleração da curva de depreciação. Consumidores de carros zero irão absorver grande perda de preços no momento da revenda, enquanto consumidores de veículos com idade maior que o ponto de inflexão não terão grandes perdas de preço.

Abstract

We study the secondary car market using cars sales data in on-line advertisements. The secondary car market is important for introducing a lot of partial substitutes for new cars. But despite its importance, there are relatively few studies about the secondary car market. We contribute to the literature by illustrating how Big Data techniques can contribute to economic research. We inspired ourselves by the paradigm of Big Data and built a database using *Python* programming and *Web Scraping* and then we segment the used cars market using an unsupervised algorithm. This reveals clusters with low correlation with the new cars market segments. It is well known that, in order to achieve savings costs, the cars manufacturers uses commonality in production. We built over the literature on commonality models, hedonic price framework and hierarchical models to show that the effect of commonality still survives in the secondary cars market. Finally, we characterized the loss of cars prices using the exponential growth models, where the inflection point determines the time of deceleration in the depreciation path. We find that zero car buyers will absorb great loss in price at the time of resale, while used cars buyers of vehicles older than the inflection point will not have big losses in prices at the time of resale.

Keywords: Big Data, Secondary Market, Commonality, Hierarchical Hedonic Model, Car Depreciation, Non-linear models

Sumário

LISTA DE FIGURAS	xi
LISTA DE TABELAS	xiii
1 INTRODUÇÃO	15
1.1 Objetivo	15
1.2 Organização do trabalho	16
2 BIG DATA E MERCADO DE CARROS NO BRASIL	18
2.1 Big Data	18
2.2 Mercado de Carro Usados no Brasil	21
2.3 Raspagem de Dados	22
2.4 Base de Dados	25
3 SEGMENTAÇÃO DO MERCADO DE CARROS USADOS	27
3.1 Introdução	27
3.2 Revisão da Literatura	29
3.2.1 Segmentação Industrial	30
3.2.2 Importância Econômica da Segmentação	32
3.2.3 Mercado de Bens Duráveis	34
3.2.4 Sistemas de Classificação de Veículos	38

3.3	Dados Utilizados	40
3.4	Metodologia: Clusters K-Means	41
3.5	Resultados	42
3.6	Conclusões	51
4	MODELOS DE COMUNALIDADE E CANIBALIZAÇÃO	53
4.1	Introdução	53
4.2	Revisão da Literatura	54
4.2.1	Comunalidade e Canibalização	55
4.3	Metodologia	59
4.3.1	Modelo de Preços Hedônicos	59
4.3.2	Modelo Hierárquico	64
4.3.3	Verificação Empírica dos Efeitos da Comunalidade	67
4.4	Dados	70
4.5	Resultados dos Modelos	71
4.6	Conclusões	77
5	MODELAGEM DA DEPRECIAÇÃO DE VEÍCULOS	79
5.1	Introdução	79
5.2	Revisão da Literatura	82
5.2.1	Políticas de Descarte de Veículos Usados como Sucata	82
5.2.2	Modelo de Consumidores Míopes	87
5.2.3	Modelo Contábil de Depreciação de Bens Duráveis	88
5.3	Metodologia	92
5.3.1	Modelos de Crescimento Não-Lineares	92
5.3.2	Pontos de Inflexão	98
5.4	Dados Utilizados	99

5.5	Resultados	100
5.6	Conclusões	105
6	CONCLUSÃO	109
	REFERÊNCIAS BIBLIOGRÁFICAS	112
	APÊNDICE A – CAPÍTULO 2	127
A.1	Análise Descritiva da Base de Dados	127
	APÊNDICE B – CAPÍTULO 3	134
B.1	Tabelas da Análise de Cluster	134
	APÊNDICE C – CAPÍTULO 4	136
C.1	Resultados dos Modelos Hierárquicos Hedônicos	136
	APÊNDICE D – CAPÍTULO 5	140
D.1	Derivadas Parciais dos Modelos de Crescimento Exponencial	140
D.2	Derivação dos Pontos de Inflexão	142
D.3	Estimativas dos Parâmetros dos Modelos	145
	ANEXO A – ROTINAS	152
A.1	Rotinas R: Segmentação de Mercado	152
A.2	Rotinas SAS: Modelo Hedônico Hierárquico	154
A.3	Rotinas SAS: Modelos de Crescimento Exponencial	157

Lista de Figuras

FIGURA 2.1 – Conjunto de Peças de um Veículo	22
FIGURA 2.2 – Estrutura de um projeto Scrapy	24
FIGURA 3.1 – Soma de Quadrados e Número de Clusters	43
FIGURA 3.2 – Distribuição do Preços por Cluster	46
FIGURA 3.3 – Distribuição dos Quilômetros por Cluster	47
FIGURA 3.4 – Distribuição da Idade dos Carros por Cluster	47
FIGURA 3.5 – Distribuição da Potência dos Carros por Cluster	48
FIGURA 3.6 – Distribuição da Soma dos Itens Opcionais por Cluster	48
FIGURA 3.7 – Distribuição do Tempo de Anúncio por Cluster	49
FIGURA 3.8 – <i>Ternary Plot</i> envolvendo Quilometragem, Preços e Idade	49
FIGURA 3.9 – <i>Ternary Plot</i> envolvendo Quilometragem, Preços e Potência	50
FIGURA 3.10 – <i>Ternary Plot</i> envolvendo Quilometragem, Preços e Ítems de Luxo	50
FIGURA 5.1 – Evolução Temporal da Depreciação de Valor	83
FIGURA 5.2 – Preço e Idade de um Bem Durável	91
FIGURA 5.3 – Curvas de Crescimento Exponencial	94
FIGURA 5.4 – Evolução dos Preços do Audi-A6 (Perua), Land Rover-Discovery (Utilitário), Toyota-Hilux (Picape) e Mercedes-Benz-SLK (Convertível)	101

FIGURA 5.5 – Evolução dos Preços dos Veículos Kia-Carnival (Minivan), Toyota-Corolla (Sedã), Fiat-Palio (Perua) e Volkswagen-Passat (Cupê)	102
FIGURA A.1 –Histograma do Logaritmo dos Preços por Tipo de Vendedor	127
FIGURA A.2 –Histograma dos Quilômetros Rodados por Tipo de Vendedor	128
FIGURA A.3 –Box-Plot dos Preços dos Veículos Usados por Ano	128
FIGURA A.4 –Box-Plot do Preços por Potência e Tipo de Vendedor	132
FIGURA A.5 –Box-Plot do Preço por Marca e Potência Abaixo de 1.4	132
FIGURA A.6 –Box-Plot do Preço por Marca e Potência Entre 1.4 e 1.6	133
FIGURA A.7 –Box-Plot do Preço por Marca e Potência Acima de 2.0	133
FIGURA B.1 – <i>Ternary Plot</i> do Tempo de Anúncio, Idade e Ítens de Luxo por Cluster	135

Lista de Tabelas

TABELA 2.1 – Preço, Quilometragem e Idade dos Veículos	25
TABELA 3.1 – Sistemas Internacionais de Classificação de Veículos	39
TABELA 3.2 – Distribuição das Categorias de Carros Através dos Clusters	44
TABELA 3.3 – Características dos Clusters: Tamanho, Preço Médio, Quilometragem Média, Potência Média, Idade Média	45
TABELA 3.4 – Caracterização dos Clusters	51
TABELA 4.1 – Matriz de Correlação de Pearson	70
TABELA 4.2 – Estrutura dos Modelos Hierárquicos Hedônicos - MHH	72
TABELA 4.3 – Estimativas dos Modelos Hedônicos Hierárquicos com Dois Níveis	76
TABELA 5.1 – Modelos Não-Lineares de Crescimento Exponencial	95
TABELA 5.2 – Determinação dos Valores Iniciais	98
TABELA 5.3 – Modelos matemáticos não-lineares	99
TABELA 5.4 – Modelos de Crescimento Não-Linear: Logistic	100
TABELA 5.5 – Modelos de Crescimento Não-Linear: Logistic	103
TABELA 5.6 – Modelos de Crescimento Não-Linear: Richards	104
TABELA 5.7 – Gráficos dos Valores Estimados e Observados	107
TABELA 5.8 – Gráficos dos Valores Estimados e Observados	108

TABELA A.1 –Quantidade de Carros, Preços Médios, Quilômetros Rodados e Coeficiente de Variação das Principais Marcas	129
TABELA A.2 –Relação de variáveis disponíveis na Base de Dados	130
TABELA A.3 –Principais Modelos de Cada Montadora na Base de Dados	131
TABELA B.1 –Características dos Itens de Conforto e Segurança nos Clusters	134
TABELA B.2 –Correlações por Clusters e Tipo de Vendedor	135
TABELA C.1 –Elasticidades dos Itens Opcionais para MHH5 com Dois Níveis	137
TABELA C.2 –Modelos de Preços Hedônicos Hierárquicos com Três Níveis	138
TABELA C.3 –Elasticidades dos Itens Opcionais para MHH5 com Três Níveis	139
TABELA D.1 –Derivadas Parciais em relação à β_0	140
TABELA D.2 –Derivadas Parciais em relação à β_1	141
TABELA D.3 –Derivadas Parciais em relação à β_2	141
TABELA D.4 –Derivadas Parciais em relação à β_3	142
TABELA D.5 –Modelos de Crescimento Não-Linear: Exponencial Negativa, Monomolecular e Mitcherlich	146
TABELA D.6 –Modelos de Crescimento Não-Linear: Gompertz	147
TABELA D.7 –Modelos de Crescimento Não-Linear: Von-Bertalanffy	148
TABELA D.8 –Modelos de Crescimento Não-Linear: Weibull	149
TABELA D.9 –Critério de Informação de Akaike e Coeficiente de Determinação	149
TABELA D.10 –Gráficos dos Valores Estimados e Observados	150
TABELA D.11 –Gráficos dos Valores Estimados e Observados	151

1 Introdução

1.1 Objetivo

O mercado secundário de carros é importante por ampliar o espectro de bens disponíveis para os consumidores (GAVAZZA *et al.*, 2014), introduzido uma série de substitutos parciais aos veículos novos (ESTEBAN; SHUM, 2007). A compra do veículo é a segunda principal despesa das famílias, após da aquisição da casa própria (ANDRIKOPOULOS; MARKELLOS, 2015). A existência de um mercado de carros usados possui implicações para as montadoras de veículos, pois em situações de instabilidade econômica, os consumidores podem decidir por um veículo usado, adiando a compra de um carro novo (DAVIS; KAHN, 2010).

Nos EUA e no Brasil, mais de 75% das vendas de carros ocorrem no mercado secundário de veículos (GAVAZZA *et al.*, 2014). As vendas de carros zero quilômetro também podem envolver a negociação de um veículo usado, dado como parte do pagamento (BUSSE *et al.*, 2012). O mercado de carros usados também é importante devido à participação de diversos agentes econômicos, tais como concessionárias, seguradoras, locadoras, financeiras, fabricantes de peças de reposição (BETTS; TARAN, 2006). Apesar da sua importância, o mercado de carros usados ainda é relativamente negligenciado (PRIETO *et al.*, 2015). Nesta tese contribuimos com a literatura estudando o mercado secundário de carros sob três pontos de vista: segmentação de mercado (Capítulo 2), comunalidade (Capítulo 3) e depreciação de preços (Capítulo 4).

A democratização da internet ocasionou mudanças permanentes nos hábitos dos con-

sumidores. O hábito de utilizar a internet produz um rastro de informações digitais, denominado na literatura de *digital footprint* (GARFINKEL; COX, 2009). São informações que se acumulam em servidores de redes sociais, sites de busca, comentários deixados em *blogs*, cadastros em sites de vendas, histórico de consulta por produtos, entre outras possibilidades.

Nos últimos anos ocorreram também importantes avanços em análise de dados e ciência da computação, com o desenvolvimento de novas técnicas de análise, softwares com algoritmos mais velozes, novas arquiteturas de bancos de dados e ferramentas de extração de dados da internet. O acúmulo de informações na internet vem acompanhada de uma nova cultura em análise de dados, chamada de *Big Data*. O advento do *Big Data* trouxe contribuições importantes em diversos setores, tais como Medicina, Biologia, Indústria e Marketing. Entretanto, ainda são escassas as aplicações de Big Data em pesquisa econômica, especialmente em um momento onde os cientistas econômicos são desafiados a adentrar nos aspectos computacionais de *Big Data* como forma de dialogar adequadamente com os demais setores envolvidos na produção de informações (VARIAN, 2014). Este trabalho procura contribuir para a literatura ilustrando como as técnicas de captura de informações e análise de dados típicas de *Big Data* podem contribuir para pesquisa econômica.

1.2 Organização do trabalho

As contribuições do Capítulo 2 incluem um debate sobre o paradigma do *Big Data* em pesquisa econômica, apresentação dos detalhes metodológicos da técnica de *Web Scraping* e a descrição da base de dados resultante.

Contribuímos para a literatura no Capítulo 3, utilizando os anúncios de veículos na internet para segmentar o mercado de carros usados. Fazendo uso da filosofia de *Big Data* e utilizamos um algoritmo não-supervisionado, o qual permite a emergência espontânea de padrões do banco de dados. Os achados mostram a existência de nichos relevantes no mercado de carros usados, não antecipados pelas segmentações de veículos novos. No

lugar das segmentações baseadas em design e tamanho, o mercado de carros é mais bem segmentado utilizando quilometragem, idade, potência e tempo de anúncio.

Buscamos a intuição necessária para estudar os efeitos da comunalidade no mercado secundário de carros no Capítulo 4, percorrendo as literaturas de canibalização, comunalidade, preços hedônicos e modelos hierárquicos. A especificação econométrica resultante permite capturar os efeitos da comunalidade sobre os preços utilizando a marca dos veículos como efeito aleatório. Os parâmetros estimados do modelo hierárquico hedônico mostram que a percepção dos consumidores a respeito da comunalidade sobrevive no mercado secundário.

Contribuímos com a literatura no Capítulo 5, realizando uma modelagem da depreciação de valor utilizando uma classe de modelos de crescimento exponencial, os quais permitem caracterizar a depreciação como um processo não-linear possuindo ponto de inflexão e uma assíntota. Os parâmetros dos modelos de crescimento podem ser utilizados pelos agentes de mercado para embasar importantes decisões de compra e venda em negociações envolvendo carros usados. Os modelos não-lineares também podem ser utilizados para previsão dos preços em função da idade, podendo embasar o estabelecimento de uma idade de elegibilidade em políticas de descarte de veículos antigos. Mostramos que o ponto de inflexão e a estimativa de depreciação variam em função do tipo de veículo. Veículos mais caros tendem experimentar menores depreciações antes do ponto de inflexão em relação aos veículos mais baratos. Consumidores de veículos zero quilômetro irão absorver grande parte da perda de preço no momento da revenda. Consumidores de veículos usados, com idade maior que o ponto de inflexão, não experimentarão grandes perdas de investimento no momento da revenda.

O Capítulo 6 apresenta as conclusões finais.

2 Big Data e Mercado de Carros no Brasil

2.1 Big Data

Garfinkel e Cox (2009) definem *digital footprint*¹ como sendo toda e qualquer informação criada na rede, produzida de forma consciente ou irrefletida, disponível on-line, e que pode ser ligada ao nome real do autor através de algum identificador. São exemplos de atividades que produzem pegadas digitais os comentários em um fórum on-line, uma busca por palavra chave, uma marcação de localização no aparelho celular, preenchimento de um cadastro on-line ou um clique de “like” em um vídeo compartilhado numa rede social. Nos anos recentes, sensores de mensuração foram incorporados ao mundo físico em dispositivos tecnológicos, tais como telefones celulares, *tablets*, *notebooks* e aparelhos de localização em automóveis (MANYIKA *et al.*, 2011). Os rastros digitais passaram também a incluir as informações geradas automaticamente como subproduto da utilização de dispositivos tecnológicos conectados à rede (BOYD; CRAWFORD, 2012).

Uma preocupação das empresas é como utilizar as informações dos rastros digitais associando-as com seus próprios bancos de dados (BOLLIER, 2010; MANYIKA *et al.*, 2011). Pesquisadores de diversas áreas têm afirmado que a análise estatística desses rastros digitais pode melhorar as decisões de negócios das corporações. Nos últimos anos, novas técnicas computacionais têm sido desenvolvidas para enfrentar os desafios de analisar grandes quantidades de informações, tais como algoritmos, softwares de análise, ferramentas

¹rastros digitais

de raspagem de dados na internet e novos conceitos de armazenamento de dados.

Esta abundância de dados permite que as empresas identifiquem com precisão até mesmo os pequenos nichos de mercado, permitindo níveis de discriminação de preços cada vez mais detalhados. Por exemplo, as informações sobre empresas concorrentes podem ser obtidas através de *Web Scraping*², complementando os bancos de dados das firmas com informações disponíveis na internet. Informações sobre clientes potenciais podem não existir nos bancos de dados internos da firma, mas podem ser construídas utilizando os perfis de usuários de uma rede social. Os agentes interessados em coletar essas informações precisarão minerar os dados de interesse em campos não-estruturados, tais como: textos livres, palavras chave, marcações de localização, endereço de IP³, perfil de usuários em redes sociais, dados gerados por sensores automáticos, entre outras possibilidades. Em seguida, todas essas informações precisarão ser analisadas e incorporadas às decisões da firma, utilizando métricas, representações gráficas, classificação em estratos entre outras formas de análise. Nos últimos anos, corporações como Oracle, IBM e Microsoft investiram aproximadamente US\$15 bilhões em softwares especializados em análise de dados. Nos EUA, o setor de tecnologia da informação (TI) e análises de dados cresce acima da média dos demais setores de TI (CUKIER, 2010).

O atual estágio de desenvolvimento tecnológico permite que as empresas armazenem diariamente grandes quantidades de dados com custos baixos. A capacidade per-capita tecnológica mundial de armazenamento de dados praticamente dobra a cada 40 meses (HILBERT; LOPEZ, 2011). Uma versão adaptada da lei de Mooran⁴ aponta que a quantidade de informações digitais armazenada aumenta dez vezes a cada cinco anos (CUKIER, 2010). Com o desenvolvimento algoritmos mais velozes e a diminuição nos custos de armazenamento de dados, a preocupação das empresas se volta para a análise dessas grandes quantidades de dados disponíveis. As organizações precisam lidar agora com grandes volumes de dados que não podem descartar nem ignorar.

²Montagem de bancos de dados utilizando informações existentes em sites da internet

³Endereço de Protocolo de Internet: Rótulo numérico atribuído a cada dispositivo existente na internet

⁴Moore (1965) previu que o número de transistores em um circuito integrado denso dobraria aproximadamente a cada 18 meses anos. Na década de 70, este período foi revisado para dois anos. Esse padrão também foi verificado para a capacidade de armazenamento de dados (HILBERT; LOPEZ, 2011).

O *Big data* representa uma mudança de paradigma em ciências sociais e ciência da computação, implicando na criação de novas linhas de pesquisa e em uma redefinição na forma como é construído o conhecimento (KITCHIN, 2014; VARIAN, 2014). Tradicionalmente, as análises eram feitas em dados com periodicidade bem definida. Os censos eram realizados a cada dez anos, as pesquisas domiciliares, tais como POF⁵ e PNAD⁶, eram feitas anualmente, ou seja, período de tempo de abrangência fixa. As informações eram relativamente organizadas, existindo em bancos de dados estruturados, criados com amostras probabilísticas, obedecendo aos pressupostos de independência e estacionariedade e procurando responder um número de perguntas bem específicas.

Existem várias definições de *Big Data*, algumas destacando os aspectos de desenvolvimento computacional, a natureza dos bancos de dados envolvidos, as ferramentas e técnicas estatísticas, até definições que articulam conceitos epistemológicos (KITCHIN, 2014). De forma geral, podemos caracterizar *Big Data* como um conjunto de procedimentos de análise, formas de processamento e armazenamento de informações, novas ferramentas de manipulação de dados, novos algoritmos, os quais são empregados geralmente em grandes bancos de dados, tendo como objetivo transformar todas as informações disponíveis em ganhos de produtividade.

Uma definição de *Big Data* muito adotada é relacionada ao princípio dos três “V”s: Volume, Velocidade e Variedade (DAVENPORT; DYCHE, 2013):

- (i) Volume: As fontes de informações digitais são de alto volume, obtidas através de registros de transações eletrônicas, de interações entre usuários, sensores automáticos embutidos em máquinas, rastreamento de IP entre outras possibilidades. Os custos de armazenamento de dados são decrescentes e as empresas não têm problemas em acumular informações.
- (ii) Variedade: Os dados estão disponíveis em planilhas, textos livres, códigos HTML e XML, arquivos de imagem ou vídeo, metadados, coordenadas GPS, dados de sensores, arquivos em formato PDF entre outros formatos. Não há muito controle

⁵Pesquisa de Orçamento Familiar.

⁶Pesquisa Nacional por Amostragem de Domicílios.

sobre o formato de entrada dos dados. Por causa desta característica, pode ser necessário utilizar processamento de linguagem natural (BIRD *et al.*, 2013) para se extrair informações de textos livres, ou expressões regulares quando a informação está contida em códigos HTML⁷.

- (iii) Velocidade: Em algumas situações, a construção da informação gerencial precisa ser feita em tempo real. O termo “velocidade” se refere à dinâmica em que os dados são gerados, coletados, processados e apresentados.

A forma de extração dos dados da internet pode influenciar na estrutura de dependência entre as observações e determinar no tipo de análise a ser conduzida (BOYD; CRAWFORD, 2012). Varian (2014) ilustra como algumas ferramentas típicas de Big Data podem contribuir em pesquisa econômica, enquanto Einav e Levin (2014) citam algumas diferenças análises econômicas tradicionais e análises utilizando *Big Data*.

2.2 Mercado de Carro Usados no Brasil

O mercado secundário de veículos é mais movimentado que o mercado de carros novos, representando entre 66% (STOLYAROV, 2002) e 75% (GAVAZZA *et al.*, 2014) de todos os veículos transacionados nos EUA. A existência de um mercado secundário de carros é responsável pela introdução de uma série de substitutos imparciais para o mercado de carros novos, influenciando a estrutura de funcionamento do setor automobilístico (ESTEBAN; SHUM, 2007).

No Brasil a produção nacional de carros ultrapassou 3 milhões de unidades em 2008, consolidando o país como um grande mercado mundial (BARROS; PEDRO, 2011). Em 2012, a indústria automotiva brasileira representou 21% do Produto Interno Bruto Industrial. O setor automobilístico é importante também por dinamizar a economia através do envolvimento de outros setores produtivos (CASTRO *et al.*, 2014).

Em 2013, a frota brasileira, de mais de 3,5 milhões de automóveis, tinha em média

⁷*Hyper Text Markup Language*

8,4 anos de idade, onde 43% dos veículos tinham menos de 5 anos de idade, 39% com idade entre 6 e 15 anos e os 10% dos veículos restantes tinham mais de 16 anos de idade. Desta forma, temos um enorme estoque de automóveis usados formando um importante mercado de revenda, caracterizado por grandes investimentos em campanhas de marketing (RODRIGUES, 2013).

Conforme representação da Figura 2.1, os veículos são compostos por diversos componentes. Visando obter economias de custo, as montadoras utilizam comunalidade para veículos de uma mesma marca (VERBOVEN, 1999) e também entre veículos de marcas distintas (SULLIVAN, 1998). Fabricantes de peças de reposição costumam produzir componentes compatíveis com várias marcas e modelos (OPLAND, 2007). A existência de peças de reposição compatíveis com vários modelos pode afetar a percepção dos consumidores sobre a comunalidade mesmo entre os veículos usados e afetar os preços. Retornaremos nesse assunto no Capítulo 4, onde investigamos os efeitos da comunalidade sobre os preços de veículos usados.



FIGURA 2.1 – Conjunto de Peças de um Veículo

2.3 Raspagem de Dados

O banco de dados analisado nesta tese foi obtido através de procedimentos de *Web Scraping*. Trata-se de uma técnica de busca de informações da Internet e organização

em formato de banco de dados estruturado, por exemplo, planilha Excel, arquivo CSV⁸ ou arquivo de texto. No *Python*, existem várias ferramentas para raspagem de dados da Internet, tais como *BeautifulSoup*, *Django*, *Selenium*, *Requests*, *Mechanize*, *Urllib2* e *Scrapy*.

Cada ferramenta possui vantagens e desvantagens relacionadas à facilidade de aprendizagem, documentação, desempenho e possibilidade de sobrecarga dos sites. Como o *Python* é uma linguagem aberta, foi importante considerar a estabilidade dessas ferramentas no que diz respeito à atualidade da documentação e ausência de erros de execução. Geralmente isto é garantido quando há uma grande comunidade de usuários regulares da linguagem, reportando os erros de execução e trabalhando cooperativamente para encontrar novas soluções. Este é o caso da ferramenta *Scrapy* no *Python*.

Scrapy é uma interface de código aberto incorporada ao *Python*, que permite raspagem de alta performance de informações na internet. A raspagem de dados ocorre nos “bastidores”, portanto, os sites não são inteiramente carregados no computador, mas somente algumas informações de interesse. Para isto, indica-se no código de programação do *Python* quais os caminhos HTML relevantes dentro da página web de interesse. Desta forma, evita-se a sobrecarga dos sites, o que é uma característica desejável. Mais especificamente, foi utilizado nesta tese o *Python* versão 2.7 e *Scrapy* versão 0.24.

A raspagem de dados inicia com a criação de um projeto-*Scrapy*, através do comando no DOS: *scrapy startproject tutorial*, onde “tutorial” é um nome definido pelo usuário. Este comando cria uma pasta contendo a estrutura apresentada na Figura 2.2:

Existem várias páginas na internet especializadas em anúncios de veículos on-line. Estas páginas geralmente possuem uma estrutura onde há uma página principal com a listagem de todos os veículos. Na listagem principal costuma vir uma foto, o ano e o preço de cada carro. Esta página possui várias telas, que podem ser ordenadas por preço ou ano, por exemplo. Ao se clicar em um anúncio específico desta listagem, o usuário é encaminhando para uma segunda página onde aparecem as informações mais detalhadas

⁸CSV: *Comma-Separated Values*

```
tutorial/  
  scrapy.cfg  
  tutorial/  
    __init__.py  
    items.py  
    pipelines.py  
    settings.py  
    spiders/  
      __init__.py  
    ...
```

FIGURA 2.2 – Estrutura de um projeto Scrapy

sobre o veículo, tais como quilometragem, cidade de venda, potência do motor, opcionais, fotos, condições de pagamento e informações do vendedor.

A programação *Scrapy-Python* precisou seguir a arquitetura do site. A navegação das telas principais com a listagem de todos os anúncios de veículos precisou ser realizada através de um *loop* sobre os argumentos do endereço web. Em cada *loop* aplica-se o comando *rules*, inserido dentro de uma classe *CrawlSpider*. Este sistema de leitura foi capaz de trazer os endereços web de todas as subpáginas contendo as informações mais detalhadas de cada veículo, abrindo os sites, anúncio por anúncio e coletando as informações relevantes. Utilizando os comandos *xpath*, foram fornecidos os caminhos HTML onde estavam contidas as informações dentro de cada página HTML, tais como preço, quilometragem e ano. Os caminhos HTML eram direcionados para os campos de variáveis que foram criadas no arquivo *Item.py*. Estas variáveis eram armazenadas em um arquivo CSV, onde podiam ser lidas posteriormente no SAS⁹ ou no R¹⁰.

Conforme literatura de Big Data, a limpeza dos dados é uma etapa importante quando se trabalha com dados coletados na internet, pois podem existir campos inválidos devido a existência de anúncios incompletos ou informações declaradas nos lugares errados nas páginas de internet. A etapa de limpeza foi realizada no *Statistical Analysis Software*, versão 9.0.

⁹SAS: Sistema gerenciador de bancos de dados integrando recuperação de dados, gerenciamento de arquivos, análises estatísticas, gráficos e geração de relatórios.

¹⁰R: Linguagem aberta para análise estatística, desenvolvida originalmente por Ross Ihaka e Robert Gentleman na Universidade de Aukuland.

2.4 Base de Dados

A base de dados de anúncios de veículos obtida por meio de *Web Scraping* nos meses de junho e agosto de 2015, totalizou 190 mil veículos, onde 84,4% se referem a carros usados e 15,6% são carros novos. Nos meses de julho e agosto, 26,5% dos anúncios se repetem, ou seja, trata-se provavelmente de um mesmo carro que não foi vendido, enquanto 73,5% dos anúncios são únicos. A maior parte dos anúncios transforma-se em negociações efetivas de vendas ou são excluídos da internet. O período de tempo mediano que um carro fica anunciado é de 28 dias.

A Tabela A.2 no Apêndice A.1 apresenta uma descrição completa de todas as variáveis disponíveis.

Foi observado que alguns anúncios continham valores incompatíveis, por exemplo, um telefone armazenado no campo de preço. Observou-se também anúncios com informações faltantes nos campos de quilometragem e preço, e por último, campos com valores extremos. Após uma limpeza nos dados, a base passou ter 127.233 carros, sendo 20.348 carros zero quilômetros e 106.885 carros usados (Tabela 2.1).

TABELA 2.1 – Preço, Quilometragem e Idade dos Veículos

	Freq. (%)	Preço (R\$)		Quilometragem		Idade
		Média	Coef. Var.	Média	Coef. Var.	Média
Concessionárias	22.151 (17%)	61.878,6	95,4	56.286,1	82,0	2,1
Lojas	78.478 (62%)	54.282,2	126,7	70.307,1	105,9	4,8
Particulares	26.604 (21%)	47.996,1	94,2	77.145,4	88,0	7,6
Carros Zeros	20.348 (16%)	88.849,0	109,3	-	-	-
Carros Usados	106.885 (84%)	47.711,3	108,1	70.345,1	99,9	5,9
Total	127.233	54.290,3	116,1	70.345,1	99,9	4,9

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

Existem três tipos de anunciantes: concessionárias, lojas e particulares. A maior parte dos anúncios (62%) é feito por lojas, as quais apresentam também a maior variabilidade de preços (126,7%). Os anúncios de particulares correspondem a 21% de todos os anúncios realizados na internet e possuem o menor preço médio (R\$47.996,1), o menor coeficiente

de variação dos preços (94,2), a maior quilometragem média (77.145,4) e a maior idade média (7,6). Nos anúncios de concessionárias constam os carros com menor idade média, menor quilometragem e maior preço médio (Tabela 2.1).

Os preços dos veículos anunciados por particulares possuem maior variância que preços dos veículos anunciados por concessionárias. Os carros vendidos por particulares e concessionárias são ligeiramente menos rodados que os carros vendidos por lojas, entretanto, a quantidade de anúncios feitos por lojas é superior à quantidade de anúncios realizados por particulares e concessionárias.

A Tabela A.1 (Apêndice A.1) mostra as frequências das principais marcas no banco de dados, juntamente com o preço médio, quilômetros rodados, idade média e os respectivos coeficientes de variação. Existe uma grande variabilidade nos preços e na quantidade de quilômetros rodados conforme a marca. As cinco principais marcas em quantidade de carros são CHEVROLET, VOLKSWAGEN, FIAT e FORD, as quais concentram 52% dos anúncios de carros usados. As marcas PORSCHE, LAND ROVER, BMW e MERCEDES-BENZ apresentaram maior preço médio e menor idade média que as demais marcas, mas correspondem somente a 7% dos anúncios de carros na internet.

Os principais modelos são FIAT-PALIO (3192), FORD-FIESTA (2853), HONDA-CIVIC (2664), PEUGEOT-207 (3007) e VOLKSWAGEN-GOL (3361) (Tabela A.3 - Apêndice A.1). De forma geral, é possível identificar três grupos de veículos. Um grupo de veículos com preço médio menor, alta variabilidade nos preços e grande quantidade de modelos, tais como PALIO, FIESTA, CIVIC, PEUGEOT-207 e GOL. Existe um grupo intermediário que também possuem uma quantidade razoável de veículos, tais como TOYOTA, HYUNDAI e NISSAN. Por fim, temos um grupo de montadoras especializadas e com menor frequência de carros, tais como LAND-ROVER, MERCEDES-BENZ, BMW e PORSCHE. Este último grupo possui preços médios variando entre R\$ 84.746,5 e R\$ 218.084,3 (Tabela A.3 - Apêndice A.1).

3 Segmentação do Mercado de Carros Usados

3.1 Introdução

Segmentação de mercado é importante para vantagem competitiva das empresas (SUDARSHAN; WINTER, 1998), possibilitando adoção de diferentes estratégias de vendas para diferentes segmentos de mercado (MAZUR; STANIEC, 2000). A segmentação de mercado permite tratar um mercado heterogêneo como se fosse composto por vários mercados menores (HRUSCHKA, 1986), identificando grupos de consumidores com respostas homogêneas aos lançamentos de produtos e campanhas de marketing (CHERON; KLEINSCHMIDT, 1985; LAUGHLIN; TAYLOR, 1991; CHAN, 2008). A identificação desses grupos de consumidores passou a ocupar lugar central nas estratégias de marketing modernas, fornecendo as bases teóricas para as campanhas de marketing, desenho de produtos e gestão de clientes (LAUGHLIN; TAYLOR, 1991).

Uma das dificuldades na segmentação de carros usados é a limitada disponibilidade no conjunto de dados geralmente disponíveis para os pesquisadores. Informações sobre o perfil demográfico dos proprietários, histórico de compras de veículos, comportamento ao volante, entre outras informações, geralmente são difíceis de obter. Nesta tese lidamos com a limitação no conjunto de dados disponíveis utilizando uma base de dados construída através de *web scraping*, mesmo assim, não temos informações sobre o perfil dos consumidores, exceto pela variável “tipo de vendedor”.

Contribuímos para a literatura utilizando um algoritmo de classificação não-supervisionado, o qual permite a emergência espontânea de padrões, sem a necessidade de fornecimento de uma classificação prévia. A principal diferença entre segmentação de carros usados e carros zero quilômetro encontra-se nas variáveis que determinam os clusters. Enquanto os segmentos mais gerais de carros zero quilômetro são determinados principalmente pelo tamanho e design, os segmentos de carros usados são explicados pela quilometragem, idade e potências dos veículos.

A análise de cluster responderá se os segmentos de carros usados possuem correlação com as classificações pré-existentes de carros novos. Demonstraremos também em que medida os clusters de carros usados são caracterizados por variáveis como idade, quilômetros e tempo de anúncio na internet. Para compreender melhor este mercado, avaliamos a composição dos clusters em termos de itens opcionais de luxo e analisamos ainda as distribuições do preço, quilometragem e idade dos veículos dentro dos clusters segundo o tipo de anunciante.

Os resultados mostram a existência de nichos de mercado economicamente relevantes, tal como o cluster 1, que possui valor de mercado de R\$ 1.274.273.137 para os veículos anunciados por lojas. Os clusters resultantes podem ser facilmente caracterizáveis em termos de preço, quilometragem e idade. Por exemplo, o cluster 7 é composto por veículos populares, contendo baixa relação entre quilometragem e idade, enquanto o cluster 3 é formado por veículos de luxo de baixa quilometragem.

A distribuição de preços entre os clusters e entre os tipos de vendedores sugere a inexistência de arbitragem nesse mercado. Utilizando os critérios de preço e quilometragem, foi possível identificar os clusters que são capazes de canibalizar o mercado de carros zero. As segmentações do mercado de carros zero, baseadas em design e tamanho, parecem que não capturam os segmentos de veículos usados (Tabela 3.2). Informações como quilometragem, idade e potência são capazes de distinguir mais adequadamente os segmentos de carros usados (Gráficos 3.8 e 3.9). Os resultados encontrados sugerem que as montadoras deveriam considerar o posicionamento que os veículos zero quilômetro terão no mercado secundário e qual potencial de canibalização que esses veículos podem exercer sobre os

carros novos.

A literatura de segmentação de mercado analisou, anteriormente, o mercado de automóveis considerando a escolha dos consumidores entre veículos à gasolina e a diesel (VERBOVEN, 2002), rivalidade de projetos de produção dentro das montadoras (VERBOVEN, 1999), mudança de comportamento dos consumidores na escolha entre carros e outros meios de transporte (DARGAY; HANLY, 2007), identificação de parâmetros de demanda utilizando a segunda melhor escolha de veículo (BERRY *et al.*, 2004), o poder da marca e escolha entre automóveis gêmeos (SULLIVAN, 1998), classificação de negociantes de veículos em concessionárias (MAZUR; STANIEC, 2000) e avaliando a lei do preço único no mercado de carros Europeu (LUTZ, 2004). Em modelos de seleção adversa, onde os agentes de mercado passam de vendedores a compradores, a literatura avaliou a intensificação do efeito de seleção adversa com o envelhecimento dos carros (KIM, 2000; GOLDBERG; VERBOVEN, 2001). Entretanto, apesar de existirem muitos artigos analisando o mercado de carros novos, ainda existe uma lacuna de estudos sobre o mercado de carros usados (GAVAZZA *et al.*, 2014).

Na Seção 3.2 apresentamos uma revisão da literatura em segmentação de mercado, onde justificamos nossa clusterização utilizando características dos produtos, ao invés de características dos consumidores. Na Subseção 3.2.3 analisamos o modelo de segmentação de bens duráveis de Waldman (1996), onde discernimos implicações da segmentação para canibalização de produtos. Na Subseção 3.2.4 mostramos que as segmentações e sistemas de classificações de veículos são baseados em tamanho e design. Em seguida, apresentamos os dados de utilizados na análise de cluster (Seção 5.4), comentamos sobre o algoritmo não-supervisionado (Seção 3.4), analisamos os clusters obtidos (Seção 3.5) e apresentamos as conclusões finais do capítulo (Seção 3.6).

3.2 Revisão da Literatura

A maior parte das análises de segmentação de mercado tem utilizado dois procedimentos sequenciais (WANG, 1996): (i) Identificação dos segmentos; e (ii) Determinação

da importância econômica de cada segmento (ARABIE *et al.*, 1981; CHAN, 2008). Os métodos de segmentação podem envolver uma etapa de classificação à priori, utilizando alguma característica já existente. Entretanto, quando os grupos não podem ser facilmente identificados à priori, ou quando tais classificações não correspondem aos anseios dos pesquisadores, então se utilizam técnicas estatísticas de clusterização (YANKELOVICH; MEER, 2006). A atratividade econômica dos segmentos pode ser dada, por exemplo, pela identificação de clientes potenciais, clientes fiéis, ocasionais, insatisfeitos ou consumidores com renda mais alta.

As abordagens de segmentação de mercado encontram novos desafios na atualidade, uma vez que as novas tecnologias e a variedade de novos produtos disponíveis já saciaram as demandas mais imediatas dos consumidores (YANKELOVICH; MEER, 2006). As empresas precisam lidar com a identificação de atributos mais subjetivos (CHAN, 2008), onde as preferências dos consumidores apresentam maior variabilidade, não sendo facilmente ordenadas por atributos de preço-qualidade (ABEELE *et al.*, 1990). Os segmentos de mercado não se deixam mais capturar facilmente por informações demográficas, econômico-sociais, geográfica ou pelas tipologias psicográficas (DATTA, 1996; YANKELOVICH; MEER, 2006). A expansão do acesso à internet também fez surgir novos hábitos de consumo, independentes de idade ou fronteiras geográficas, enfraquecendo as abordagens de segmentação baseadas em dados demográficos (YANKELOVICH; MEER, 2006).

3.2.1 Segmentação Industrial

No presente trabalho, devido à inexistência de informações sobre o perfil dos consumidores, buscou-se inspiração na literatura de segmentação industrial e em sua ênfase na utilização de fontes alternativas de dados. A base de dados possui anúncios realizados por lojas, particulares e concessionárias. Estes agentes de mercado compartilham a característica identificada por Kim (2000), revezando constantemente sua posição entre vendedores e compradoras de veículos. Ao segmentar os anúncios de vendas da internet, estamos caracterizando indiretamente os consumidores existentes neste mercado, os quais

podem ser particulares, lojas ou concessionárias. Ao identificar os clusters de produtos homogêneos, estamos indiretamente identificando os grupos de consumidores homogêneos associados a cada cluster (CHAN, 2008).

Apesar de existirem algumas similaridades entre segmentação industrial e segmentação de consumidores (CHERON; KLEINSCHMIDT, 1985), a segmentação industrial encontra maior desafio ao lidar com a complexidade dos processos decisórios de clientes industriais (YANKELOVICH; MEER, 2006), tendo que discernir entre variáveis macro e micro, ou entre variáveis sobre o perfil ou o processo dos clientes industriais (SUDARSHAN; WINTER, 1998). Ao utilizar informações corporativas, muitas vezes é necessário percorrer várias camadas de informações dos clientes industriais (SHAPIRO; BONOMA, 1984), por exemplo, qual a composição acionária dos clientes, quais os processos produtivos, qual o perfil dos gerentes, qual a posição de mercado dos clientes industriais e quais são segmentos finais de consumidores. Para levantar esta quantidade de informações e estabelecer abordagens de segmentação mais personalizadas é necessário incorrer em altos custos de coleta de informações (ARABIE *et al.*, 1981; SUDARSHAN; WINTER, 1998).

Os altos custos de captura de informações justifica o emprego de fontes alternativas de informações (SIMKIN, 1994; LEESON; SOBEL, 2008). Berry *et al.* (2004) procuram utilizar fontes alternativas de informações para identificação de parâmetros de demanda em uma classe de produtos diferenciados¹. Os autores utilizam o histórico de consulta por produtos para identificar a segunda melhor escolha de veículo, estabelecendo assim as relações de preferências dos consumidores no mercado de automóveis.

Fornecedores de bens duráveis têm interesse em identificar quais as decisões gerenciais e processos produtivos das empresas-clientes se manifestam em adoção de novas tecnologias (LAUGHLIN; TAYLOR, 1991), utilizando segmentação de mercado para desenhar produtos que atendam às demandas mais específicas (SUDARSHAN; WINTER, 1998). Na Subseção 3.2.3 apresentamos brevemente um modelo de mercado composto por um monopolista que fornece para várias empresas, as quais competem pelo mercado de consumidores

¹Nesse modelo, os produtos são vistos como uma coleção de características e os consumidores escolhem a coleção de características que maximizam sua utilidade.

fnais. Da mesma forma, [Plank \(1985\)](#) e [Laughlin e Taylor \(1991\)](#) estudam um mercado composto por um monopolista fornecedor de bens duráveis para empresas-clientes, as quais comercializam produtos diversificados em um ambiente mais competitivo. Este mercado guarda algumas similaridades relevantes com o mercado de veículos usados, onde existe uma pequena quantidade de montadoras com alta concentração de mercado (Tabela [A.1](#)). Por outro lado, o mercado de revenda é formado por uma grande quantidade de agentes, muitos dos quais comercializam veículos novos e usados.

A literatura sobre segmentação de automóveis tem-se concentrado no mercado de carros zero, apesar da importância do mercado secundário de automóveis, são mais escassos os trabalhos sobre o mercado de carros usados ([PRIETO et al., 2015](#)). A segmentação do mercado de carros justifica-se pelo reconhecimento de variáveis inexistentes no mercado de carros novos: quilometragem e idade. Desta forma, as segmentações elaboradas para os carros zero não necessariamente acomodam o mercado de carros usados ([DINIS; SILVA, 2015](#)), onde a variabilidade na quilometragem e na idade pode revelar novos clusters perfis de consumidores. Esses segmentos de carros usados são de interesse para os agentes econômicos atuantes nesse setor, tais como concessionárias, consumidores, seguradoras, locadora de veículos, fornecedores de peças de reposição e oficinas de reparo e manutenção.

3.2.2 Importância Econômica da Segmentação

Apesar de existirem vários métodos de segmentação de mercado, não existem muitas bússolas para avaliação do custo-benefício de cada segmento. A literatura têm apontado problemas entre a identificação de clusters em uma base de dados e os custos de uma estratégia de marketing adequada ([CHAN, 2008](#)).

[Kleinberg et al. \(1998\)](#) propõem um método para avaliar a importância econômica de uma segmentação de mercado no contexto de *data mining*, onde o valor das informações extraídas da base de dados é mensurado do ponto de vista microeconômico. As correlações e clusters são relevantes na medida em que podem sugerir ações lucrativas para os tomadores de decisão. A firma enfrenta o problema de escolha das ações y pertencentes

ao domínio de decisões D , que maximizem a função de utilidade $f(y)$: $\max_{y \in D} f(y)$. Assumindo que o domínio de decisão D é endógeno às firmas, a função objetivo $f(y)$ pode ser escrita como:

$$f(y) = \sum_{i \in C} f_i(x)$$

Onde C é o conjunto dos consumidores e $f_i(y)$ é a contribuição do consumidor i para a utilidade da firma. Representando o conjunto de informações de perfil disponíveis para os consumidores por x_i , podemos reescrever o problema de maximização de lucro:

$$\max_{y \in D} \sum_{i \in C} g(y x_i)$$

Utilizando agregação linear, pode-se substituir $\sum_{i \in C} g(y x_i)$ na expressão acima por $g(y, x)$, onde x é alguma agregação linear de x_i .

Supondo que uma revendedora de carros decide dividir o mercado em dois segmentos, aplicando estratégias diferentes de marketing para cada segmento. Vamos representar o espaço de decisão para os dois segmentos de mercado por $D \times D = D^2$. Para cada consumidor i e para cada decisão possível $y \in D$, a revendedora recebe um lucro de $c_i \times y$ no conjunto dos pares de decisões $(y_1 y_2) \in D^2$. A revendedora decide escolher dentro do conjunto de consumidores C os i consumidores para os quais vale a relação: $c_i \times y_1 > c_i \times y_2$. Para estes i consumidores a revendedora irá aplicar a estratégia de marketing y_1 , e para os consumidores restantes a revendedora irá aplicada a estratégia y_2 . Considerando as ações das revendedoras em cada segmento, a maximização de utilidade pode ser representada por:

$$f_i(y) = \max [c_1 \times y_1 \ c_2 \times y_2]$$

Ou seja, a função de utilidade da revendedora envolva agora uma forma de não-linearidade. Podemos representar a maximização de lucros da revendedora ao escolher a estratégia

$(y_1 y_2)$ por:

$$\max_{((y_1 y_2) \in C^2)} \sum_{i \in C} [c_1 y_1 c_2 y_2]$$

Neste caso, em vez de experimentar alguma clusterização arbitrária e depois verificar se tais escolhas geram lucratividade, a revendedora por escolher qual a divisão de clusters mais lucrativa tendo em vista as ações de marketing no conjunto de decisão D .

3.2.3 Mercado de Bens Duráveis

Produtores de bens duráveis costumam delinear suas linhas de produtos segmentando o mercado com base em atributos de valoração de qualidade. As firmas não são capazes de satisfazer todos os consumidores com um único produto. No mercado de carros, a lucratividade dos agentes de mercado se origina em consumidores heterogêneos, com alta valoração dos atributos de qualidade e que estão dispostos a pagar preços maiores pelos carros novos. Entretanto, as dificuldades em identificar os segmentos de mercado impedem que os agentes de mercado compreendam todas as consequências da adoção de uma determinada segmentação de mercado (BERRY *et al.*, 2004).

Canibalização ocorre quando os consumidores de um determinado segmento escolham um produto que foi desenhado para outro segmento (DESAI, 2001). Em tal situação, o planejamento da firma falha em atender adequadamente a demanda de cada segmento de mercado. Do ponto de vista da segmentação, o mercado secundário de automóveis e o mercado de carros novos estão conectados através do processo de canibalização, onde os carros usados pode exercer pressão sobre os preços dos carros novos (ZHAO; JAGPALB, 2006). A literatura reconhece que as políticas de precificação das fabricantes de automóveis novos deveriam considerar a persistência do estoque de carros usados ao longo do tempo (STOLYAROV, 2002). Entretanto, existem relativamente poucos trabalhos que investiguem os efeitos de canibalização do mercado de carros usados sobre o mercado de carros novos, especialmente quando o fabricante de carros novos atende vários segmentos de mercado (ZHAO; JAGPALB, 2006). Uma releitura dos modelos de canibalização podem

fornecer importantes intuições sobre o processo de canibalização entre carros usados e novos, por exemplo, substituindo-se os produtos de baixa qualidade pelos carros usados e os produtos de alta qualidade pelos carros novos. Os modelos de [Waldman \(1996\)](#) e [Erat e Kavadias \(2006\)](#) ilustram uma situação onde a segmentação de mercado está associada à intensificação do processo de canibalização.

[Erat e Kavadias \(2006\)](#) analisam o processo de canibalização em um mercado composto por um monopolista que provê tecnologia para um segmento de clientes-industriais, os quais competem no mercado final. O monopolista decide pela forma de introdução de novas tecnológicas, podendo-se comprometer com uma agenda de lançamentos futuros, adiar os lançamentos de novas tecnologias ou oferecer upgrade. Existem dois segmentos de clientes industriais: (i) Empresas com alta capacidade produtiva, cujo valor da produção é influenciado pela tecnologia adotada; (ii) Empresas com baixa capacidade produtiva, cujo valor da produção não é influenciado pela tecnologia adotada. A estratégia de introdução tecnológica do monopolista irá induzir determinado comportamento dos clientes industriais. Os clientes industriais precisam decidir se adotarão imediatamente as novas tecnologias, considerando os impactos da tecnologia sobre a qualidade dos produtos finais. Os resultados encontrados por [Erat e Kavadias \(2006\)](#) mostram que o monopolista pode escolher preços e níveis de qualidade que induzam a adoção parcial de tecnologia pelos clientes industriais.

[Waldman \(1996\)](#) considera uma segmentação baseada em qualidade de bens duráveis, onde a qualidade é representada pela durabilidade. Existe um monopolista e dois segmentos de consumidores ($j = 1, 2$) neste mercado. O monopolista produz com diferentes durabilidades. O monopolista produz com qualidade $Q > 0$ e custo marginal $c(Q)$, onde $c(0) = 0$, $c'(0) = 0$, $c'(Q) > 0$ e $c''(Q) > 0$. Não existem custos fixos de produção. Novas unidades produzidas são consideradas de maior qualidade que unidades antigas. O atributo de qualidade dos bens está associado à durabilidade ou velocidade de deterioração. Existe um mercado de segunda mão onde unidades antigas são transacionadas. O modelo pode ser contextualizado no mercado de veículos novos e usados.

Seja v_j Q a função de utilidade do consumidor j resultante do consumo de um bem

com durabilidade Q e seja $v_2 > v_1 > 0$, considere que cada indivíduo consome zero ou 1 unidade. Seja um cliente no segmento j e $(P_j^*; Q_j^*)$ o respectivo vetor preço-qualidade. Procurando extrair todo o excedente potencial do consumidor a firma define os preços como sendo: $P_j^* = v_j Q_j^*$ e escolhe o nível de qualidade Q_j^* que maximiza o lucro:

$$\max_{Q_j} v_j Q_j^* - c(Q_j^*) \quad \Rightarrow \quad c'(Q_j^*) = v_j$$

Assim, Q_j^* é o nível de qualidade que maximiza os lucros da firma, ou seja, Q_j^* satisfaz $c'(Q_j^*) = v_j$. Considerando a possibilidade de arbitragem entre os consumidores, os preços dos bens usados limitam os preços dos bens novos. O monopolista não pode escolher preços e quantidade dos produtos novos tratando cada consumidor isoladamente e desconsiderando o mercado de segunda mão.

Para ilustrar esta limitação, suponha que o monopolista decida oferecer para cada consumidor o produto que discrimina preços perfeitamente. Seja o menu de preço e qualidade que extrai todo o excedente do consumidor dado por: (P_1^*, Q_1^*) e (P_2^*, Q_2^*) . Nesta situação o consumidor 2 irá decidir pelo menu (P_1^*, Q_1^*) ao invés do menu (P_2^*, Q_2^*) . Isto ocorre porque consumidor 2 recebe excedente zero ao consumir (P_2^*, Q_2^*) e um excedente estritamente positivo ao consumir (P_1^*, Q_1^*) . Assim vemos que ao decidir os níveis de qualidades dos produtos (Q_1^*, Q_2^*) o monopolista se depara com uma restrição atuando sobre o preço que pode ser cobrado do produto de maior qualidade. Dada a restrição de auto-seleção, os preços ótimos que o monopolista poderá oferecer são $[P_1^*, P_2^* + v_2 (Q_2^* - Q_1^*)]$. Para o produto 2, qualquer preço maior que $P_1^* + v_2 (Q_2^* - Q_1^*)$ resultará na migração dos consumidores no segmento 2 para o produto 1.

Seja $V_1 = \frac{v_2 N}{1 + N}$, onde $N = \frac{n_1}{n_2}$. Considerando um menu de preços-qualidade (P_j, Q_j) adquiridos pelo consumidor j em equilíbrio, temos:

- Se $v_1 > V_1$ temos os seguintes equilíbrios de mercado:
 1. Ambos os consumidores consomem.
 2. $P_1 = v_1 Q_1^*$, $Q_1 < Q_1^*$, $P_2 = v_1 Q_1 + v_2(Q_2 - Q_1)$, e $Q_2 = Q_2^*$.

• Se $v_1 \leq V_1$ então temos os seguintes equilíbrios:

1. Apenas consumidores do grupo 2 compram.
2. $P_2 = P_2^*$ e $Q_2 = Q_2^*$.

Como $P_2 = v_1 Q_1 + v_2(Q_2 - Q_1)$ é uma restrição que atua sobre o preço do produto 2, o monopolista tem incentivo para diminuir a qualidade oferecida para segmento 1, visando aumentar o preço cobrado do consumidor 2. Consequentemente, surge uma distorção na qualidade oferecida para o segmento 1. Esta distorção de qualidade é dada por: $Q_1 < Q_1^*$. Ao atender os dois consumidores, o monopolista encontra uma restrição nos preços que pode cobrar do consumidor de alto-padrão. Por causa da maior facilidade dos consumidores de alto padrão em substituir entre as unidades de novas e antigas, o preço das unidades antigas no mercado de segunda mão atua como limitador do preço das novas unidades. Devido ao elo entre os preços de bens novos e antigos, o monopolista tem um incentivo para fabricar bens duráveis de menor durabilidade que o ótimo social. Esta redução de durabilidade pode ser tão intensa a ponto de levar o monopolista a produzir com durabilidade zero, eliminando assim o mercado de segunda mão².

É fácil verificar a multiplicação de possibilidades de migração de consumidores entre segmentos de produtos. Por exemplo, considere três segmentos ordenados segundo seu grau de valoração de qualidade: $v_1 > v_2 > v_3$. A migração dos consumidores entre produtos pode apresentar uma dinâmica onde o consumidor pertencente ao primeiro segmento escolha consumir o produto pertencente ao terceiro ou segundo segmento, tornando a análise do modelo teórico acima mais complexa. A multiplicação de possibilidades de migração entre segmentos ressalta a importância de definir segmentos de mercados adequadamente.

²Se o monopolista investe zero em durabilidade, então não existem unidades antigas sendo transacionadas no período seguinte

3.2.4 Sistemas de Classificação de Veículos

Diversas classificações de automóveis foram desenvolvidas ao longo dos anos, categorização de veículos com as finalidades de cobrança de impostos, regulamentação ambiental entre outros. Por exemplo, no Reino Unido, os veículos são taxados com base na potência do motor, peso do carro, tipo de combustível e emissões de poluentes. Um dos exemplos é a classificação *iso 3833-1977*, que define categorias conforme características técnicas, e tem sido aplicada para todos os veículos de circulação rodoviária nos EUA. Por sua vez, a Administração de Segurança de Tráfego em Estradas dos Estados Unidos (NHTSA³) utiliza uma classificação de veículos baseando-se na tara⁴, *curb weight*⁵, na capacidade máxima de armazenamento de combustível, no tipo de óleo utilizado, no sistema de arrefecimento e presença de ar condicionado.

Na Europa, as locadoras de veículos: Avis, Budget, Alamo, Europcar, Hertz, Sixt e National utilizam um sistema de segmentação denominado ACRISS⁶. Essa classificação visa facilitar locações em diferentes países, onde os consumidores escolhem uma categoria de carro ao invés de um modelo específico.

O sistema de codificação ACRISS possui quatro dígitos que identificam: categoria de tamanho, tipo de transmissão, tipo de combustível e presença de ar condicionado (OPLAND, 2007). Outro exemplo é o Sistema de Segmentação da Comissão Europeia⁷ baseado principalmente no tamanho dos veículos e contendo as seguintes categorias: (A) Carros Mini, (B) Pequenos, (C) Médios, (D) Grandes, (E) Executivos, (F) Luxo, (S) Esportivos e (M) Carros Multi-propósito. A Tabela 3.1, retirada de Dinis e Silva (2015), mostra uma tentativa de correlação entre os diferentes sistemas de classificação de veículos.

Além dos sistemas internacionais de classificação de veículos, as montadoras de veículos costumam desenvolver seus próprios sistemas. Entretanto, não é uma tarefa fácil posicionar a classificação de uma montadora dentro dos sistemas internacionais, ou mesmo

³NHTSA: The United States National Highway Traffic Safety Administration

⁴Capacidade de carga em quilogramas

⁵Peso do veículo sem carga

⁶Association of Car Rental Industry Systems Standards

⁷Euro Cars Segments

TABELA 3.1 – Sistemas Internacionais de Classificação de Veículos

	Pequeno/Econômico <i>Small/cheap</i>	Médio <i>Medium</i>	Grande/Luxuoso <i>Large/Expensive</i>
Fabricantes de Automóveis	-Generalistas	-Prestígio	-Especialistas
SMMT: Society of Motor Manufacturers and Traders	-Supermini -Mini	-Médio Baixo -Médio Alto -Executivo	-Luxo -Esportivo Especial - <i>Dual-purpose</i> - <i>Multi-purpose</i>
JAMA: Japanese Automobile Manufacturers Association	-Minicars -Small Cars	-Padrão	-Padrão
SIAM: Society of Indian Automobile Manufacturers	-A1, A2, A3, A4	-A5, A6, B1	-B2, SUV
EU: European Union	-A: Mini carros -B: Carros Pequenos	-C: Médio -D: Grande -E: Executivos	-F: Luxo -G- <i>Multi-purpose</i> -G-Esportivo utilitário
ACAP: Automobile Association from Portugal	-A:Econômico -B:Inferior	-C: Inferior Médio -D: Médio Superior -E: Superior	-F:Luxo -G:SUV -H:MPV

dentro dos sistemas de classificadas das demais montadoras.

A necessidade de maior segurança em acidentes de trânsito resultou em projetos de veículos contendo zonas de deformação eficientes, levando ao aumento de tamanho dos carros (OPLAND, 2007). Paralelamente a isto, a intensificação da competição tem levado a constantes modificações de projeto nos carros, as quais resultaram em aumento de tamanho. Apesar do advento dos minicarros, nas demais categorias houve um considerável aumento de tamanho. Opland (2007) mostra que, desde a década de 70, a Volkswagen manteve intactos os nomes dos veículos, mas aumentando consideravelmente a largura, altura e peso dos carros. Assim, mesmo quando consideramos um único sistema classificação, as diferentes gerações de um mesmo veículo podem pertencer a categorias diferentes.

Dinis e Silva (2015) realizam uma comparação entre as classificações adotadas por alguns agentes do mercado global de automóveis⁸, mensurando o nível de dificuldade dos

⁸International Organization of Automobile Manufacturers, Society of Motor Manufacturers and Tra-

consumidores em situar os carros dentro dos diferentes sistemas de classificação internacionais, classificações das montadoras: Renault, Peugeot e Volkswagen e classificações adotadas pelas locadoras Europcar e Avis. Os consumidores têm muita dificuldade em associar os modelos de veículos dentro de cada classificação e correlacionar as diferentes classificações. Estes resultados combinam as observações de Opland (2007), sobre a existência de ambiguidade nas referências aos “pequenos carros de família” (*small family cars*) encontradas em revistas automotivas.

Diferentes safras de carros usados possuem diferentes designs e inovações tecnológicas. Entretanto, ao serem comercializados no mercado secundário, os veículos usados levam consigo as “etiquetas” classificatórias que receberam quando eram novos. A análise de cluster de veículos usados apresentada neste capítulo apontará para novas dimensões na segmentação de carros usados, avaliando em que medida as classificações pré-existentes conseguem capturar os segmentos de carros zero.

3.3 Dados Utilizados

Na Seção 2.4 são apresentados maiores detalhes sobre a montagem do banco de dados utilizando informações disponíveis na internet.

Existem ao todo 82 marcas distintas de veículos usados. Entretanto, conforme apresentado na Tabela A.1, existem treze marcas de veículos que concentram 88% dos anúncios de veículos usados, sendo elas: Chevrolet (16%), Volkswagen (14%), Fiat (13%), Ford (10%), Peugeot (7%), Honda (6%), Toyota (4%), Renault (4%), Mitsubishi (4%), Citroen (3%), Hyundai (3%), Mercedes-Benz (3%) e Land Rover (2%). As marcas restantes agrupam 12% dos anúncios de veículos usados, formando um grupo pulverizado, contendo desde marcas mais exclusivas do mercado de alto luxo (Porsche e American), marcas que foram extintas do mercado de carros novos (Gurgel e Envemo) e marcas que foram incorporadas por montadoras maiores (Daihatsu e Buick). Por este motivo, consideramos excluir da

ders, Japan Automobile Manufacturers Association, Society of Indian Automobile Manufacturers, Association of European Automobile Manufacturers e Automobile Association of Portugal

análise de clusters as marcas de veículos que possuem menos de 0,5% de participação no mercado de carros usados. Após a exclusão dos veículos zero quilômetro, a base de dados passou a contar com 104 mil observações, as quais correspondem a 38 marcas distintas e 430 modelos de carros.

3.4 Metodologia: Clusters K-Means

Os parâmetros a serem estimados pelo método *K-means* são as posições, os centróides e a posição relativa de cada observação em relação aos centróides. Dado um conjunto de observações x_1, x_2, \dots, x_n , onde cada observação x_i representa um vetor de S características ($s < n$). Os parâmetros do método *K-means* são obtidos através de minimização da função custo dada pela Equação 3.4.1. A função custo é uma soma de quadrados, também conhecida como soma de quadrados dentro dos clusters:

$$\arg_S \min \sum_{j=1}^k \sum_{i=1}^n \|x_i - \mu_j\|^2 \quad (3.4.1)$$

onde, k é o número de clusters, n é o número de observações, μ_j é o j -ésimo centróide e $\|x_i - \mu_j\|^2$ é a função distância.

Primeiro, é necessário definir o número de clusters k , o que é feito analisando graficamente a soma de quadrados para diferentes números de clusters. Isto é realizado no Gráfico 3.1. Após a definição do número de clusters, o algoritmo define centróides iniciais selecionando aleatoriamente k vetores da amostra fornecida. Em seguida, a cada observação é atribuída uma correspondência a um dos vetores centróides segundo critério de proximidade utilizando distância euclidiana.

$$d_{ik} = \sqrt{\sum_{j=1}^p x_i - \mu_k}$$

Se a distância entre uma observação i e o seu centróide k é maior do que a distância entre esta mesma observação e um segundo centróide j , então o algoritmo irá substituir

o centroide da observação i para o cluster j ([HACKELING, 2014](#)). Após a atribuição de cada observação ao centroide mais próximo, calculam-se novamente os valores dos centroides, desta vez utilizando todo o conjunto de dados, não apenas observações adquiridas aleatoriamente. Após a atualização dos valores dos centroides, repete-se o processo de atribuição de cada observação a um dos centroides e atualizam-se novamente os valores dos centroides, repetindo-se o algoritmo um número pré-definido de vezes. O módulo *kmeans* do *R* utiliza como padrão o algoritmo de inicialização de Forgy, que consiste em escolher aleatoriamente t observações e utilizar tais observações para calcular as médias iniciais dos centroides ([LANTZ, 2015](#)).

3.5 Resultados

Apresentamos nesta subseção os resultados obtidos através do método *K-means*. Optou-se por excluir os carros zero quilômetro da análise, uma vez que as variáveis quilometragem e idade não apresentam variabilidade. Para escolher o número de clusters, rodamos a análise de cluster repetidas vezes utilizando diferentes números de cluster ([Gráfico 3.1](#)). Em cada rodada, calcula-se a soma de quadrados dentro dos clusters ([Equação \(3.4.1\)](#)). As somas de quadrados dentro dos clusters são comparadas com a variância total da base de dados, que equivale à soma de quadrados de um único cluster.

Utilizamos dois conjuntos de informações: variáveis “originais” e “construídas”. As variáveis originais foram coletadas diretamente nos anúncios de internet e compreendem: Preço do veículo, Quilometragem, Idade do carro, Tempo de anúncio e Potência do motor. As variáveis construídas são: Preço elevado ao quadrado, Quilometragem elevada ao quadrado, Idade elevada ao quadrado, Razão entre quilômetro e idade e Soma dos itens de luxo e segurança: Air bag, Alarme, Ar condicionado, Sistema de som com CD ou MP3, Computador de bordo, Controle de velocidade, Banco de couro, Desembaçador traseiro, Direção hidráulica, Sensor de estacionamento, Freio ABS, Rodas de liga leve, Regulagem de banco, Retrovisor elétrico, Tração nas quatro rodas, Trava elétrica, Vidro elétrico e Regulagem de volante.

Existe a possibilidade de incluir variáveis categóricas no método *K-means*, desde que estas variáveis sejam transformadas previamente em variáveis binárias. Entretanto, a literatura desencoraja utilização de muitas variáveis binárias no método *K-means* (LANTZ, 2015). Por este motivo, optou-se por somar as variáveis binárias dos itens de luxo e de segurança. Outra justificativa para somar as variáveis binárias dos itens de luxo é dada pela forma de comercialização desses opcionais. Os itens opcionais são geralmente ofertados através de incrementos de uma cesta básica de opcionais, onde as cestas sofisticadas contêm todos os itens das cestas mais simples. Assim, a soma dos itens de luxo e segurança possui sentido econômico, sendo interpretada como uma variável *proxy* para o luxo total do carro.

A decisão pelo conjunto final de variáveis foi feita após vários testes e comparando a razão entre a soma de quadrado dentro dos clusters e a soma de quadrados total. Conforme recomendação da literatura, todas as variáveis, originais e criadas, foram padronizadas utilizando o comando *scale* do *R*.

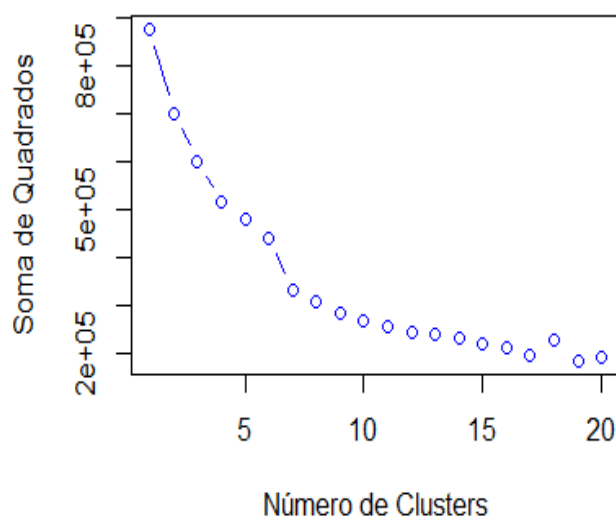


FIGURA 3.1 – Soma de Quadrados e Número de Clusters

O Gráfico 3.1 mostra a soma de quadrados para diferentes números de clusters. A soma de quadrados parece estabilizar após 12 clusters, entretanto, uma quantidade de clusters muito elevada complicaria demasiadamente a etapa posterior de caracterização dos clusters. Optamos por escolher 7 clusters, já que após este limiar o percentual de

queda da soma de quadrados é pequeno. O algoritmo *K-Means* foi ajustado utilizando módulo de mesmo nome no *R*, onde definiu-se 150 interações. A razão entre a soma de quadrado dentro dos clusters (466.547) e a soma de quadrados total (1.052.220) menos 1 é igual a 56%, ou seja, a divisão de clusters é capaz de explicar mais da metade de toda a variabilidade no conjunto de dados.

A Tabela 3.2 apresenta uma comparação entre os 7 clusters obtidos e a categoria de design, enquanto a Tabela 3.3 apresenta a quantidade de veículos por clusters, os preços médios, quilometragem média e idade média.

Segundo informações das Tabelas 3.2 e 3.3, 29% dos Conversíveis estão classificados no cluster 1, que possui preço médio de 30 mil reais, idade média de 3,7 anos e potência 2,0. O cluster 1 abriga 42% dos veículos Cupês e 41% dos Utilitários. Dado o perfil de preços, quilometragem, potência e itens de luxo, nomeamos os veículos do cluster 1 como sendo carros seminovos de alto padrão (Tabela 3.4).

TABELA 3.2 – Distribuição das Categorias de Carros Através dos Clusters

Categoria	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Total
Conversível	2580 29%	1440 16%	949 11%	1716 20%	39 0%	63 1%	1963 22%	8750
Cupê	2542 42%	638 11%	582 10%	1787 29%	21 0%	10 0%	482 8%	6062
Hatchback	9356 24%	5830 15%	7 0%	5779 15%	262 1%	0 0%	18498 47%	39732
Minivan	1299 22%	1139 20%	12 0%	1539 26%	42 1%	0 0%	1805 31%	5836
Perua	1464 36%	637 16%	243 6%	1280 32%	17 0%	0 0%	387 10%	4028
Picape	3415 26%	1797 14%	90 1%	4900 37%	56 0%	0 0%	2899 22%	13157
Sedã	3251 29%	1235 11%	155 1%	2326 21%	82 1%	0 0%	4263 38%	11312
Utilitário	6705 41%	1214 7%	583 4%	6214 38%	33 0%	0 0%	1597 10%	16346

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

O cluster 2 possui veículos com baixos preços médios (R\$ 24.123), alta quilometragem

(120 mil) e antigos (11,8 anos), havendo a possibilidade de abrigar veículos anteriormente utilizados para trabalho, por exemplo, taxis e locadoras. Segundo Tabela 3.2, não existe nenhuma contração percentual de categoria de veículo que sobressaia no cluster 2. O cluster 2 possui alta concentração de itens de luxo mais difundidos: Ar condicionado, Direção hidráulica, Freio ABS, Retrovisor elétrico e Vidro elétrico. Dada todas estas características, nomeamos o cluster 2 como sendo carros sucatas (Tabela 3.4).

Conforme o Gráfico 3.9, o cluster 3 distinguiu-se consideravelmente dos demais clusters em relação ao perfil quilometragem-preço-potência. Segundo a Tabela 3.2, 11% dos Conversíveis e 10% dos Cupês encontram-se no cluster 3. Neste cluster encontram-se veículos luxuosos, caros, de alta potência, porém muito rodados. Nomeamos o cluster 3 de luxuosos de uso diário. (Tabela 3.4).

O cluster 4 possui carros baratos, idade mediana, bastante rodados e comercializados principalmente por lojas e particulares. Segundo a Tabela B.1 (Apêndice B.1), o cluster 4 possui baixo percentual dos opcionais: Computador de bordo, Controle de velocidade, Sensor de estacionamento, Freio ABS, Banco de couro e Tração nas quatro rodas. Nomeamos o cluster 4 como sendo Veículos medianos, não comercializados por concessionárias (Tabela 3.4).

TABELA 3.3 – Características dos Clusters: Tamanho, Preço Médio, Quilometragem Média, Potência Média, Idade Média

Cluster	Freq.	Preço Médio	Quilometragem Média	Idade Média	Potência Média
1	30.612	68.075	31.013	3,7	2,0
2	13.930	24.123	120.491	11,8	1,8
3	2.621	257.132	26.555	5,3	4,4
4	25.541	50.045	86.981	6,5	2,2
5	552	37.786	55.980	7,1	1,6
6	73	110.803	8.134	5,4	4,6
7	31.894	30.249	53.379	5,1	1,4

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

O cluster 5 possui poucos veículos. O perfil de preço, quilometragem, idade e potência de motor não parecem inicialmente distinguir nenhuma das características deste cluster

em relação aos demais (Gráfico 3.8). Entretanto, analisando a distribuição do tempo de anúncio por cluster e tipo de vendedor, vemos que o cluster 5 possui tempo de anúncio bem mais alto que os demais clusters (Gráfico 3.7). Por este motivo, nomeamos o cluster 5 como sendo veículos encalhados (Tabela 3.4).

O cluster 6 forma um segmento de mercado muito específico, possuindo muitos Conversíveis e Cupês. Este cluster possui preço médio de 110 mil, baixíssima quilometragem, abrigando 40% dos veículos com tração nas quatro rodas, 60% dos veículos com banco de couro e 56% dos veículos com computador de bordo (Tabela B.1 - Apêndice B.1). Inicialmente, os clusters 6 e 1 parecem comportar o mesmo tipo de veículo, porém uma análise mais detalhada aos Gráficos 3.2, 3.3 e 3.4 revela que os clusters possuem distribuições distintas de preços, quilometragem, idade, assim como no perfil de vendedor. Todas estas informações permitem caracterizar o cluster 6 como sendo carros de colecionadores, provavelmente o segundo veículo do domicílio.

O cluster 7 é formado por carros populares com preço médio de 31 mil reais e potência média de 1,3 (Tabela 3.3).

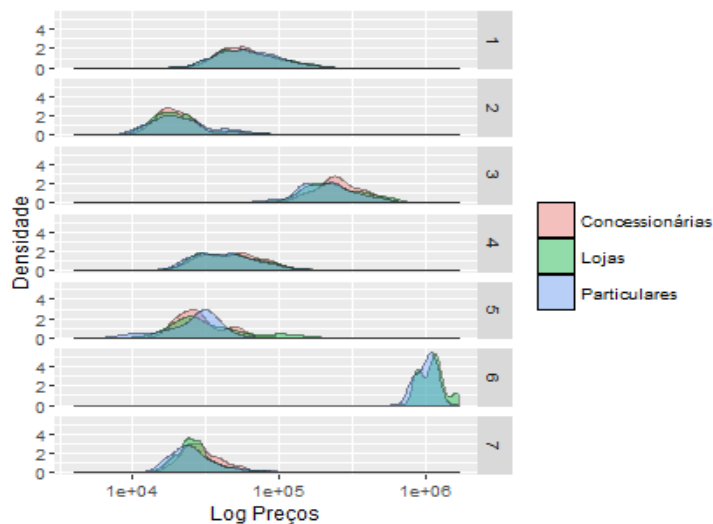


FIGURA 3.2 – Distribuição do Preços por Cluster

As Figuras 3.2, 3.3, 3.4, 3.5, 3.6 e 3.7 apresentam a distribuição dos preços, quilômetros rodados, idade, potência dos clusters, soma dos itens de luxo e segurança e tempo de anúncio na internet. As distribuições são apresentadas para cada cluster, classificando também por tipo de vendedor: Particular, Concessionária e Lojas. As distribuições entre

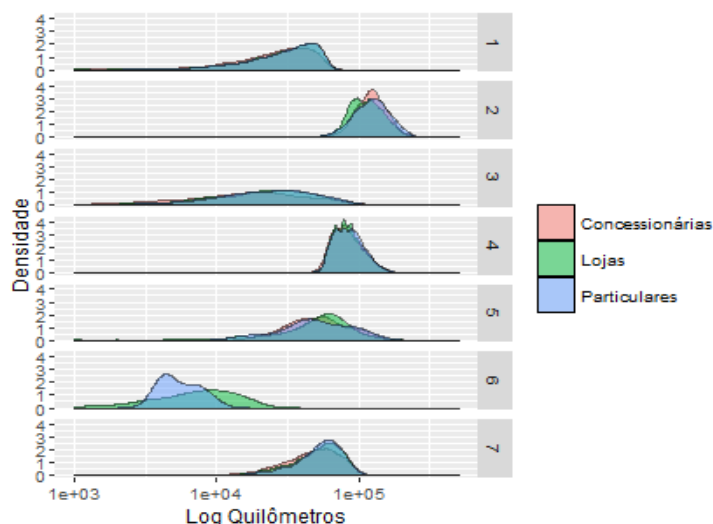


FIGURA 3.3 – Distribuição dos Quilômetros por Cluster

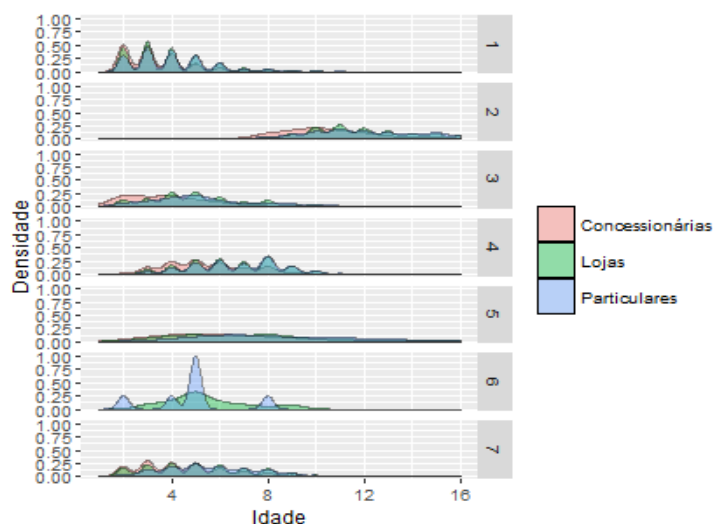


FIGURA 3.4 – Distribuição da Idade dos Carros por Cluster

os tipos de vendedores consideram que não existem grandes oportunidades de arbitragem no mercado de carros usados.

As distribuições dos preços (Gráfico 3.2), quilometragem (Gráfico 3.3) e idade (Gráfico 3.4) diferem consideravelmente entre os clusters, entretanto, não diferem muito por tipo de vendedor dentro de cada cluster.

Os clusters 2 (Carros antigos e baratos) e 4 (Veículos de idade mediana, não comercializados por concessionárias) apresentam distribuições de preço (Gráfico 3.2) e quilometragem (Gráfico 3.3) parecida. Entretanto, as distribuições da idade (Gráfico 3.4), potência (Gráfico 3.5) e itens de luxo e segurança (Gráfico 3.6) são diferentes. Esta impressão é con-

firmada pelos gráficos *ternary plot*, onde vemos que os clusters 2 e 4 são bem distintos em relação ao perfil quilometragem-preço-potência (Gráfico 3.9) e quilometragem-preço-itens de luxo (Gráfico 3.10).

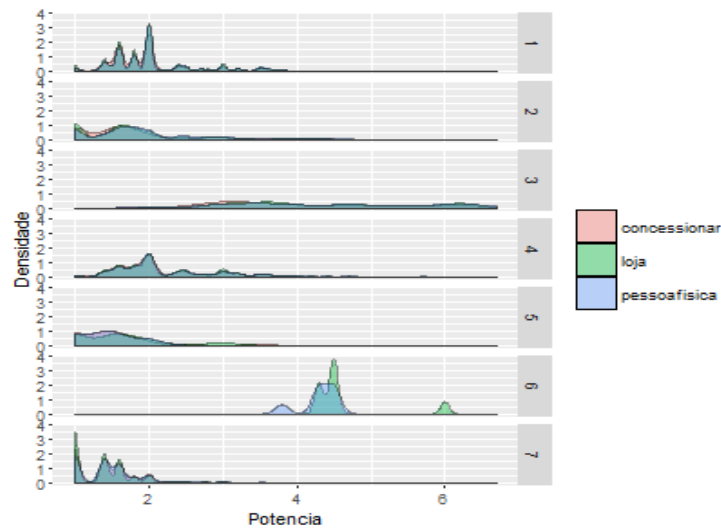


FIGURA 3.5 – Distribuição da Potência dos Carros por Cluster

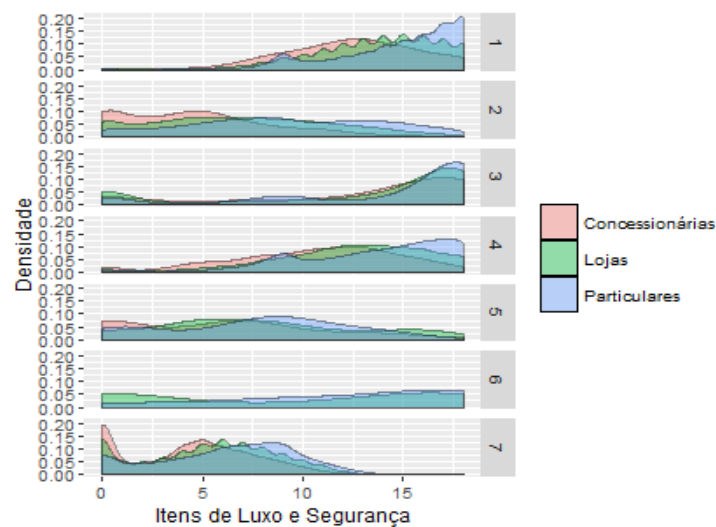


FIGURA 3.6 – Distribuição da Soma dos Itens Opcionais por Cluster

O Gráfico 3.7 revela que os clusters possuem distribuições semelhantes, com exceção do cluster 5, que possui tempo de anúncio consideravelmente maior que os demais clusters.

Os Gráficos 3.8, 3.9, 3.10 e B.1 mostram que as observações dentro de cada clusters possuem razoável separação, uma vez que as cores formam regiões facilmente distinguíveis umas das outras.

O Gráfico 3.8 mostra que o cluster 2 é composto por carros mais antigos com menores

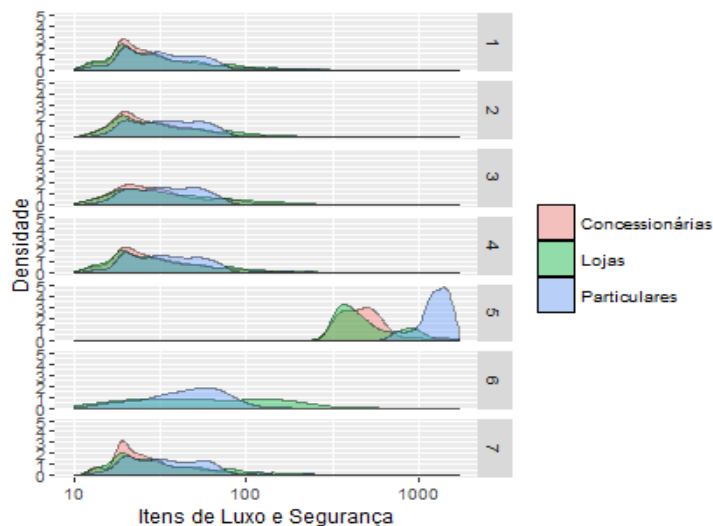


FIGURA 3.7 – Distribuição do Tempo de Anúncio por Cluster

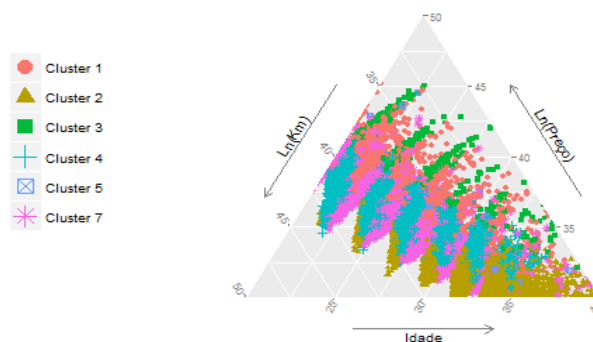


FIGURA 3.8 – Ternary Plot envolvendo Quilometragem, Preços e Idade

preços. O cluster 4 possui preços e idade intermediários, distribuindo-se por todas as idades. De fato, segundo a Tabela 3.3, o cluster 4 possui preço médio de R\$50 mil e quilometragem média de 86 mil quilômetros, enquanto o cluster 2 possui preço médio de R\$ 25 mil e quilometragem média de 123 mil.

Utilizando a intuição obtida em Kleinberg *et al.* (1998), apresentamos na Tabela B.2 (Apêndice B.1), as correlações existentes por cluster e tipo de vendedor, acompanhadas do valor total de mercado para cada segmento e da frequência de veículos dentro de cada segmento. Desta forma, pretendemos avaliar quais as correlações estão associadas a valores de mercado relevantes.

Considerando os anúncios feitos por concessionárias, os segmentos de mercado relevantes são os clusters 1 (seminovos de alto padrão) e 7 (veículos populares), os quais

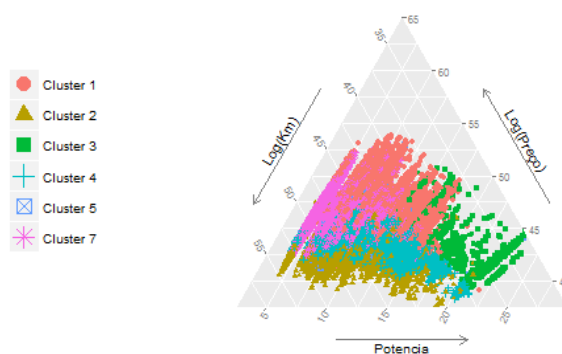


FIGURA 3.9 – Ternary Plot envolvendo Quilometragem, Preços e Potência

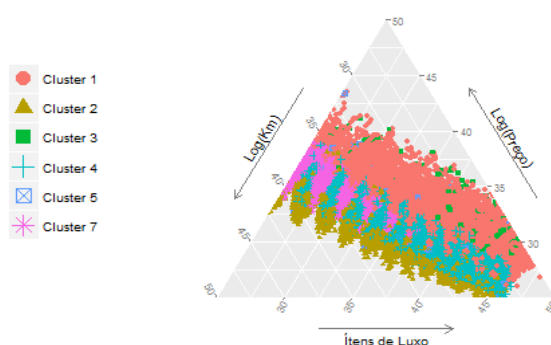


FIGURA 3.10 – Ternary Plot envolvendo Quilometragem, Preços e Ítens de Luxo

representam, respectivamente, R\$ 226.501.003 e R\$ 203.774.910 de valor total. O cluster 1 possui uma quantidade pequena de veículos, sugerindo certa facilidade de alcance de campanhas de marketing direcionadas. Os anúncios feitos por lojas possuem o cluster 6 como segmentos de mercado relevante, o qual apresenta, simultaneamente, baixa frequência e alto valor médio por carro (R\$ 1.118.165).

Enquanto correlação entre preços e quilometragem na base de dados é de -28%, as correlações desagregadas por cluster e tipo de vendedor revelam variações por cluster e tipo de vendedor. Os anúncios feitos por Lojas, no cluster 3, possuem correlação de -45,2% entre idade e quilometragem, enquanto no cluster 6 a correlação é de -51,3%. Os veículos anunciados por Particulares no cluster 5 possuem correlação entre idade e quilometragem de -53,6%. As correlações entre quilometragem e idade são negativas, exceto para os veículos classificados no cluster 3 anunciados por Lojas e Concessionárias (Tabela B.2 - Apêndice B.1). A soma dos itens de luxo e segurança também apresentou correlações relevantes com o preço, por exemplo, para veículos anunciados por Concessionárias

(62,4%), Lojas (57,0%) e Particulares (91,4%) classificados no cluster 5.

3.6 Conclusões

A análise de clusters mostrou a existência de nichos relevantes no mercado de carros usados, tais como os carros luxuosos de uso diário (cluster 3) e os carros de colecionadores (cluster 6), que possuem veículos de alta potência e preços elevados. Apesar das semelhanças, estes clusters diferem em termos de tamanho de mercado, sugerindo a possibilidade de fusão em uma campanha de marketing.

A análise de cluster revelou um cluster de veículos encalhados com tempo de anúncio excessivamente superior aos demais clusters (Gráfico 3.7). Por exemplo, os carros antigos, baratos e não comercializados por concessionárias (cluster 2) também possuem alta quilometragem, mas o tempo de anúncio não é tão prolongado quanto no cluster 5 (veículos encalhados).

Com base em todas as observações feitas para os clusters do mercado de carros, apresentamos a Tabela 3.4, que contém uma tipologia para cada um dos clusters. Na tabela abaixo optamos ainda por ordenar os clusters pelo seu valor de mercado.

TABELA 3.4 – Caracterização dos Clusters

Cluster	Característica	Valor de Mercado
1	Seminovos de alto padrão comercializados por lojas.	2.083.918.390
4	Veículos medianos comercializados por lojas e particulares.	1.278.189.927
7	Veículos populares.	964.762.061
3	Luxuosos de uso diário.	673.942.615
2	Sucatas: Carros antigos e baratos.	336.029.373
6	Cupês e conversíveis de final de semana: Segundo carro do domicílio.	80.885.900
5	Encalhados	20.858.102

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

De forma geral, a análise de cluster revelou ainda que as classificações mais tradici-

onais desenhadas para o mercado de carros zero quilômetro (Tabela 3.2) não capturam adequadamente os segmentos revelados no mercado secundário. Entretanto, devido às limitações de informações na base de dados, não foi possível comparar os clusters obtidos com outras classificações de veículos (Tabela 3.1).

Também ficou evidente que o algoritmo de classificação não-supervisionado possui sensibilidade suficiente para separar os carros de colecionadores (cluster 6) de outros clusters que também possuem valor elevado. Da mesma forma, foi possível separar os veículos encalhados (cluster 5) de outros carros com o mesmo perfil de idade-preços-quilometragem (cluster 7), mas com tempo de venda mais curto.

Considerando os critérios de preço e quilometragem, os clusters 6 e 1 possuem os maiores potenciais de canibalização do mercado de carros zero quilômetro. É claro, estas conclusões não levam em conta as mudanças de design e status que os veículos mais recentes possuem (ABRAHAM *et al.*, 2015), economia de consumo de combustível (VERBOVEN, 1999), fidelidade por marca (SULLIVAN, 1998) e durabilidade (QI; SAWHILL, 2014), geralmente associados pelos consumidores aos veículos zero quilômetro.

4 Modelos de Comunalidade e Canibalização

4.1 Introdução

A demanda dos consumidores por preços mais baixos e maior variedade de produtos tem impellido os fabricantes a utilizarem processos produtivos baseados em comunalidade para obter economias nos custos na fabricação de produtos diferenciados ([SALVADOR *et al.*, 2002](#)). Por isto, a comunalidade se tornou um importante instrumento de gestão na fabricação de produtos diversificados em um ambiente de crescente demanda por produtos customizados ([WAZED *et al.*, 2010](#)).

Na indústria automotiva, as montadoras General Motors ([DESAI, 2001](#)), Volkswagen ([HEESE; SWAMINATHAN, 2006](#)) e Nissan ([KIM; CHHAJEDB, 2000](#)) relatam obtenção de economias de custos devido ao uso de comunalidade. Os veículos Toyota-Camry e Lexus-ES300 ([DESAI, 2001](#)) e Volkswagen-Passat e Audi-A4 ([HEESE; SWAMINATHAN, 2006](#)) compartilham a mesma plataforma, apesar de serem produzidos por montadoras diferentes.

Entretanto, um elevado grau de comunalidade pode agravar o efeito de canibalização ([SRINIVASAN *et al.*, 2005](#)), pois afeta a percepção de exclusividade dos produtos ([CRATTY; SAHUTSKE, 2003](#)), diminuindo a capacidade das firmas de extrair prêmios devido à diferenciação ([DESAI, 2001](#)). A literatura relatou que canibalização pode levar à retirada prematura de produtos do mercado ([RUEBECK, 2005](#)) e ao fracasso no lançamento de novos modelos de carros ([HEERDE *et al.*, 2012](#)), sendo este um importante efeito a ser

considerado. Por último, sabe-se que os carros usados são substitutos parciais dos carros novos (ESTEBAN; SHUM, 2007), conseqüentemente, as montadoras de automóveis podem sofrer canibalização do mercado secundário de carros.

Apesar de sua importância, ainda são relativamente escassos os trabalhos empíricos sobre os efeitos da comunalidade (WAZED *et al.*, 2010) e da expansão na variedade de produtos (DESAI, 2001). A literatura também tem se concentrado no mercado de veículos zero quilômetro, negligenciando o mercado de carros usados (PRIETO *et al.*, 2015). Contribuímos com a literatura averiguando empiricamente os impactos da comunalidade sobre os preços de veículos usados. Os resultados obtidos mostram efeitos significantes da comunalidade sobre os preços dos veículos usados, mesmo após a inclusão de variáveis importantes na explicação dos preços, tais como idade, potência e quilometragem.

Após uma revisão do modelo de canibalização e comunalidade de Kim e Chhajedb (2000) (Seção 4.2), preços hedônicos (Seção 4.3.1) e modelos hierárquicos (Seção 4.3.2), apresentamos uma especificação econométrica capaz de capturar os impactos da comunalidade sobre os preços de carros usados (Seção 4.3.3). Na Subseção 4.5 apresentamos a análise dos modelos hierárquicos hedônicos, onde o efeito aleatório da marca dos veículos é interpretado como sendo o efeito da comunalidade. Na Seção 4.6 apresentamos as conclusões finais do Capítulo 4.

4.2 Revisão da Literatura

Visando consolidar nossa percepção a respeito da comunalidade e canibalização, nesta seção apresentamos uma revisão da literatura baseada no modelo de Kim e Chhajedb (2000). Os leitores mais ansiosos podem seguir adiante, retomando a leitura a partir da descrição dos dados (Seção 4.4), análise dos resultados (Seção 4.5) e conclusões finais (Seção 5.6).

4.2.1 Comunalidade e Canibalização

Kim e Chhajedb (2000) estimam que 80% dos custos de produção, 50% da qualidade, 50% do tempo de espera por pedidos e 80% dos negócios são influenciados pela complexidade de design dos produtos, bem como pela variedade de produtos. A utilização de modularidade na produção possibilita às firmas atenderem às necessidades de clientes heterogêneos e obter economias de custo.

Entretanto, o uso da comunalidade pode tornar semelhantes produtos que foram originalmente desenhados para mercados distintos e afetar a utilidade dos consumidores (DESAI, 2001). Se os consumidores acharem suficientemente atrativo podem escolher produtos que foram originalmente projetados para outro segmento, caracterizando o processo de canibalização (MOORTHY; PNG, 1992).

Os modelos de comunalidade geralmente consideram um mercado composto por um monopolista que serve dois segmentos: “alto padrão” e “baixo padrão”. As valorações dos consumidores de alto padrão v_h e baixo padrão v_l aumentam linearmente com a qualidade. As funções de utilidade dos consumidores de alto e baixo padrão são dadas, respectivamente, por: $v_h q_h$ e $v_l q_l$, onde $v_h > v_l$. O tamanho dos mercados de alto e baixo padrão são dados, respectivamente, por n_h e n_l . Todos os consumidores procuram maximizar o próprio excedente, o qual é definido como a utilidade de um produto com qualidade q_i ; $i \in (h, l)$ menos o preço pago por este produto.

Kim e Chhajedb (2000) adicionam ao modelo de Moorthy e Png (1992) o componente de comunalidade q_m , os fatores de desconto β_d e valoração de utilidade β_p . O custo de produção unitário é crescente em q_i , sendo dado por: $c(q_i) = c q_i^2$, onde c é uma constante. A qualidade geral do bem (q_i) é determinada pelo design de comunalidade (q_m) e pelo design personalizado ($q_i - q_m$). O design de comunalidade afeta a utilidade do consumidor através de um fator de desconto ($0 \leq \beta_d < 1$) e um fator de prêmio ($0 \leq \beta_p < 1$) que modificam a percepção de qualidade dos consumidores.

A presença de comunalidade no produto de baixo padrão torna-o mais parecido com o produto de alto padrão, afetando positivamente a utilidade dos consumidores de baixo

padrão. A presença de comunalidade no produto de alto padrão torna-o mais parecido com o produto de baixo padrão, afetando negativamente a utilidade dos consumidores de alto padrão. A qualidade do produto percebida pelos consumidores de alto padrão será dada por: $q_h - \beta_d q_m$, enquanto a qualidade percebida dos consumidores de baixo padrão será: $q_l + \beta_p q_m$.

Do lado da oferta ocorre uma economia de custo de $\alpha c f(q_m)$ devido à utilização de comunalidade. A função de custo total será dada por: $c [q^2 - \alpha f(q_m)]$, onde $0 \leq \alpha < 1$. A função de economia de custos $f(q_m)$ é não decrescente em q_m e α é um parâmetro de economia de custo.

As firmas produzem com qualidade q_l e q_h e utilizam um design modular comum q_m . Considere a situação onde o monopolista introduz um produto para cada segmento de mercado. Neste caso o monopolista tenta maximizar seus lucros através do sistema:

$$\text{Max}_{P_l, P_h, q_l, q_h, q_m} n_l(P_l - c q_l^2 + \alpha c f(q_m)) + n_h(P_h - c q_h^2 + \alpha c f(q_m)) \quad (4.2.1a)$$

Sujeito à:

$$v_l(q_l + \beta_p q_m) - P_l \geq v_l(q_h - \beta_d q_m) - P_h \quad (4.2.1b)$$

$$v_h(q_h + \beta_p q_m) - P_h \geq v_h(q_l + \beta_d q_m) - P_l \quad (4.2.1c)$$

$$v_l(q_l + \beta_p q_m) \geq P_l \quad (4.2.1d)$$

$$v_h(q_h - \beta_d q_m) \geq P_h \quad (4.2.1e)$$

$$q_h \geq 0, \quad q_l \geq 0, \quad q_h \geq q_m, \quad q_l \geq q_m, \quad q_m \geq 0 \quad (4.2.1f)$$

onde, $v_l(q_l + \beta_p q_m)$ é a função de utilidade dos consumidores de baixo padrão e $v_h(q_h + \beta_p q_m)$ é a função de utilidade dos consumidores de alto padrão. Os excedentes dos consumidores de baixo e alto padrão são dados, respectivamente, por: $v_l(q_l + \beta_p q_m) - P_l$ e $v_h(q_h + \beta_p q_m) - P_h$.

As restrições auto-seleção (Equações (4.2.1b) e (4.2.1c)) asseguram que os consumidores escolham os produtos do próprio segmento de mercado. As restrições de participação (Equações (4.2.1d) e (4.2.1e)) asseguram que os consumidores tenham utilidade positiva

na escolha do produto de seu segmento de mercado.

Para analisar quais restrições serão atendidas com igualdade, suponha que o monopolista tenta extrair todo o excedente do consumidor de alto padrão: $P_h = v_h(q_h + \beta_p q_m)$. Substituindo na restrição de auto-seleção temos:

$$\begin{aligned} v_h(q_h + \beta_p q_m) - v_h(q_h + \beta_p q_m) &\geq v_h(q_l + \beta_d q_m) - P_l \\ v_h(\beta_p q_m - \beta_p q_m) &\geq v_h(q_l + \beta_d q_m) - P_l \\ v_h(q_l + \beta_d q_m) - P_l &\leq 0 \end{aligned}$$

A desigualdade acima mostra que o consumidor de alto padrão possui utilidade estritamente positiva no consumo de um produto de baixo padrão. A tentativa de extrair todo o excedente do consumidor do segmento de alto padrão leva à migração dos consumidores para os produtos de baixo padrão.

Antevendo a canibalização, o monopolista escolhe extrair todo o excedente do consumidor de baixo padrão, ou seja: $P_l = v_l(q_l + \beta_d q_m)$. Substituindo P_l na equação de auto-seleção, temos uma desigualdade envolvendo o preço cobrado do consumidor de alto-padrão: $P_h \geq v_l(q_h + \beta_d q_m)$. Esses são os limites inferiores e superiores do preço que pode ser cobrado do segmento de alto padrão:

$$v_l(q_h - \beta_d q_m) \leq P_h \leq v_h(q_h + \beta_d q_m) - [v_h - v_l](q_l + \beta_p q_m)$$

Rearranjando a desigualdade acima:

$$v_l(q_h + \beta_d q_m) \leq v_h(q_h + \beta_d q_m) - [v_h - v_l](q_l + \beta_p q_m) \quad (4.2.2)$$

$$q_h - q_l - (\beta_p - \beta_d) q_m \geq 0 \quad (4.2.3)$$

$$q_h - \beta_d q_m > q_l + \beta_p q_m \quad (4.2.4)$$

A Equação (4.2.4) mostra a necessidade da firma em assegurar que a qualidade do produto de alto padrão seja maior que a qualidade do produto de baixo padrão: $q_h -$

$\beta_d q_m > q_l + \beta_p q_m$. Substituindo $P_l = v_l (q_l + \beta_p q_m)$ e transformando a Equação (4.2.1c) em uma igualdade:

$$\begin{aligned} v_h(q_h - \beta_d q_m) - P_h &\geq v_h(q_l + \beta_p q_m) - v_l(q_l + \beta_p q_m) \\ P_h &= v_h q_l - q_l[v_h - v_l] - q_m(v_h \beta_p + \beta_p[v_h - v_l]) \end{aligned}$$

Considerando P_l e a possibilidade de canibalização, vemos que o monopolista cobra o maior preço possível do consumidor de alto padrão.

Em [Desai \(2001\)](#), quanto maior a diferença entre v_h e v_l menores os efeitos da canibalização. A introdução dos parâmetros de mudança de valoração β_p e β_d , em [Kim e Chhajedb \(2000\)](#), torna o processo de canibalização mais dependente da avaliação de qualidade do segmento de alto padrão v_l .

De forma geral, a canibalização leva a diminuição da qualidade do produto destinado ao segmento de baixo padrão q_l . Considerando a canibalização exercida pelo mercado secundário no mercado de carros novos, as montadoras poderiam influenciar na qualidade dos carros usados diminuindo a qualidade das peças de reposição ou, numa situação extrema, descontinuando a fabricação de peças de reposição para veículos usados. Outra possibilidade é a introdução de inovações tecnológicas e em design que modifiquem a avaliação de utilidade dos consumidores a respeito dos carros usados.

Em [Kim e Chhajedb \(2000\)](#), valores altos de β_d e β_p aumentam a propensão do consumidor de alto-padrão em migrar para o produto de baixo-padrão. Considerando o mercado de carros novos e usados, podemos afirmar que os parâmetros β_d e β_p são as fontes de canibalização devido à similaridade de produtos. Quanto mais dessemelhantes em termos de design, tecnologia e itens opcionais forem os carros novos em relação aos carros usados, menor será o valor de β_d e maior será a distância entre v_l e v_h .

Na próxima seção vamos apresentar uma breve revisão da literatura de preços hedônicos e modelos de preços hierárquicos, os quais serão utilizados para capturar os efeitos da comunalidade no mercado secundário. As equações de preços dos consumidores de alto e

baixo padrão no modelo de comunalidade de [Kim e Chhajed \(2000\)](#) serão importantes para especificação do modelo hierárquico hedônico na Subseção [4.3.3](#).

4.3 Metodologia

Apresentamos uma breve revisão da literatura de preços hedônicos (Subseção [4.3.1](#)) e modelos hierárquicos (Subseção [4.3.2](#)). Posteriormente, esses modelos serão utilizados na especificação de um modelo econométrico que capture os efeitos da comunalidade no mercado secundário de carros (Subseção [4.3.3](#)).

4.3.1 Modelo de Preços Hedônicos

Os modelos hedônicos procuram estabelecer a relação entre preços de produtos diferenciados e os atributos presentes ([ANGLIN; GENCA, 1996](#)). Os coeficientes estimados são chamados “preços implícitos”, e representam a disposição do consumidor em pagar por um incremento em algum dos atributos ([SILVER; HERAVI, 2003](#)).

Seja uma função de utilidade U dos serviços fornecidos por um carro. Assumindo um processo em dois estágios, onde os serviços dos automóveis são função de um vetor de características z , podemos representar a utilidade U por:

$$U = U(z) \tag{4.3.1}$$

$$z = h(x_1, x_2, \dots, x_N, t) = h(x, t) \tag{4.3.2}$$

onde t indica que a conversão de características físicas em serviços U pode mudar com o tempo devido à tecnologia.

Preferencialmente, o vetor de características observáveis x deve incluir o maior conjunto de informações disponíveis. [Gordon \(1990\)](#) inclui no vetor x o espaço interior, indicador de qualidade, tipo do estofamento, *Ride Quality*¹, espaço de porta malas e

¹*Ride Quality* é o grau de proteção e conforto experimentado pelos ocupantes de um veículo diante de instabilidades na superfície das estradas. Carros que perturbam os ocupantes com tremores, trancos

*Automobile Handling*², tipo de sistemas de freios, tempo de aceleração, desempenho na economia de combustível, custos de reparos e decaimento físico.

Para períodos de tempo curtos pode-se desconsiderar as mudanças tecnológicas representadas por t e incluir os atributos físicos x diretamente na função de utilidade U .

Rosen (1974) considera um problema de equilíbrio de mercado onde os preços dos atributos orientam as decisões de consumidores e vendedores. Os produtos são descritos por um vetor de características z : $z = (z_1, z_2, \dots, z_n)$, onde z_i é a quantidade da característica i presente no produto. Os agentes tratam $p(z)$ como um parâmetro fixo na maximização de utilidade e lucro. O equilíbrio de mercado surge através das interações entre compradores e vendedores, revelando uma função $p(z)$ que indica os preços implícitos das características z_1, z_2, \dots, z_n .

Sejam dois produtos com características iguais e preços diferentes. O consumidor racional irá descartar o produto mais caro. Assumindo que $p(z)$ é uma função contínua que possui derivada de segunda ordem, e sejam z_a, z_b e z_c os valores arbitrários do vetor de características, suponha que $z_a = \frac{1}{\lambda} z_b$ e $p(z_a) < \left[\frac{1}{\lambda} p(z_b) \right]$, onde $\lambda > 1$. Então λ unidades de um modelo com vetor de características z_a possui o mesmo benefício que um modelo oferecendo z_b .

Entretanto, o modelo contendo z_a custa estritamente menos. Supondo $z_a < z_b < z_c$ e $p(z_b) > \delta p(z_b) + (1 - \delta)p(z_c)$, onde $0 < \delta < 1$ e $z_b = \delta z_a + (1 - \delta)z_c$, vemos que a cesta contendo as características z_b pode ser obtida por um preço estritamente menor que $p(z_b)$. No lugar de adquirir z_b , o consumidor escolhe δ unidades de um produto contendo z_a e $1 - \delta$ unidades de um produto contendo z_c . Consequentemente, os produtos localizados nas regiões de concavidade de $p(z)$ não serão viáveis (ROSEN, 1974).

Entretanto, os bens não são divisíveis, não sendo possível obter a combinação: $\delta z_b + (1 - \delta)z_a$, ou seja, os consumidores não podem substituir um determinado produto por uma combinação convexa de frações de outros produtos.

e ruídos possuem uma baixa qualidade de passeio.

²*Automobile Handling* é o ângulo de manobra das rodas de um veículo na execução de manobras transversais.

Sabendo que os consumidores adquirem um único produto com vetor de características z , seja uma função de utilidade estritamente côncava: $U(x; z_1, z_2, \dots, z_n)$, onde x são todos os outros produtos consumidos, o preço de x é igual uma unidade e $p(z)$ é mensurado em termos da renda y : $y = x + p(z)$. A maximização da utilidade sujeita à restrição orçamentária requer que x e z satisfaçam as condições de primeira ordem abaixo:

$$\frac{\partial p}{\partial z_i} = p_i = \frac{\frac{\partial U}{\partial z_i}}{\frac{\partial U}{\partial x}} = \frac{U_{z_i}}{U_x}$$

A utilidade ótima é alcançada adquirindo-se um produto que ofereça a mesma combinação desejada de características.

O gasto que o consumidor está disposto a realizar em (z_1, z_2, \dots, z_n) para alcançar determinado nível de utilidade, dada a renda y , é representado por: $\theta(z; u, y)$. Esta função-oferta define uma família de superfícies de indiferença relacionando z_i com o valor monetário de cada produto.

Os modelos de preços hedônicos podem conter atributos relevantes para os consumidores, mas não disponíveis nos bancos de dados ([ARGUEA; HSIAO, 1993](#)). Por exemplo, os anúncios de veículos na internet não mencionam quais veículos são “primeiro dono”, sendo este um atributo valorizado pelos consumidores devido a expectativa de conservação desses carros. Apesar de não ser mencionado nos anúncios, isto será provavelmente revelado aos consumidores no momento das negociações. Outro exemplo de atributo ocultos que são valorizados pelos consumidores são os carros com ruído excessivo no motor.

[Bajari et al. \(2012\)](#) propõem uma abordagem para tratamento dos atributos não-observados que evoluem no tempo. Os veículos são caracterizados por um vetor x_j , $j = 1, \dots, J$ de atributos observáveis e um escalar ξ_j que representa o atributo oculto que evolui no tempo.

Utilizando o pressuposto de equilíbrio [Rosen \(1974\)](#), os preços podem ser escritos como $p_j = p(x_j, \xi_j)$. Ao se estimar a função hedônica $p(x_j, \xi_j)$ utilizando x_j e omitindo ξ_j , o pressuposto $E[\xi_j | x_j] = 0$ de uma regressão OLS é falso. Isto porque os atributos

não-observados podem evoluir com o tempo, por exemplo, o ruído no motor pode se tornar mais intenso. Para obter estimativas consistentes [Bajari *et al.* \(2012\)](#) utilizam três pré-supostos de identificação:

- (a) Em qualquer momento do tempo o preço pode ser escrito como função dos atributos observáveis e não-observáveis e os resíduos da regressão hedônica contém informação que pode ser utilizada para previsão dos preços.
- (b) A evolução do atributo omitido é um processo de Markov de primeira ordem.
- (c) Os consumidores são racionais e antevem o impacto da evolução temporal dos atributos omitidos nos preços, ou seja, os consumidores não erram sistematicamente ao antever os preços implícitos futuros dos atributos omitidos.

Considere um mercado formado por consumidores e vendedores racionais e um vetor $x_{jt}, t = 1, \dots, T$ contendo K atributos observáveis e um único atributo oculto $\xi_{j,t}, t = 1, \dots, T$ que evolui com o tempo.

Do ponto de vista dos consumidores, o produto j pode ser completamente caracterizado pelo vetor 1 por $K + 1$: $(x_{jt}, \xi_{j,t})$. A incerteza a respeito da evolução de $\xi_{j,t}, t = 1, \dots, T$ pode ser oriunda de duas fontes: (i) Características omitidas que evoluem no tempo e (ii) Preços implícitos das características não observadas que evoluem no tempo. Por exemplo, (i) o ruído no motor pode desaparecer após uma manutenção e (ii) os consumidores podem passar a valorizar menos o atributo “único dono”.

A interação entre compradores e vendedores determina os preços de equilíbrio: $p_j = p(x_{jt}, \xi_{j,t})$, onde $p(\cdot)$ é a função de preços hedônicos. Como não podemos fazer uso do pressuposto: $E(\xi_{j,t}|x_{jt}) = 0$, utilizamos as repetições dos anúncios de vendas ao longo do tempo. O preço p_{jt} é observado nos períodos $t = 1, 2, \dots, T$ e temos um sistema de preços

dado por:

$$\begin{aligned}\ln(p_{j,1}) &= \alpha_1 + \beta_1 x_{j,1} + \xi_{j,1} \\ &\dots \\ \ln(p_{j,T}) &= \alpha_1 + \beta_1 x_{j,T} + \xi_{j,T}\end{aligned}$$

onde $\xi_{j,t}$ segue um processo de Markov de primeira ordem: $\xi_{j,t} = \gamma_{t-1,t}\xi_{j,t-1} + \eta_{j,t-1,t}$, onde $\gamma_{t-1,t}$ $\xi_{j,t-1}$ é a esperança condicional do atributo e $\eta_{j,t-1,t}$ é um componente estocástico do atributo omitido.

O pressuposto de identificação (c) diz que: $E[\eta_{j,t-1,t}|It] = 0$, onde I representa a informação disponível para o comprador no período t . Reescrevendo a equação de preços hedônicos e utilizando as informações dos períodos prévios, podemos eliminar o termo $\xi_{j,t}$. Vamos escrever $\xi_{j,t}$ como função de $\xi_{j,t-1}$ e substituir em $\ln(p_{j,1}) = \alpha_1 + \beta_1 x_{j,1} + \xi_{j,1}$:

$$\begin{aligned}\ln(p_{j,t}) &= \alpha_t + \beta_t x_{j,t} + \xi_{j,t} \\ &= \alpha_t + \beta_t x_{j,t} + \eta_{j,t-1,t} [\ln(p_{j,t-1}) - \alpha_{t-1} - \beta_{t-1} x_{j,t-1}] + \gamma_{t-1,t} \xi_{j,t-1} \\ &= \alpha_t + \eta_{j;(t-1,t)} \alpha_{t-1} + \gamma_{t-1,t} \ln(p_{j,t-1}) - \gamma_{t-1,t} \beta_{t-1} x_{j,t-1} + \beta_t x_{j,t} + \eta_{j;(t-1,t)}\end{aligned}$$

Se $x_{j,t}$ e $\eta_{j,t-1,t}$ estiverem correlacionados, os parâmetros de uma regressão OLS são viciados. Para obter parâmetros consistentes estima-se a equação acima através de mínimos quadrados não-lineares em dois estágios (2SNLS), onde o primeiro estágio é dado por:

$$x_{j,t} = \pi_{0,t,t-1} + \pi_{1,t,t-1} x_{j,t-1} + \pi_{2,t,t-1} \mathbf{w}_{j,t} + v_{j,t,t-1}$$

onde $E[v_{j,t,t-1}|\mathbf{I}_t] = 0$ e $\mathbf{w}_{j,t}$ representa um conjunto de variáveis instrumentais observadas.

O procedimento de estimação em dois estágios proposto por [Bajari et al. \(2012\)](#) não é necessário quando acredita-se que as características ocultas não evoluem no tempo, podendo ser ignorado para períodos de tempo curto. Quando não existe um período de tempo longo, a estimação do modelo hedônico em dois estágios não é possível, sendo esta

a situação deste trabalho.

Na Subseção 4.3.3 vamos fazer uso da intuição do modelo de [Bajari *et al.* \(2012\)](#) e da estrutura dos modelos hierárquicos utilizando marca dos veículos é definida como efeito aleatório. Na próxima seção apresentamos alguns detalhes metodológicos do modelo hierárquico.

4.3.2 Modelo Hierárquico

Introduzimos o modelo hierárquico considerando inicialmente uma única marca j , um único modelo de carro i e uma única variável explicativa. Seja um modelo linear relacionando o logaritmo dos preços ($\ln p_i$) com o logaritmo da quilometragem ($\ln km_i$): $\ln(p_i) = \alpha + \beta \ln(km_i) + e_i$, onde α é o intercepto, β é o coeficiente de inclinação e e_i é o erro. Considerando todos os veículos i e todas as marcas k , podemos escrever um modelo linear relacionando preços e quilômetros para todos os veículos e marcas:

$$\ln(p_i) = \alpha_k + \beta_k \ln(km_i) + e_{ik} \quad (4.3.3)$$

Se não existirem muitas marcas na base de dados, podemos utilizar mínimos quadrados e estimar vários interceptos e coeficientes de inclinação, um para cada marca, obtendo o conjunto de parâmetros α_k e β_k , $k = 1, \dots, n$.

Se existe uma grande quantidade de níveis k , a estimação de um modelo hierárquico é uma alternativa mais compacta e elegante em relação à estimação de um modelo de regressão para cada marca ([SNIJDERS; BOSKER, 2012](#)). Utilizando a marca dos veículos como hierarquia, podemos representar o modelo hierárquico utilizando a mesma notação da Equação (4.3.3). Se quisermos interpretar os resultados do modelo hierárquico para toda a população, e não somente para os dados da amostra, então a literatura recomenda especificar um modelo hierárquico com efeitos aleatórios ([HOX, 2010](#); [DEMIDENKO, 2013](#)).

Vamos considerar o caso de um modelo hierárquico onde não existem variáveis expli-

cativas para os efeitos aleatórios. Substituindo $\alpha_k = \beta_{0k}$ e $\beta_k = \beta_{1k}$ temos os parâmetros:

$$\begin{aligned}\beta_{0k} &= \gamma_{00} + \mu_{0k} \\ \beta_{1k} &= \gamma_{10} + \mu_{1k} \\ E(\mu_{0k}) &= E(\mu_{1k}) = 0 \\ Var(\mu_{0k}) &= \sigma_{\mu_{0k}}^2 \\ Var(\mu_{1k}) &= \sigma_{\mu_{1k}}^2 \\ Cov(\mu_{0k}, \mu_{1k}) &= \sigma_{(\mu_{0k}, \mu_{1k})}\end{aligned}$$

Incluindo quilometragem como variável aleatória, o seu coeficiente de inclinação passa a ser $\beta_j = \beta_{1j} + \mu_{1j}$ e o Modelo (4.3.3) pode ser escrito como:

$$\ln(p_{ik}) = \beta_{0k} + \beta_{1k} \ln(km_{ik}) + e_{0ik} \quad (4.3.4)$$

$$\ln(p_{ik}) = [\gamma_{00} + \mu_{0k}] + [\gamma_{10} + \mu_{1k}] \ln(km_{ik}) + e_{0ik} \quad (4.3.5)$$

$$\ln(p_{ik}) = \gamma_{00} + \gamma_{10} \ln(km_{ik}) + \{\mu_{0k} + \mu_{1k} \ln(km_{ik})\} + e_{0ik} \quad (4.3.6)$$

onde $Var(e_{0ik}) = \sigma_{e_0}^2$. A Equação acima explicita os preços dos veículos em função de um componente de efeito fixo ($\gamma_{00} + \gamma_{10} \ln(km_{ik})$) e de um componente aleatório ($\mu_{0k} + \mu_{1k} \ln(km_{ik})$).

A Equação (4.3.6) requer a estimação de dois efeitos fixos (γ_{00} e γ_{10}) e quatro parâmetros aleatórios ($\sigma_{\mu_{0k}}^2, \sigma_{\mu_{1k}}^2, \sigma_{\mu_{0k}, \mu_{1k}}$ e μ_{1k}). Devido à presença dos componentes aleatórios, precisamos recorrer a procedimentos mais específicos para estimação dos parâmetros (HOX, 2010).

Vamos apresentar os componentes $\sigma_{\mu_{0k}}^2$ e $\sigma_{e_0}^2$ de um modelo hierárquico contendo um efeito aleatório e nenhuma variável explicativa. Este modelo é também denominado de “modelo de componentes de variância” (WOLTMAN *et al.*, 2012) e permite avaliar se real-

mente é necessário utilizar um modelo hierárquico.

$$Cov(\mu_{0k} + e_{0jk}, \mu_{0k} + e_{0jk}) = Cov(\mu_{0k}, \mu_{0k}) = \sigma_{\mu 0}^2 \quad (4.3.7)$$

$$\rho = \frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + \sigma_{e 0}^2} \quad (4.3.8)$$

A Equação (4.3.7) apresenta a covariância entre duas observações i e j pertencentes à mesma marca k , enquanto a Equação (4.3.8) fornece o coeficiente de correlação intraclasse (ICC), ou seja, a correlação entre duas observações i e j pertencentes à mesma marca k .

Um coeficiente de correlação intraclasse muito alto implica que os parâmetros do Modelo (4.3.6) não podem ser estimados por OLS. Representando o modelo hierárquico segundo a notação de Demidenko (2013) e incluindo outras variáveis explicativas como efeitos fixos, temos:

$$\ln(p_{ik}) = \beta_0 + \beta_1 \ln(km_{ik}) + \left[\sum_{h=2}^p \beta_h x_{hik} \right] + (\mu_{0k} + \mu_{1k} \ln(km_{ik}) + e_{0ik})$$

onde, $\sum_{h=2}^p \beta_h x_{hik}$ contém os efeitos fixos. No presente estudo, os efeitos fixos são todas as características disponíveis no banco de dados, e possuem a mesma interpretação dos parâmetros de um modelo de preços hedônicos.

É mais fácil visualizar expressão para as estimativas dos parâmetros utilizando a notação de matrizes bloco-diagonal. Vamos exemplificar considerando duas marcas de veículos “A” e “B”. Por simplicidade de notação, considere que marca “A” contém 3 veículos e a marca “B” contém dois veículos. As matrizes de variância-covariância de A e B são dadas por:

$$V = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

$$A = \begin{bmatrix} \sigma_{\mu 0}^2 + \sigma_{e 0}^2 & \sigma_{\mu 0}^2 & \sigma_{\mu 0}^2 \\ \sigma_{\mu 0}^2 & \sigma_{\mu 0}^2 + \sigma_{e 0}^2 & \sigma_{\mu 0}^2 \\ \sigma_{\mu 0}^2 & \sigma_{\mu 0}^2 & \sigma_{\mu 0}^2 + \sigma_{e 0}^2 \end{bmatrix}$$

$$B = \begin{bmatrix} \sigma_{\mu 0}^2 + \sigma_{e 0}^2 & \sigma_{\mu 0}^2 \\ \sigma_{\mu 0}^2 & \sigma_{\mu 0}^2 + \sigma_{e 0}^2 \end{bmatrix}$$

Seja $I_{(n)}$ uma matriz identidade de dimensão $n \times n$ e seja $J_{(n)}$ uma matriz de constantes iguais a 1 e dimensão $n \times n$. Podemos escrever a matriz de covariância do modelo hierárquico na forma:

$$V_2 = \begin{bmatrix} \sigma_{\mu 0}^2 J_{(3)} + \sigma_{e 0}^2 I_{(3)} & 0 \\ 0 & \sigma_{\mu 0}^2 J_{(2)} + \sigma_{e 0}^2 I_{(2)} \end{bmatrix}$$

Uma das possibilidades de estimação dos parâmetros do modelo hierárquico é através do método de máxima verossimilhança com algoritmo de mínimos quadrados generalizados iterativo (GOLDSTEIN, 2011).

Utilizando mínimos quadrados generalizados e notação bloco-diagonal, os parâmetros dos efeitos fixos são dados por: $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$, onde X contém as variáveis explicativas e uma coluna igual a 1 e Y é a variável dependente.

O procedimento de estimação iterativo requer o fornecimento de valores iniciais plausíveis para os efeitos fixos, obtidos geralmente através do ajuste de um modelo OLS.

4.3.3 Verificação Empírica dos Efeitos da Comunalidade

Nesta subseção apresentamos uma releitura do modelo de canibalização e comunalidade de Kim e Chhajedb (2000) utilizando o modelo hierárquico (Subseção 4.3.2) e de preços hedônicos (Subseção 4.3.1). Nosso objetivo é integrar essas abordagens em uma equação econométrica única que permita capturar os efeitos da comunalidade nos preços.

Sejam um monopolista e dois segmentos de mercado denominados: “alto padrão” e “baixo padrão”. Vamos assumir que o segmento de alto padrão é representado pelos consumidores de carros zero quilômetro, enquanto o segmento de baixo padrão é representado pelos consumidores de carros usados. Temos os seguintes parâmetros do modelo teórico:

- (i) Valoração da qualidade dos consumidores de alto padrão: v_h .

- (ii) Valoração da qualidade dos consumidores de baixo padrão: v_l .
- (iii) Utilidades dos consumidores de alto padrão: $v_h q$.
- (iv) Utilidades dos consumidores de baixo padrão: $v_l q$.
- (v) Tamanho do mercado de alto padrão: n_h .
- (vi) Tamanho do mercado de baixo padrão: n_l .
- (vii) O custo de fornecimento de uma unidade: $c(q) = c q^2$.
- (viii) Design modular: q_m .
- (ix) Fator desconto: $0 \leq \beta_d < 1$.
- (x) Fator prêmio: $0 \leq \beta_p < 1$.

O parâmetro β_d diminui a utilidade dos consumidores de alto padrão devido à similaridade com o produto de baixo padrão. O parâmetro β_p aumenta a utilidade dos consumidores de baixo padrão devido à similaridade com o produto de alto padrão. Para evitar a migração entre produtos o monopolista escolhe:

$$P_l = v_l (q_l + \beta_d q_m) = v_l q_l + v_l \beta_d q_m$$

$$P_h = v_h q_h - q_l (v_h - v_l) - q_m (v_h \beta_d + \beta_p [v_h - v_l])$$

Vamos nos concentrar na equação do mercado de carros usados: $P_l = v_l(q_l + \beta_d q_m) = v_l q_l + v_l \beta_d q_m$. Como as características observadas do carro (potência do motor, quilometragem, número de válvulas, idade entre outras) são os atributos de qualidade do veículo, podemos representar a qualidade geral do carro usado q_l pela função: $q_l = g(\text{potência}_l, \text{km}_l, \text{válvulas}_l, \dots)$, e obter os parâmetros do modelo de comunalidade estimando um modelo de preços hedônicos com o formato: $\ln(p_{il}) = \alpha_0 + \alpha_1 q_{il} + \alpha_2 q_{im} + \xi_{il}$.

O modelo hedônico nos fornece então:

$$\begin{aligned}
 v_l &= \alpha_1 \\
 \beta_d &= \alpha_2 \\
 v_h &= \lambda_1 \\
 \lambda_3 &= (v_h \beta_d + \beta_p [v_h - v_l]) \\
 v_h &= \lambda_1 \\
 \beta_p &= \frac{\lambda_3 - v_h \beta_d}{[v_h - v_l]}
 \end{aligned}$$

Assumindo por conveniência que $g(\cdot)$ é uma função linear das características de qualidade observáveis do carro, temos: $q_{il} = \phi_{0l} + \phi_{1l} \text{potencia} + \phi_{2l} \text{km} + \phi_{3l} \text{valvulas} + \phi_{4l} \text{idade} + \dots$

Incluindo no modelo hedônico e usando a mesma padronização de [Kim e Chhajedb \(2000\)](#): $v_l = 1 = \alpha_1$, temos a equação final:

$$\begin{aligned}
 \ln(p_{il}) &= \alpha_0 + \alpha_1 [\phi_{0l} + \phi_{1l} \text{potencia} + \phi_{2l} \text{km} + \phi_{3l} \text{valvulas} + \phi_{4l} \text{idade} + \dots] \\
 &\quad + \alpha_2 q_{im} + \xi_{il} \\
 &= \alpha_0 + \phi_{0l} + \phi_{1l} \text{potencia} + \phi_{2l} \text{km} + \phi_{3l} \text{valvulas} + \phi_{4l} \text{idade} + \dots + \alpha_2 q_{im} + \xi_{il}
 \end{aligned}$$

A literatura aponta que a presença de comunalidade é mais forte para veículos da mesma marca ([VERBOVEN, 2002](#)). Visando capturar os efeitos de comunalidade sobre os preços, vamos substituir a variável q_m pela variável categórica de marca dos veículos segundo uma estrutura hierárquica. A variável marca será definida como efeito aleatório no modelo hierárquico, possibilitando interpretar os parâmetros de variabilidade no modelo como sendo o efeito da comunalidade. A Tabela [4.2](#) contém as especificações econométricas resultantes, as quais serão nomeadas neste trabalho de “modelo hierárquico hedônico”.

4.4 Dados

A base de dados utilizada para estimação do modelo hierárquico hedônico possui 131.650 mil anúncios de venda de 675 modelos de veículos usados e 82 marcas, coletados nos meses de junho e agosto de 2015. Maiores detalhes sobre a montagem da base de dados são apresentados na Seção 2.4.

Foram excluídas aproximadamente 30 mil observações relativas aos veículos zero quilômetro. Adicionalmente foram excluídos os veículos com mais de 25 anos de uso e as observações com valores extremos ou inconsistentes no preço e na quilometragem. A base final resultante passou a ter 88.810 observações, correspondendo a 508 modelos e 38 marcas.

Segundo a Tabela 4.1, quilometragem, tempo de anúncio, idade e tempo de mercado da montadora são negativamente correlacionados com os preços dos carros. A quilometragem possui correlação de 30,5% com os preços, enquanto a idade dos carros possui correlação de 46,6% e o tempo de anúncio possui correlação de -2,5%. A potência do motor (63,3%), a quantidade de válvulas (31,9%) e o número de portas do carro (8,1%) possuem correlação positiva com os preços. O tempo de mercado da marca ($wTime$) possui correlação negativa com os preços dos carros. As principais correlações com o preço envolvem quilometragem, idade, potência e número de válvulas.

TABELA 4.1 – Matriz de Correlação de Pearson

Coef.	ln(Preço)	ln(Km)	Anúncio	Idade	Potência	Válvulas	Portas	$wTime$
ln(Preço)	1,000	-0,305	-0,025	-0,466	0,633	0,319	0,081	-0,158
ln(Km)	-0,305	1,000	0,028	0,453	0,019	-0,054	0,027	0,008
Anúncio	-0,025	0,028	1,000	0,058	-0,005	-0,025	0,022	0,017
Idade	-0,466	0,453	0,058	1,000	0,098	-0,077	-0,146	0,060
Potência	0,633	0,019	-0,005	0,098	1,000	0,212	0,017	-0,067
Válvulas	0,319	-0,054	-0,025	-0,077	0,212	1,000	0,056	-0,113
Portas	0,081	0,027	-0,022	-0,146	0,017	0,056	1,000	0,079
$wTime$	-0,158	0,008	0,017	0,060	-0,067	-0,113	0,079	1,000

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

4.5 Resultados dos Modelos

Veículos da mesma marca possuem maiores níveis de comunalidade, pois os fabricantes agrupam os produtos em famílias que contenham maior grau de comunalidade interna (JANS *et al.*, 2008), e comercializam os veículos oferecendo um modelo básico e incrementos ao modelo básico (VERBOVEN, 1999). Para capturar os efeitos de comunalidade sobre os preços, vamos definir a marca dos veículos como hierarquia do modelo de preços hedônicos (Subseção 4.3.3). Os efeitos aleatórios da marca são interpretados como sendo os efeitos da comunalidade.

A especificação da forma funcional de modelos de preços hedônicos pode estar sujeita às escolhas metodológicas dos pesquisadores (REIS; SILVA, 2006). Entretanto, a literatura sugere incluir um conjunto amplo de variáveis explicativas (ARGUEA; HSIAO, 1993).

Reis e Silva (2006) utilizam como variáveis explicativas: marca, modelo, país de origem da montadora, existência de ar condicionado, número de portas, capacidade de mala, tipo de pintura, *air bag*, assento do motorista com ajuste lombar, banco do condutor com regulação elétrica, bancos desportivos, estofamento de couro, rodas de liga leve, tipo de combustível, vidros elétricos, sistema de áudio e equipamento de rádio, volante com ajuste de altura, sistema de direção assistida, computador de bordo, potência do motor, torque do motor, tipo de transmissão (automático/manual), faróis de neblina, velocidade máxima atingida, tempo de aceleração de 0 à 100 quilômetro por hora, relação entre peso e potência, consumo de combustível urbano e na estrada e autonomia de combustível do veículo.

No mercado de carros usados, Prieto *et al.* (2015) utiliza quilometragem, potência do motor, tipo de motor (diesel ou gasolina), segmento de mercado à qual pertence o veículo, cor do carro, pintura metálica e presença dos itens opcionais de luxo e segurança: Freio ABS, *Cruise Control*³, Computador de Bordo e Ar Condicionado. A inclusão do tipo de combustível como variável explicativa é justificada também pelo aumento das preocupações ambientais dos consumidores (PARRY *et al.*, 2007) e surgimento de regulações

³*Cruise Control*: Sistema de programação de velocidade.

ambientais determinando a cobrança de taxas por emissão de poluentes (KAGAWAA *et al.*, 2013; FENG *et al.*, 2013).

A fidelização dos consumidores por algumas marcas, devido à associação com qualidade (BRONNENBERG *et al.*, 2012), levanta a possibilidade das marcas estarem capturando também os efeitos de fidelização (SULLIVAN, 1998). A fidelidade dos consumidores por marcas não invalida sua utilização para captura do efeito de comunalidade. Primeiramente, os parâmetros estimados para a potência e quilometragem dos carros usados sugerem que a fidelidade por marca não representa efeitos tão importantes no mercado de carros usados, dados outros atributos dos veículos. Mesmo se houver tal fidelidade no mercado secundário, pode-se argumentar que os consumidores estariam associando toda a marca com qualidade, e não somente alguns modelos isolados.

A Tabela 4.2 apresenta as especificações dos modelos de preços hedônicos utilizados para verificação dos impactos da comunalidade sobre os preços.

TABELA 4.2 – Estrutura dos Modelos Hierárquicos Hedônicos - MHH

Modelo	Nível	Especificação dos Níveis Hierárquicos
MHH0	2: Anúncio	$\ln(p_{ik}) = \beta_{0k} + e_{ik}$
	1: Marca	$\beta_{0k} = \gamma_{00} + \mu_{0k}$
MHH1	2: Anúncio	$\ln(p_{ik}) = \beta_{0k} + \beta_{1k} potencia_{ik} + \beta_{2k} \ln(km_{ik}) + e_{ik}$
	1: Marca	$\beta_{0k} = \gamma_{00} + \mu_{0k}$
		$\beta_{1k} = \gamma_{10}, \dots, \beta_{2k} = \gamma_{20}$
MHH2	2: Anúncio	$\ln(p_{ik}) = \beta_{0k} + \beta_{1k} potencia_{ik} + \beta_{2k} \ln(km_{ik}) + e_{ik}$
	1: Marca	$\beta_{0k} = \gamma_{00} + \gamma_{01} wTime_k + \gamma_{02} Europa_k + \mu_{0k}$
		$\beta_{1k} = \gamma_{10}, \dots, \beta_{2k} = \gamma_{20}$
MHH3	2: Anúncio	$\ln(p_{ik}) = \beta_{0k} + \beta_{1k} potencia_{ik} + \beta_{2k} \ln(km_{ik}) + e_{ik}$
	1: Marca	$\beta_{0k} = \gamma_{00} + \gamma_{01} wTime_k + \gamma_{02} wEuropa_k + \mu_{0k}$
		$\beta_{1k} = \gamma_{10} + \mu_{1k}, \dots, \beta_{2k} = \gamma_{20} + \mu_{2k}$
MHH4	2: Anúncio	$\ln(p_{ik}) = \beta_{0k} + \beta_{1k} potencia_{ik} + \sum_{p=2}^5 \beta_{ip} X_{pik} + e_{ik}$
	1: Marca	$\beta_{0k} = \gamma_{00} + \gamma_{01} Europa_k + \mu_{0k}$
		$\beta_{1k} = \gamma_{10} + \mu_{1k}, \dots, \beta_{2k} = \gamma_{20}$
MHH5	2: Anúncio	$\ln(p_{ik}) = \beta_{0k} + \beta_{1k} potencia_{ik} + \sum_{p=2}^P \beta_{ip} X_{pik} + e_{ik}$
	1: Marca	$\beta_{0k} = \gamma_{00} + \gamma_{01} Europa_k + \mu_{0k}$
		$\beta_{1k} = \gamma_{10} + \mu_{1k}$

O modelo *MHH0* não contém nenhuma variável explicativa além do intercepto e do efeito aleatório da marca. O modelo *MHH0* tem como objetivo calcular o coeficiente de correlação intraclasse (ICC) e verificar qual o percentual de variabilidade é devido à estrutura de hierárquica envolvendo marcas (nível 1) e veículos (nível 2). O coeficiente de correlação intraclasse (ICC) indica a intensidade da correlação entre os preços de veículos de uma mesma marca, sendo interpretado como o efeito da comunalidade. Valores muito pequenos indicam que não existe hierarquia nos dados, podendo-se ajustar um modelo OLS. Valores altos indicam que as observações são correlacionadas dentro de cada marca e o ajuste de um modelo OLS não é adequado.

Em *MHH1*, incluímos como efeitos fixos dois importantes determinantes dos preços: potência e quilometragem. Os efeitos fixos possuem a mesma interpretação dos parâmetros de um modelo OLS. Em *MHH1* um valor significativo para o coeficiente de correlação intraclasse indica que os efeitos da comunalidade sobrevivem mesmo após a inclusão da potência e quilometragem dos carros.

Em *MHH2*, a estrutura hierárquica fica mais complexa ao incluir-se como variáveis explicativas do efeito aleatório tempo de mercado (*wTime*) e origem europeia da marca (Europa=0 ou 1). Estas variáveis encontram-se no intercepto aleatório β_{0k} .

Em *MHH3*, incluímos como variável explicativa da marca (efeito aleatório) a potência do veículo e a quilometragem. O parâmetro μ_{1k} diz que o efeito da potência sobre os preços não é fixo, mas possui variabilidade devido às marcas. Interpreta-se μ_{1k} como sendo o efeito da potência sobre as marcas, mesmo contabilizando as percepções dos consumidores a respeito da comunalidade. O parâmetro μ_{2k} diz que o efeito da quilometragem sobre os preços possui variabilidade devido a marca. Como a potência é um atributo de fábrica, observamos que seus impactos sobre os efeitos aleatórios foram mais significativos que os impactos da quilometragem (Tabela 4.3).

Em *MHH4*, excluimos a quilometragem das variáveis explicativas dos efeitos aleatórios, passando a ser um efeito fixo. Além da quilometragem ($X_{p=2,ik}$), são utilizados como efeitos fixos: tempo de anúncio na internet ($X_{p=3,ik}$), idade ($X_{p=4,ik}$) e válvulas ($X_{p=5,ik}$).

Em *MHH4*, vemos que a correlação intraclasse se mantém forte mesmo após a inclusão de outros determinantes dos preços (Tabela 4.3).

Em *MHH5* temos a mesma especificação hierárquica de *MHH2*, mas utilizando a potência como efeito aleatório e incluindo como efeitos fixos os itens de luxo e segurança⁴. O modelo *MHH5* permitirá interpretar os efeitos fixos como em um modelo de preços hedônicos, incorporando, ao mesmo tempo, a estrutura de hierarquia dos dados. Por este motivo, adotamos a nomenclatura “modelo hierárquico hedônico”.

Abaixo encontra-se a representação, em equação única, dos modelos hierárquicos hedônicos da Tabela 4.2.

$$\text{MHH0: } \ln(p_{ik}) = \gamma_{00} + \mu_{0j} + e_{ik}$$

$$\text{MHH1: } \ln(p_{ik}) = \gamma_{00} + \gamma_{10} \textit{potencia}_{ik} + \gamma_{20} \ln(km_{ik}) + \mu_{0j} + e_{ik}$$

$$\text{MHH2: } \ln(p_{ik}) = \gamma_{00} + \gamma_{01} \textit{wTime}_k + \gamma_{02} \textit{Europa}_k + \gamma_{10} \textit{potencia}_{ik} + \gamma_{20} \ln(km_{ik}) + \mu_{0j} + e_{ik}$$

$$\text{MHH3: } \ln(p_{ik}) = \gamma_{00} + \gamma_{01} \textit{wTime}_k + \gamma_{02} \textit{Europa}_k + \gamma_{10} \textit{potencia}_{ik} + \mu_{1j} \textit{potencia}_{ik} + \gamma_{20} \ln(km_{ik}) + \mu_{2j} \ln(km_{ik}) + \mu_{0j} + e_{ik}$$

$$\text{MHH4: } \ln(p_{ik}) = \gamma_{00} + \gamma_{01} \textit{Europa}_k + \gamma_{10} \textit{potencia}_{ik} + \mu_{1j} \textit{potencia}_{ik} + \gamma_{20} \ln(km_{ik}) + \gamma_{30} \textit{anuncio}_{ij} + \gamma_{40} \textit{idade}_{ij} + \gamma_{40} \textit{valv}_{ij} + \mu_{0k} + e_{ik}$$

$$\text{MHH5: } \ln(p_{ik}) = \gamma_{00} + \gamma_{01} \textit{wEuropa}_k + \gamma_{10} \textit{potencia}_{ik} + \mu_{1j} \textit{potencia}_{ik} + \left[\sum_{p=2}^P \beta_{ip} X_{ik} \right] + \mu_{0k} + e_{ik}$$

O valor *ICC* do modelo *MHH0* aponta que 20,9% da variabilidade nos preços devem-se às diferenças nas marcas dos veículos (Tabela 4.3), sendo este um valor relevante (BELL *et al.*, 2013). O valor do *ICC* indica que os consumidores percebem a comunalidade em

⁴Os itens de luxo e segurança são: idade, tempo de anúncio, categoria de design, tipo de vendedor, válvulas, combustível, freios *ABS*, *Air Bag*, alarme, ar condicionado, ar quente, regulagem de bancos, controle de velocidade, bancos de couro, desembaçador traseiro, direção hidráulica, sensor de estacionamento, som com CD ou MP3, computador de bordo, rodas de liga leve, regulagem de retrovisor, tração nas quatro rodas, trava elétrica, vidro elétrico, regulagem de volante, número de portas. A descrição completa das variáveis encontra-se na Tabela A.2.

veículos de uma mesma marca.

Em *MHH1* e *MHH2*, a inclusão de efeitos fixos parece diluir os efeitos da comunalidade. Em *MHH1*, o coeficiente de correlação intraclasse é de 11,4%, enquanto em *MHH2*, o coeficiente de correlação intraclasse é de 8,4%. Entretanto em *MHH2*, 44,1% da variabilidade devem-se à origem europeia das marcas. Assim, a percepção de comunalidade dos consumidores parece estar associada à origem das marcas, mas ainda impacta nos preços. O tempo de mercado não apresentou significância estatística, sendo excluída dos próximos modelos.

Em *MHH3*, o impacto da quilometragem não possui variação devida à marca. Assim, a quilometragem foi excluída da lista de efeitos aleatórios, sendo mantida como efeito fixo em *MHH4* e *MHH5*.

Em *MHH3*, um aumento de uma unidade na potência está associada a uma elevação de 0,373% nos preços, sendo este impacto variável entre as marcas (Tabela 4.3). A inclusão do tempo de mercado da marca como efeito aleatório não apresentou significância estatística.

Os modelos *MHH4* e *MHH5* apresentaram os menores valores para os critérios de informação de Akaike (AIC) e Bayesiano (BIC), sendo estes os modelos indicados para análise final.

Em *MHH4*, um aumento de uma unidade na potência dos motores está associado a um aumento de 0,41% nos preços, e este impacto não é fixo, mas possui variabilidade entre as marcas. Desta forma, podemos falar que o efeito da potência sobre os preços depende da percepção de comunalidade dos consumidores.

Em *MHH4*, os efeitos aleatórios são potência e a origem europeia da montadora. O parâmetro aleatório para o tempo de mercado (*wTime*) não apresentou significância estatística. O parâmetro de origem europeia é estatisticamente significante, sendo responsável por 51,1% da variabilidade de preços. Os consumidores provavelmente percebem intensidades de comunalidade diferentes, dependendo da origem da montadora. Um aumento de 1% na quilometragem dos carros possui um impacto sobre os preços de -0,035% (Tabela 4.3). O tempo de anúncio também possui impacto positivo sobre os preços, indicando que

TABELA 4.3 – Estimativas dos Modelos Hedônicos Hierárquicos com Dois Níveis

Explicativas	MHH0	MHH1	MHH2	MHH3	MHH4	MHH5
Intercepto	11,0 213,5***	11,7 382,7***	11,6 140,9***	11,6 241,2***	10,4 905,3***	10,46 927,4***
Potência	-	0,41 174,3***	0,41 176,6***	0,373 23,3***	0,409 28,3***	0,311 26,8***
ln(Km)	-	-0,16 129,8***	-0,16 -131,9***	-0,157 -134,3***	-0,035 -38,0***	-0,037 -41,6***
Tempo de Anúncio	-	-	-	-	0,0001 6,7***	0,0001 8,1***
Idade	-	-	-	-	-0,084 -278,4***	-0,078 -251,6***
Válvulas	-	-	-	-	0,012 65,8***	0,008 43,1***
Europa	-	-	0,310 3,6***	0,526 5,3 ^{ns}	0,775 9,4 ^{ns}	0,656 9,7***
Tempo de Mercado (<i>wTime</i>)	-	-	0,633 1,3 ^{ns}	0,583 0,8 ^{ns}	-	-
ICC(Marca)	20,9%	11,4%	8,4%	10,0%	-	-
ICC(Potência)	-	-	-	2,2%	3,5%	2,9%
ICC(Europa)	-	-	44,1%	28,0%	51,1%	44,4%
AIC	140.385	95.966	94.939	87.695	23.473	9.965
BIC	140.388	95.969	94.944	87.70 0	23.484	9.970

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “*ns*” Não possui significância estatística.

a demora em vender um veículo está associada a preços mais altos.

Em *MHH5*, a variabilidade explicada pela inclusão da potência foi de 44,4%. As estimativas dos itens de luxo e segurança do modelo *MHH5* serão apresentadas na Tabela C.1 do Apêndice C.1. As variáveis indicadoras são interpretadas utilizando a transformação: $[\exp \beta - 1] \times 100\%$.

Os preços dos carros são valorizados pela presença de *air bag* (1,3%), ar condicionado (3,4%), computador de bordo (2,5%), sensor de estacionamento (4,7%), controle de velocidade (8,2%), bancos de couro (8,2%), direção hidráulica (0,8%), freios ABS (4,3%), rodas de liga leve (7,4%), regulagem retrovisor (3,2%), tração nas quatro rodas (5,1%) e vidro elétrico (0,1%) (Tabela C.1 - Apêndice B.1). O sistema de regulagem de banco (-1,8%),

trava elétrica (-4,1%) e desembaçador traseiro (-4,7%) possuem elasticidades estimadas negativas e estatisticamente significantes, provavelmente devido à ampla difusão desses itens em todos os tipos de carros.

No Apêndice C.1, a Tabela C.2 apresenta os parâmetros estimados para diferentes especificações do modelo hierárquico hedônico, contendo três níveis: veículo (nível 3), modelo (nível 2) e marca (nível 1). Quilometragem, potência e idade dos carros apresentaram significância estatística diante de diferentes especificações do modelo hierárquico hedônico, e os valores do coeficiente de correlação intraclasse são acentuados. Estes resultados mostram que a estrutura hierárquica captura os efeitos da comunalidade e não os efeitos de fidelidade por marca.

4.6 Conclusões

Neste capítulo contribuímos para a literatura avaliando os efeitos da comunalidade no mercado de carros usados. Utilizando intuição obtida em Kim e Chhajed (2000), estimamos um modelo de preços hedônicos estruturado segundo um modelo hierárquico, tendo como objetivo capturar os efeitos da comunalidade através da variável marca. Os parâmetros do modelo hierárquico hedônico foram estimados utilizando uma base de dados de anúncios na internet, a qual foi construída através de técnicas de *web scraping*.

A metodologia empregada permitiu constatar a existência de efeitos significantes da comunalidade sobre os preços dos veículos usados. Os efeitos da comunalidade são consistentes diante de diferentes especificações do modelo hierárquico hedônico e preservam-se mesmo após a inclusão de importantes determinantes dos preços de veículos usados, tais como quilometragem, idade e potência.

Potência, quilometragem e idade dos carros apresentaram significância estatística diante de diferentes especificações do modelo hierárquico hedônico, indicando enfraquecimento do efeito de fidelidade por marca no mercado secundário (Tabela 4.3). A intensidade dos parâmetros estimados para potência, quilometragem e idade assegura que o

efeito aleatório de marca captura os efeitos da comunalidade entre os carros usados.

Os resultados indicam que a percepção dos consumidores a respeito da comunalidade, utilizada pelas montadoras de veículos, estende-se para o mercado secundário, afetando a escolha de modelo e marca de veículo.

5 Modelagem da Depreciação de Veículos

5.1 Introdução

A depreciação de preços dos veículos usados é importante na decisão entre alocação de veículos para venda e arrendamento (ANDRIKOPOULOS; MARKELLOS, 2015). Quando um veículo é devolvido ao final de um contrato de *leasing* a montadora precisa revendê-lo no mercado secundário (HUANG *et al.*, 2001). Portanto, um decaimento muito radical nos preços dos veículos usados torna menos atrativa a estratégia de *leasing* e mais atrativa a venda de veículos.

O entendimento do processo de depreciação dos veículos também é importante para a decisão racional dos consumidores. A compra de um automóvel zero representa a alocação de parte significativa da renda do domicílio (ABRAHAM *et al.*, 2015), e os consumidores costumam entregar o carro antigo como parte do pagamento por um carro novo (BUSSE; SILVA-RISSO, 2010). Consumidores racionais deveriam considerar a expectativa de depreciação futura dos veículos nas negociações de compra e venda de veículos.

O entendimento do processo de depreciação dos carros também é importante no desenho de políticas de descarte de veículos velhos como sucata, por exemplo, no estabelecimento de uma idade mínima de elegibilidade. Estas políticas têm sido adotadas em alguns países devido a motivações ambientais relacionadas com a emissão de poluentes de combustíveis fósseis dos carros antigos (FENG *et al.*, 2013).

Para modelagem do processo de depreciação dos carros, foram estimados os chamados modelos de crescimento exponencial (GOSHU; KOYA, 2013; ANNADURAI *et al.*, 2000; GRIMM, 2009; FEKEDULEGN *et al.*, 1999), cujos parâmetros estimados permitem calcular o ponto de inflexão da curva de depreciação para várias categorias de veículos (Perua, Minivan, Sedã, Utilitário, Picape, Hatchback, Cupê e Conversível). O ponto de inflexão é interpretado como sendo um momento de transição, quando o veículo termina seu ciclo de depreciação mais intensa, e o processo de decaimento nos preços passa a ser mais lento. O ponto de inflexão é interpretado como a idade do veículo onde a decisão de compra do carro usado seria benéfica para um comprador consciente. Os consumidores de carros novos poderiam avaliar os pontos de inflexão para veículos de diferentes categorias, decidindo pela compra em função do horizonte de tempo em que pretendem revender o veículo novo. Os parâmetros dos modelos também permitem estabelecer a depreciação máxima pelo veículo, podendo embasar políticas de descarte de veículos antigos, fornecendo um parâmetro para o prêmio pago pelo veículo antigo.

Os resultados encontrados mostram que os modelos de crescimento exponencial revelam interessantes padrões na dinâmica de depreciação dos carros. Os pontos de inflexão e a depreciação associada a este ponto apresentam consideráveis variações dependendo da categoria de veículo. Segundo o modelo Richards, um veículo Toyota-Hilux (Picape) possui ponto de inflexão em 5 anos de uso, momento em que atinge também a desvalorização máxima de 78% do valor inicial. Já um Fiat-Palio (Hatchback) possui ponto de inflexão em 3 anos de idade, entretanto, não experimenta sua desvalorização máxima no ponto de inflexão. Um veículo Audi-A6 possui ponto de inflexão no primeiro ano de vida útil, quando experimenta a desvalorização de 49%, e possui desvalorização máxima de 91%. Os resultados sugerem que a decisão dos consumidores entre adquirir um veículo novo ou usado poderia considerar o horizonte de tempo em que pretendem utilizar seus veículos, e ainda qual a perda de valor que estão dispostos a absorver no momento da revenda.

Os valores dos coeficientes de explicação mostram que os modelos de crescimento exponencial são adequados para explicar o processo de depreciação dos veículos, considerando, principalmente que os modelos possuem uma única variável explicativa. Os modelos de

crescimento exponencial podem ser utilizados para previsão do valor residual dos veículos, embasando as decisões de negócios dos agentes atuantes neste mercado, tais como seguradoras, locadoras, revendedoras e consumidores.

A evolução dos preços de carros usados foram investigados anteriormente do ponto de vista da assimetria de informação e seleção adversa (AKERLOF, 1970; ENGERS *et al.*, 2009; MURRAY; SARANTIS, 1999), do ponto de vista dos custos de transação, como impedimento para ajuste instantâneo de utilidade (FENG *et al.*, 2013), e do ponto de vista da heterogeneidade dos consumidores (BUSSE *et al.*, 2012; GAVAZZA *et al.*, 2014; ANDRIKOPOULOS; MARKELLOS, 2015; ABRAHAM *et al.*, 2015), onde consumidores pouco tolerantes à perda de desempenho dos carros (ANDRIKOPOULOS; MARKELLOS, 2015) recorrem mais frequentemente ao mercado de carros novos (ABRAHAM *et al.*, 2015), gerando os lucros no mercado secundário (GAVAZZA *et al.*, 2014) e forçando uma curva de depreciação mais acentuada nos anos iniciais de vida dos veículos (BUSSE *et al.*, 2012).

Na Seção 5.2 apresentamos uma breve revisão da literatura, abordando a depreciação de preços de veículos do ponto de vista das políticas de *scraping* (Seção 5.2.1), passando pela literatura de comportamento míope dos consumidores (Seção 5.2.2), e modelos contábeis de depreciação de bens duráveis (Subseção 5.2.3). Na Subseção 5.3.1 apresentamos algumas características dos modelos não-lineares, incluindo a expressão final para os pontos de inflexão. Demais detalhes metodológicos sobre os pontos de inflexão e a estimação dos modelos não-lineares encontram-se no Apêndice D.3. Na Seção 5.5 são apresentados os resultados estimados para os modelos não-lineares: Richards e *Logistic*, enquanto a Seção 5.6 apresenta as conclusões finais do capítulo 5. O Apêndice D.3 contém as rotinas de programação do *proc model-SAS* utilizadas para estimação dos modelos não-lineares de crescimento. Demais tabelas e gráficos contendo os resultados não reportados para os modelos: Negative Exponential, Monomolecular, Mitcherlich, Gompertz, Von Bertalanffy e Weibull, encontram-se no Apêndice do Capítulo 5.

5.2 Revisão da Literatura

Nesta seção exploramos mais detalhadamente a literatura sobre depreciação de preços de veículos já comentada na seção introdutória. Os leitores que desejarem, podem seguir adiante, retomando a leitura a partir da descrição dos dados (Seção 5.4), análise dos resultados (Seção 5.5) e conclusões finais (Seção 5.6).

5.2.1 Políticas de Descarte de Veículos Usados como Sucata

As emissões de poluentes dos veículos constituem a principal fontes de gases-estufa nos centros urbanos (KAGAWAA *et al.*, 2013). Visando desestimular comportamentos poluentes no trânsito, vários países desenvolveram regulações e criaram políticas ambientais, envolvendo vistorias programadas para emissão de poluentes, cobrança de taxas devido à emissão de CO₂, cobrança de taxas devido ao consumo extra de gasolina, e por último, implantação de programas de incentivo à troca de carros antigos por carros novos (FENG *et al.*, 2013).

Políticas de remoção de carros velhos¹ foram implantadas em diversos países. Na França, na década de 90, inicialmente era oferecido o valor de \$2.000 francos pela troca de um carro usado por um novo, sem estabelecer idade mínima de elegibilidade. Nos anos seguintes a política foi retomada, modificando a idade de elegibilidade mínima e o montante de prêmio pago para os proprietários de carros velhos. Em Adda e Cooper (2000), a mudança da idade ocorreu em função dos efeitos das políticas anteriores, na formação de estoque mais jovem de veículos. Entretanto, a literatura não debate quais os critérios quantitativos para estabelecimento de uma idade mínima de elegibilidade e do valor pago como prêmio de *scraping*. Nos EUA e França, após 10 anos de uso, os proprietários começam a considerar o descarte dos carros usados como sucata (ENGERS *et al.*, 2009).

O estoque de carros mantido pelos consumidores é persistente ao longo do tempo, ou

¹Do inglês, *scraping policy* é a retirada de circulação de um veículo usado como sucata, geralmente sendo aproveitado o material para reciclagem.

seja, se um consumidor possui um carro em um ano, então provavelmente irá possuir o mesmo carro no ano seguinte (SCHIRALDI, 2011). Esta persistência define uma composição etária e de valor do estoque de veículos usados. Devido aos custos de transação na troca de veículos, os consumidores absorvem anualmente parte da depreciação de preços até certo limiar de perda de utilidade (ALBERINI *et al.*, 1998). Após este limiar de perda de utilidade os consumidores se tornam propensos a trocar os veículos por modelos mais novos, podendo engajar em uma política de descarte de veículos antigos como sucata, caracterizando um processo de atualização da composição etária da frota de veículos.

Quando as políticas de *scraping* visam diminuir a emissão de poluentes de combustíveis fósseis (KAGAWAA *et al.*, 2013), tais políticas deveriam focar em consumidores que possuem maior tolerância à perda de utilidade do veículo devido à idade, e em proprietários de veículos próximos do limiar de depreciação máxima. No último caso, os modelos não-lineares de crescimento exponencial permitem estabelecer a idade onde o veículo atinge sua depreciação máxima, ou seja, a idade limiar onde o veículo continuará produzindo poluição ambiental sem experimentar queda estatisticamente de preço. Utilizando a idade onde ocorre a depreciação máxima do carro, é possível ainda estimar o valor residual do carro, o qual que deverá ser oferecido como prêmio de *scraping*, permitindo também avaliar a viabilidade econômica da política. Conforme representação no Gráfico 5.1, os proprietários estariam dispostos a aceitar um prêmio de *scraping* de V por veículos com idade A , concordando em retirar prematuramente tais carros de circulação e evitando a emissão de poluição associada a tais veículos nos próximos anos.

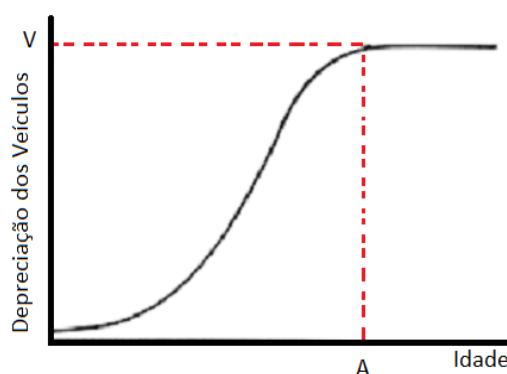


FIGURA 5.1 – Evolução Temporal da Depreciação de Valor

Utilizando a estimativa de depreciação máxima de um modelo não-linear de crescimento exponencial, substituir na variável dependente e invertendo a equação, obtemos a idade onde o veículo atinge a sua depreciação máxima. Desta forma, teríamos uma idade de elegibilidade baseada em critérios quantitativos. Como em algumas marcas que possuem maior reputação associada à resistência e durabilidade (DESAI, 2001), o ajuste dos modelos não lineares podem considerar as diferentes categorias de veículo e estabelecer a idade de elegibilidade diferentes para veículos de diferentes categorias.

Se os consumidores contabilizarem os custos de manutenção e reparos, um veículo em perfeita condições de funcionamento é descartado como sucata quando o seu valor de mercado p_i menos o valor prêmio de descarte sv_i for menor que os gastos com reparos $q r$ necessários para fazer funcionar um veículo inoperante, ou seja, o veículo será descartado como sucata se: $p_i - sv_i < q r$, onde q é a quantidade de reparos, e r é o preço dos reparos (GRUENSPECHT, 1982). Se os consumidores fazem escolhas racionais e todos os veículos possuem a mesma probabilidade de incorrer em reparos $f_i(q)$, então a taxa de *scraping* para cada tipo de veículo d_i será dada 1 menos a integral da distribuição probabilidade dos reparos sobre o intervalo onde a decisão pelos reparos é racional:

$$d_i = 1 - \int_0^{(p_i - sv_i)/r} f_i(q) dq \quad (5.2.1)$$

A Equação (5.2.1) mostra que o percentual de veículos descartados diminui com o aumento do valor de mercado dos carros usados: p_i , caracterizando uma associação inversa entre a taxa de *scraping* d_i e o valor de mercado dos carros usados p_i , mesmo diante de custos de transação e assimetria de informação. Um aumento no valor dos carros novos aumentaria a demanda por carros usados, levando à queda da taxa de *scraping* (GRUENSPECHT, 1982). Mantendo-se os preços dos carros novos constantes, as novas tecnologias incorporadas aos carros novos fazem a taxa de *scraping* aumentar, devido ao aumento no diferencial de utilidade entre veículos novos e antigos, o que também faz cair os preços dos carros usados.

Espera-se que os primeiros carros a serem descartados sejam aqueles em pior estado

de conservação (ALBERINI *et al.*, 1998), sendo estes os maiores responsáveis pela emissão de poluentes (WEE *et al.*, 2000).

Os consumidores podem fazer a decisão entre manter seu veículo ou descartá-los comparando o valor presente líquido (VPL) de um veículo usado π_t com o valor presente líquido da segunda melhor alternativa A_t . Esta segunda melhor alternativa pode ser um carro mais novo, não possuir nenhum veículo ou descartar o veículo antigo como sucata aceitando o prêmio de *scraping* (ALBERINI *et al.*, 1998). Seja $J_t = \max(A_t, \pi_t)$ e o valor presente líquido (VPL) da decisão ótima no ano t , para entender os efeitos de uma política de *scraping*, vamos dividir o valor presente líquido (VPL) de manter a propriedade do veículo antigo em dois componentes: (i) O valor presente líquido dos serviços do veículo durante o ano corrente, (ii) a taxa esperada em valores atuais da decisão ótima no ano seguinte $t + 1$:

$$\pi_t = V_t + \beta E(J_{t+1})$$

onde, β é o fator de desconto e E o valor esperado. O valor líquido dos serviços do veículo é dado pela diferença entre os benefícios e os custos advindos do veículo: $V_t = B(D_t, Q_t(Q_{t-1}, M_{t-1})) - C(D_t, EM_t)$, onde D_t é a quantidade de milhas dirigidas no ano t , Q_t é a medida de condição do veículo, M_{t-1} é a medida de manutenção do veículo no ano $t - 1$, e EM_t é o gasto esperado com manutenção no próximo ano. As condições de desempenho do veículo Q_t dependem das mesmas condições no ano anterior Q_{t-1} , bem como das condições de manutenção no ano prévio M_{t-1} . O valor presente líquido (VPL) de substituição do veículo por um novo é dado por:

$$A_t = \max_a p_t - (p_t^a + c) + V_t^a + \beta E(J_{t+1}^a)$$

onde, p_t é o preço atual do veículo e $p_t^a + c$ é o custo da segunda melhor alternativa. O custo da segunda melhor alternativa abrange também os custos de transação c . O valor dos serviços advindos do veículo a no ano corrente é dado por: V_t^a , e o valor presente líquido (VPL) de se possuir um veículo no ano $t + 1$ é dado por: $\beta E(J_{t+1}^a)$.

Como a decisão de descartar um veículo é afetada pela oferta de uma prêmio de *scraping*? Qual o impacto das expectativas futura de gastos com manutenção sobre a decisão de descartar um carro como sucata? Suponha que o consumidor decidiu manter o seu veículo: $\pi_t > A_t$, e seja S uma oferta de descarte de sucata. Considerando que o veículo antigo está sujeito as taxas extras de poluição e precisa recusar as ofertas de descarte de sucatas, então o premio de *scraping* será aceito se:

$$p_t < \pi_t - [\pi_t^a - (p_t^a + c)] \quad (5.2.2)$$

onde, $\pi_t^a = V_t^a + \beta E(J_t^a)$ é o benefício bruto da segunda melhor alternativa. A desigualdade (5.2.2) é derivada considerando que o proprietário manteria seu veículo antigo na ausência de uma política de descarte veículos antigos como sucata. A decisão de entregar o carro para descarte depende das características do veículo e também das características da segunda melhor alternativa disponível. Consumidores que já estivessem planejando trocar o veículo antigo aceitariam um prêmio de *scraping* cujo valor seja pelo menos igual ao valor de mercado do seu veículo antigo. Consumidores que não estivessem planejando trocar de carro teriam que pesar sua decisão comparando o prêmio de *scraping* com os valores esperados futuros com manutenção e reparos, taxas de poluição e depreciação dos usados nos anos seguintes.

Suponha a introdução de uma regulação ambiental prevendo um teste de emissão de poluentes, onde os proprietários devem incorrem em custos de manutenção e reparos de R em caso de poluição excedente do veículo. A expectativa destes custos diminui o valor presente líquido do veículo de p_t para $p_t - R$. Agora um potencial comprador do veículo antigo provavelmente irá verificar as condições de emissão de poluentes ambientais, e a expectativa dos custos de manutenção irá impactar negativamente o valor de mercado. O proprietário do veículo antigo desejará evitar tais custos de manutenção e reparação, o que pode ser feito aceitando um prêmio de *scraping*. O proprietário do veículo antigo decidirá pelo descarte se $\pi_t - R < S$, onde S é o prêmio de *scraping*, ou seja, a decisão pelo descarte depende da comparação do valor de mercado do veículo antigo em relação

ao valor de descarte de sucata, mais os custos de manutenção e reparos.

Os modelos não lineares de crescimento exponencial serão utilizados como critério quantitativo para estabelecimento de uma idade mínima de elegibilidade e um valor sugerido de prêmio de *scraping*. Em primeiro lugar, os modelos fornecem o valor predito dos veículos em qualquer idade, permitindo simular diferentes combinações entre idade e prêmio de *scraping*. Em segundo lugar, a estimativa de depreciação máxima, dada pelo parâmetro β_0 (Subseção 5.3.1), pode ser substituída na variável dependente, calculando-se a idade na qual o veículo atinge sua depreciação máxima: $\hat{\beta}_0 = w(d) \Rightarrow d^* = w(\hat{\beta}_0)^{-1}$.

5.2.2 Modelo de Consumidores Míopes

A literatura investiga em que medida os consumidores antevem adequadamente os custos operacionais na sua decisão de compra de bens duráveis (SAWHILL, 2010). Os consumidores são heterogêneos em relação à utilidade marginal do serviço prestado pelos carros (STOLYAROV, 2002), existindo alguns consumidores pouco tolerantes à perda de utilidade dos carros, os quais fazem escolhas que privilegiem gratificação instantânea (MEIER; SPRENGER, 2010). Parte da depreciação experimentada pelos veículos nos anos iniciais de vida seria explicada pela combinação entre consumidores heterogêneos e a falha em antever adequadamente a depreciação futura dos veículos (KAHN, 1986). Ao transacionarem os seus veículos usados como parte do pagamento de um veículo zero quilômetro, esses consumidores impulsivos estariam mais dispostos em perder preço no carro usado na negociação de compra de um carro novo, induzindo taxas de depreciação mais intensas no início do ciclo de vida de um veículo (BUSSE; SILVA-RISSO, 2010).

Entretanto, Verboven (2002) não encontrou evidências de miopia dos consumidores no mercado de carros, enquanto Sallee (2015) mostra que os consumidores antecipam adequadamente os gastos futuros com combustíveis gastos por milhas rodadas, e em Schiraldi (2011) os consumidores decidem entre veículos movidos a gasolina ou a diesel de forma compatível com a expectativa de gastos futuros com combustíveis. Também em Verboven (1999), os consumidores antecipam razoavelmente os gastos futuros com combustível na

decisão de escolha por um carro, não havendo evidências de miopia. Evidências de miopia dos consumidores também não são confirmadas por [Busse et al. \(2012\)](#), que compara as taxas de desconto implícitas na avaliação de veículos e a taxa de juros da economia.

5.2.3 Modelo Contábil de Depreciação de Bens Duráveis

Deterioração de bens duráveis pode ser definida como o declínio físico observado ao longo do tempo, ou seja, os efeitos de obsolescência e perda de desempenho dos bens usados em comparação aos modelos novos ([DESAI, 2001](#)). Por outro lado, a depreciação é a perda de valor econômico experimentada com o envelhecimento de um bem durável. A literatura geralmente adota variáveis *Proxy* para representar o desgaste de veículos. Em [Engers et al. \(2009\)](#), o desgaste do veículo é mensurado pela quilometragem anual, em [Feng et al. \(2013\)](#), o desgaste do carro é percentual de depreciação dos preços entre dois momentos do tempo, e [Barnes e Langworthy \(2003\)](#) utiliza o desgaste dos pneus como medida de desgaste.

A literatura distingue ainda os conceitos de depreciação e obsolescência, onde este último refere-se aos efeitos sobre os preços de uma “safra” de bens duráveis resultante do aparecimento de novas safras ([WYKOFF, 2004](#)). Se a quantidade de milhas dirigidas por ano pode ser utilizada como medida do fluxo de serviços advindos de um carro, ou seja, uma variável *proxy* para a qualidade do carro, então as milhas anuais dirigidas explicam grande parte das quedas nos preços dos carros usados ([ENGERS et al., 2009](#)). O efeito obsolescência relacionado às mudanças de design, incorporação de novos itens de segurança e conforto estão entre os fatores que explicam a queda nos preços dos veículos usados ao longo do tempo ([SALLEE, 2011](#)). Entretanto, o efeito de obsolescência de carros ainda é pequeno se comparado a outros bens duráveis que também experimentam mudanças radicais de design e tecnologia, ou seja, as melhorias de design e qualidade dos carros não explicam inteiramente o declínio nos preços dos carros usados ([WYKOFF, 2004](#)).

Os debates sobre os conceitos depreciação e deterioração remetem aos tempos da con-

trovêrsia das duas Cambridges² na década de 50 (WYKOFF, 2004). Como um tema secundário ao debate, procurava-se estabelecer se as medidas agregadas de capital deveriam distinguir entre obsolescência e deterioração diante das mudanças tecnológicas incorporadas às novas máquinas (COHEN; HARCOURT, 2003). O ponto de vista neoclássico não diferenciava entre obsolescência e deterioração nas medidas agregadas de capital, distinguindo tais conceitos somente em análises microeconômicas.

No presente trabalho apesar de reconhecermos a correlação entre obsolescência e depreciação (GORDON, 1990), será realizada a modelagem considerando o conceito de depreciação econômica. Embora fosse possível construir uma medida de deterioração utilizando a mesma abordagem de Engers *et al.* (2009), nosso interesse em depreciação econômica deve-se à sua manifestação como reflexo de outros efeitos além do decaimento físico, tais como, os efeitos de heterogeneidade dos consumidores na valoração da qualidade dos carros (ENGERS *et al.*, 2009), efeito de mudança de design de um modelo novo (SAWHILL, 2010), mudança de valoração de atributos pelos consumidores (OHTA; GRILICHES, 1986) e efeito de miopia dos consumidores (BUSSE *et al.*, 2012).

Ao abordar o processo de depreciação de bens duráveis do ponto de vista da deterioração física, Schiraldi (2011) considera um processo estocástico no tempo t , onde cada consumidor possui no máximo uma unidade do bem durável. Devido à degradação dos bens duráveis, cada consumidor desejará fazer a substituição por um modelo novo, recorrendo ao mercado de revenda. Os consumidores maximizam o valor esperado da utilidade intertemporal utilizando fator de desconto $\beta \in (0, 1)$.

Seja $\mathbf{J}t$ o conjunto de carros disponíveis no período t , o qual abrange também o mercado secundário: $J_t = j : j \in [\bigcup_{\tau=1}^t \mathbf{J}_\tau]$. Considerando o mercado de automóveis, em cada período existe a possibilidade do consumidor optar pelo carro *outside*, ou seja, não possuir nenhum carro: $j = 0$. Se o consumidor não adquiriu veículo no período

²Nos debates, os principais protagonistas eram Joan Robinson (Cambridge, Inglaterra) e Robert Solow (Cambridge, Massachusetts). Robinson argumentava que a análise marginal do capital é circular e enganadora. Solow defendia que a análise do capital era coerente e útil. Em relação à mensuração do capital, Robinson argumentou que não se deveria agregar uma lista de entradas de capital para construir uma medida global. Por sua vez, Solow aprovava a agregação neoclássica com base na análise marginal do capital (COHEN; HARCOURT, 2003).

anterior à t , pode optar por adquirir um veículo em $t + 1$ ou permanecer sem veículo. Se o consumidor adquiriu um veículo em t , então esse carro começa a experimentar um processo de depreciação em $t + 1$, e em cada momento do tempo o proprietário terá que decidir novamente entre manter seu carro usado, vender ou descartar³. Se o consumidor decide revender ou descartar, precisa decidir novamente se irá adquirir algum veículo entre as opções restantes no conjunto $J_t \cup 0$. O consumidor vê-se então diante do mesmo problema de decisão no período $t + 1$, precisando maximizar a utilidade esperada em um conjunto de restrições de informações e recursos. Os $j \in \mathbf{J}_t$ veículos possuem características físicas observadas x_{jt} , não observadas (pelo econometrista) ξ_{jt} , preços p_{jt} e custos de transação τ_{jt} .

O modelo de preços hedônicos de [Bajari et al. \(2012\)](#) adota uma abordagem parecida ao tratar separadamente características observáveis e não-observáveis. Os custos de transação, abrangendo os custos de busca por informação, custos financeiros e informação assimétrica, são pagos pelos consumidores. Não existem custos de transação se o consumidor escolhe o veículo *outside*: $p_{0t} = 0$. Os consumidores são heterogêneos em relação à valoração dos atributos dos carros. Em cada período t as preferências dos consumidores possuem choques aleatórios dados por: $\epsilon_{it} = (\epsilon_{i0t}, \epsilon_{i1t}, \dots, \epsilon_{iJ_t t})$. Os choques aleatórios são independentes entre (i, j, t) . O consumidor deriva uma das utilidades \tilde{u}_{it}^k ou u_{it}^k para a possível escolha em t .

Importante lembrar que os bens duráveis usados que não se encontram no mercado para revenda também possuem um valor residual ([HULTEN; WYKOFF, 1981](#)), o qual pode ser mensurado pelo valor de um bem idêntico e disponível para revenda no mercado. Entretanto, existe um debate na literatura sobre a equivalência de preços entre ativos usados disponíveis e indisponíveis para a revenda, questionando se os veículos oferecidos para revenda possuem a mesma qualidade dos veículos fora do mercado. A literatura de *lemon cars* têm se ocupado em avaliar se os veículos são comercializados à uma taxa de desconto de risco sistemático ([AKERLOF, 1970](#)). Diferenças sistemáticas de qualidade entre os veículos usados fora do mercado e disponíveis para revenda comprometeriam a expansão

³através de *scraping*

das análises de métodos de preços hedônicos e outros métodos econométricos para avaliar toda a economia. No presente estudo, isto implicaria que os veículos fora do mercado estariam sujeitos a um processo de depreciação diferente dos veículos comercializados, e conseqüentemente, a interpretação dos modelos de crescimento exponencial ficaria restrita aos veículos comercializados.

Segundo [Sawhill \(2010\)](#), bens duráveis mais antigos experimentam queda de preços por dois motivos: (i) O final do ciclo de vida produtivo do capital está mais próximo, e (ii) os ativos mais antigos são menos rentáveis, pois apresentam produtividade mais baixa e gastos com reparos e manutenção mais altos. Em qualquer ponto do tempo, a relação entre idade e preço é dada por uma curva inclinada para baixo, tal como na [Figura 5.2](#). As maiores taxas de declínio nos preços ocorrem nos anos iniciais de vida do produto. Possuir um veículo novos implica em mais depreciação que a posse de um veículo antigo, levando a um relacionamento não-linear entre preço e idade de carros ([FENG *et al.*, 2013](#)).

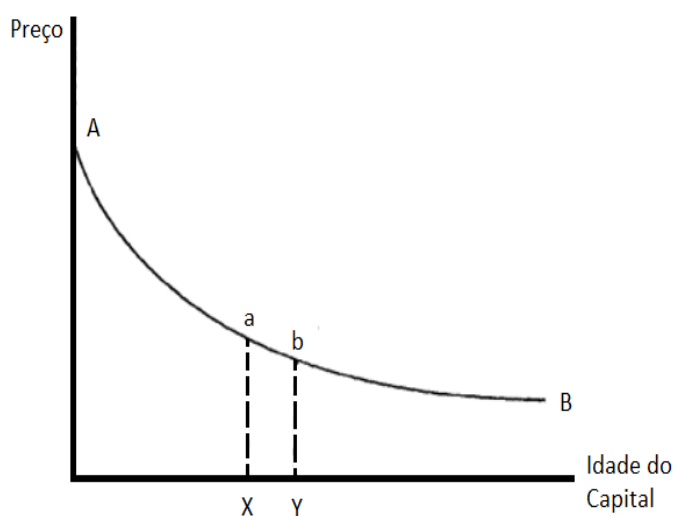


FIGURA 5.2 – Preço e Idade de um Bem Durável

Uma relação entre preço e idade representada por uma linha reta estaria retratando que o valor do ativo diminui a taxas constantes com a idade.

Na [Figura 5.2](#), o valor de um bem com idade X é representado pelo ponto “a”, enquanto o valor de um bem com idade Y é representado pelo ponto “b”. A depreciação econômica é dada pela diferença entre as projeções dos pontos “a” e “b” no eixo vertical. Quando AB é representada por uma linha reta, então a taxa de depreciação econômica é crescente

com o tempo.

Na ausência de inflação e de obsolescência física, a curva AB delinea uma trajetória de preço de um bem durável com o tempo. Dois produtos com idade X e Y teriam a mesma diferença de preços quando observados em 2010 e 2015, entretanto, devido à inflação que ocorre neste período, esta equivalência não é válida. Desta forma, a mudança de preço de um bem durável possui dois componentes, um devido à depreciação econômica e outro devido à inflação.

Na mensuração do valor de bens duráveis, além da distinção entre depreciação e inflação, deve-se distinguir também a obsolescência tecnológica. Os produtos desenvolvidos em 2015 incorporam desenvolvimentos tecnológicos e de design que os tornam superiores aos produtos desenvolvidos em 2010. Particularmente no mercado de carros, cada nova geração de veículos possui melhores tecnologias de economia de combustível, incorporando melhores itens de segurança em acidentes e design de absorção de batidas (SAWHILL, 2010). Considerando o ano que diferentes coortes de carros foram produzidas, pode-se tentar enquadrar a questão da superioridade tecnológica de uma geração sobre a geração anterior, tornando os produtos novos mais valiosos que os produtos anteriores.

5.3 Metodologia

Aqui apresentamos maiores detalhes metodológicos sobre os modelos de crescimento exponencial, incluindo estimação e interpretação dos parâmetros.

5.3.1 Modelos de Crescimento Não-Lineares

Um modelo de regressão não-linear é uma generalização do modelo de regressão linear, onde a variável explicativa é representada por uma função não-linear, por exemplo: $E(y_i) = \frac{\exp(\theta_1)}{1 + \exp(\theta_2)}$ (SOUZA, 1998), onde os parâmetros podem ser estimados através de mínimos quadrados não-lineares, sempre que os erros são aditivos (LINDSTROM; BATES, 1990; GOSHU; KOYA, 2013).

A classe de modelos de crescimento não-linear é tratada por Souza (1998), Annadurai *et al.* (2000), Grimm (2009) e Goshu e Koya (2013) para modelagem de crescimento de culturas em biologia, engenharia florestal, agricultura, demografia e agronomia. O modelo de crescimento logístico é um dos casos dos modelos de crescimento exponencial. Em demografia, usa-se o modelo logístico de crescimento para populações (KAHM *et al.*, 2010):

$$w(t) = m(t; \Theta^*) = \frac{\theta_1^*}{1 + \theta_2^* \exp[-(\theta_3^* + \theta_4^* x)]}$$

onde w é o crescimento populacional, t é o tempo em anos e $m(t; \Theta^*)$ representa a função média, a qual depende do vetor de parâmetros $\Theta^* = (\theta_1^*, \theta_2^*, \theta_3^*, \theta_4^*)$ e do vetor de variáveis explicativas t . Por exemplo, se $\theta_3^* > 0$ conforme t aumenta, o termo $\exp[-(\theta_3^* + \theta_4^* x)]$ se aproxima de zero e a função $m(t; \Theta^*)$ se aproxima da assíntota θ_1^* . Assim, ao assumir um crescimento populacional logístico, está se impondo um limite de saturação populacional dado por θ_1^* . A interpretação de θ_3^* e θ_4^* está relacionada ao ponto onde a população atinge metade do seu crescimento. Ao utilizar o modelo de crescimento logístico, assume-se simetria no valor de t onde $m(t)$ atinge a metade do crescimento: $m(t = \frac{\theta_3^*}{\theta_4^*}; \Theta^*) = \frac{\theta_1^*}{2}$, ou seja, o modelo logístico é simétrico no ponto $t = \frac{\theta_3^*}{\theta_4^*}$.

Os resultados dos parâmetros estimados dos modelos de crescimento exponencial permitiram caracterizar a dinâmica temporal entre perda de valor e a idade do veículo. Considerando o modelo de crescimento logístico⁴, o seu ponto de inflexão ($t = \frac{\theta_3^*}{\theta_4^*}$) indicaria a idade de transição onde o veículo começa o seu processo de estabilização de preço, e o parâmetro θ_1^* representa a desvalorização máxima atingida pelo carro.

A expressão matemática do ponto de inflexão do modelo de crescimento logístico assume que a mudança do processo de desvalorização ocorre na mediana da curva de desvalorização (ANNADURAI *et al.*, 2000). Entretanto, existem outros modelos de crescimento que não possuem este pressuposto. Goshu e Koya (2013) desenvolvem os pontos de infle-

⁴Não confundir com o modelo probabilístico *logit* para variáveis dependentes binárias. O modelo probabilístico logístico herda seu nome da sua função de ligação $g(\mu_i) = \frac{\ln \mu_i}{1 - \ln \mu_i}$, onde $g(\mu_i) = X_i \beta$ (DEMIDENKO, 2013)

xão para os modelos de crescimento exponencial: Logístico Generalizado, Richards, Von Bertalanffy, Brody, Gompertz, Weibull Generalizada, Monomolecular e Mitscherlich. Em todos os casos, o ponto de inflexão é obtido a partir dos parâmetros estimados e possui importante significado em diversas áreas de aplicação. O Gráfico 5.3 apresenta algumas curvas exponenciais de crescimento retiradas de Goshu e Koya (2013):

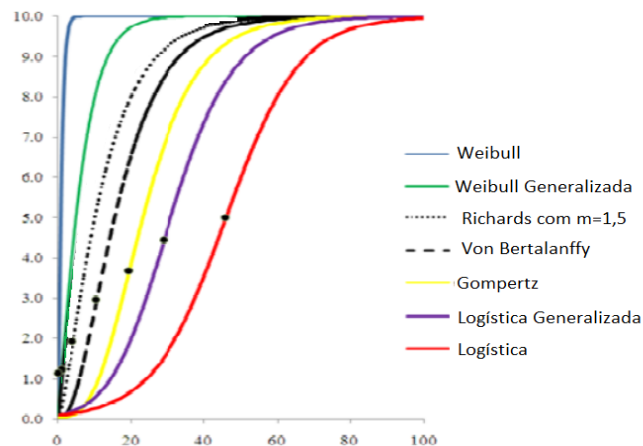


FIGURA 5.3 – Curvas de Crescimento Exponencial

Sempre que a evolução da variável dependente se caracterizar pela existência de fases e períodos, tais como adolescência, juventude e maturidade, pode-se utilizar uma função sigmoide⁵, contendo ponto de inflexão nos períodos da juventude e apresentando um ponto de estabilidade no período de maturidade (Gráfico 5.3). Tais funções são relativamente mais difíceis de estimar em relação aos modelos OLS.

Seja modelo não-linear no formato:

$$w_i = f(t_i, B) + \epsilon_i, \quad i=1,2,\dots,n.$$

onde w é a variável dependente, t é a variável independente e B é um vetor representando os parâmetros a serem estimados: $(\beta_1, \beta_2, \dots, \beta_p)$, p é o número de parâmetros a serem estimados, n é o número de observações e ϵ_i é um erro aleatório que entra na equação de forma aditiva. A estimativa de $(\beta_1, \beta_2, \dots, \beta_p)$ é obtida através da minimização da soma

⁵Função sigmoide é uma função cujo gráfico possui formato em “S”, por exemplo, $f(x) = \frac{1}{1 + \exp(-\lambda x)}$

de quadrados dos resíduos (SQRes):

$$\text{SQRes} = \sum_i^n [w_i - f(t_i, B)]^2 \quad (5.3.1)$$

Sob os pressupostos de independência das observações e distribuição normal de ϵ_i com média zero e variância σ^2 , sendo w_i e t valores fixos, a soma dos resíduos é uma função de B . As estimativas de mínimos quadrados de B são encontrados diferenciando a Expressão (5.3.1) em relação a cada um dos parâmetros $(\beta_1, \beta_2, \dots, \beta_p)$ e igualando a zero:

$$\sum_i^n w_i - f(t_i, B) \frac{\partial f(t_i, B)}{\beta_j} = 0$$

As equações resultantes são denominadas equações normais e possuem formato não-lineares (SOUZA, 1998).

A Tabela 5.1 apresenta o formato matemático das funções de crescimento exponencial mais populares conforme notação utilizada em Fekedulegn *et al.* (1999). Para a classe de modelos considerados na Tabela 5.1, não é possível obter uma solução analítica de estimação dos parâmetros, lançando-se mão de métodos iterativos para minimização da soma de quadrados dos resíduos.

TABELA 5.1 – Modelos Não-Lineares de Crescimento Exponencial

Caso	Modelo de Crescimento Não-Linear
Exponencial Neg.	$w(t) = \beta_0 (1 - \exp(-\beta_2 t))$
Monomolecular	$w(t) = \beta_0 (1 - \beta_1 \exp(-\beta_2 t))$
Mitcherlich	$w(t) = (\beta_0 - \beta \beta_2^t)$
Gompertz	$w(t) = \beta_0 [\exp(-\beta_1 \exp(-\beta_2 t))]$
Logística	$w(t) = \frac{\beta_0}{[1 + \beta_1 \exp(\beta_2 t)]}$
Von Bertalanffy	$w(t) = (\beta_0^{1-\beta_3} - \beta_1 [\exp(\beta_2 t)])^{1/(1-\beta_3)}$
Richards	$w(t) = \frac{\beta_0}{(1 + \beta_1 \exp(\beta_2 t))^{1/\beta_3}}$
Weibull	$w(t) = \beta_0 - \beta_1 \exp(-\beta_2 t^{\beta_3})$

Nem sempre os parâmetros dos modelos não-lineares de crescimento possuem fácil

interpretação. O parâmetro mais facilmente interpretado é o β_0 , o qual representa a assíntota da curva de depreciação ou ponto máximo de crescimento. O parâmetro β_1 é uma constante intrínseca, β_2 é o parâmetro que governa a taxa de crescimento da variável dependente em relação ao crescimento do tempo, e por último β_3 é denominada constante alométrica. Em biologia, a constante alométrica refere-se às mudanças nas dimensões relativas das partes de um organismo que são correlacionados com mudanças no seu tamanho como um todo. No presente estudo, os principais parâmetros de interesse da análise serão o parâmetro β_0 e o ponto de inflexão.

Os métodos iterativos de estimação dos parâmetros dos modelos não-lineares requerem o desenvolvimento de derivadas parciais em relação a cada um dos parâmetros⁶: β_0 , β_1 , β_2 e β_3 . Estas derivadas são apresentadas nas tabelas no Apêndice D.1.

5.3.1.1 Valores Iniciais

A especificação dos valores iniciais é um dos principais problemas enfrentados no ajuste de modelos não lineares (FEKEDULEGN *et al.*, 1999). Valores iniciais inadequados resultam em processos iterativos mais longos, com falta de convergência e possibilidade de obtenção de máximos locais ao invés de máximos globais (SOUZA, 1998).

Fox e Weisberg (2011) ilustram um método para determinação do valor inicial para o modelo de crescimento exponencial logístico. Supondo inicialmente o valor máximo observado na variável resposta como sendo o valor inicial para o parâmetros β_0 , os valores iniciais para os demais parâmetros podem ser obtidos reparametrizando o modelo de

⁶O procedimento *NLIN* do *SAS* requer a especificação das derivadas parciais em relação aos parâmetros dos modelos não lineares. Estas derivadas parciais devem ser inseridas na sintaxe de programação. Já o procedimento *MODEL* do *SAS* e a função *nls* do **R** não necessitam da especificação das derivadas parciais (FOX; WEISBERG, 2011).

crescimento logístico para obter-se uma equação de regressão linear:

$$\begin{aligned}
 w(t) &= \frac{\alpha}{1 + \beta \exp(k t)} & (5.3.2) \\
 \frac{w(t)}{\alpha} &= \frac{1}{1 + \beta \exp(k t)} \\
 1 - \frac{w(t)}{\alpha} &= \frac{\exp(k t)}{1 + \beta \exp(k t)} \\
 \frac{w(t)/\alpha}{1 - w(t)/\alpha} &= \frac{1}{\beta \exp(k t)} \\
 \ln \left[\frac{w(t)/\alpha}{1 - w(t)/\alpha} \right] &= k t
 \end{aligned}$$

onde a Expressão (5.3.3) pode ser estimada através de um modelo OLS para obter um valor inicial para o parâmetro k .

As dificuldades de especificação de valores iniciais podem ser contornadas entendendo-se o significado dos parâmetros dos modelos. O parâmetro β_0 é o valor máximo possível que a variável dependente pode atingir. O parâmetro β_2 é interpretado como sendo a velocidade pela qual a variável resposta atinge seu valor máximo: β_0 . Com base nesta definição, podemos obter a Expressão (5.3.3):

$$\beta_2 \Rightarrow \frac{(w_2 - w_1)/(t_2 - t_1)}{\beta_0^*} \quad (5.3.3)$$

onde w_1 e w_2 são valores da variável resposta nos períodos de tempo t_1 e t_2 , tendo β_0^* como “chute” do valor inicial de β_0 .

Para cada um dos modelos de crescimento, os valores iniciais para o parâmetro β_1 podem ser obtidos avaliando o estágio de crescimento inicial. A Tabela 5.2 mostra os valores iniciais para o parâmetro β_1 quando $t = 0$, onde $w(0)$ é a magnitude da variável resposta no momento inicial, em algumas situações igual a zero.

Fazendo $w(0) = 0$, as funções Monomolecular e Von Bertalanffy recebem parâmetro inicial β_1^* igual a 1, enquanto as funções Mitcherlich e Weibull recebem valor inicial β_1^* igual a β_0 . Não é possível fazer $w(0) = 0$ para obter os valores iniciais das funções

TABELA 5.2 – Determinação dos Valores Iniciais

Caso	Valor de $w(0)$ no momento inicial	Aproximação para β_1
Exponencial Negativa	Não Existe	-
Monomolecular	$w(0) = \beta_0 (1 - \beta_1)$	$\beta_1^* = 1 - \frac{w(0)}{\beta_0}$
Mitcherlich	$w(0) = \beta_0 - \beta_1$	$\beta_1^* = \beta_0 - w(0)$
Gompertz	$w(0) = \beta_0 \exp(-\beta_1)$	$\beta_1^* = \ln[\beta_0] - \ln[w(0)]$
Logística	$w(0) = \frac{\beta_0}{1 + \beta_1}$	$\beta_1^* = \frac{\beta_0}{w(0)} - 1$
Von Bertalanffy	$w(0) = \beta_0(1 - \beta_1)^{\frac{1}{1 - \beta_3}}$	$\beta_1^* = 1 - \left[\frac{w(0)}{\beta_0}\right]^{1 - \beta_3}$
Richards	$w(0) = \frac{\beta_0}{\frac{1}{(1 - \beta_1)^{\beta_3}}}$	$\beta_1^* = \left[\frac{\beta_0}{w(0)}\right]^{\beta_3} - 1$
Weibull	$w(0) = \beta_0 - \beta_1$	$\beta_1^* = \beta_0 - w(0)$

Gompertz, Logística e Richards, assumindo-se um valor pequeno para β_0 , mas diferente de zero. Fazendo $w(0) = 0$, a função Richards recebe valor inicial: $\beta_1^* = \beta_0^{[1 - \beta_3]}$.

5.3.2 Pontos de Inflexão

Ponto de inflexão⁷ é um ponto sobre uma curva onde a derivada de segunda ordem troca de sinal, ou seja, é o ponto onde ocorre uma mudança no sentido da concavidade, sendo uma função contínua $w(t)$ duas vezes diferenciável.

A importância do ponto de inflexão reside no fato deste representar o momento do tempo onde termina a fase de desvalorização mais intensa e começa a fase de estabilização do processo de depreciação. A Tabela 5.3 apresenta as expressões finais para os pontos de inflexão dos modelos de crescimento exponencial. As derivações matemáticas dos pontos de inflexão são apresentadas no Apêndice (D.3).

Os modelos de crescimento Gompertz, Logístico, Von Bertalanffy, Richards e Wei-

⁷Formalmente, seja I um intervalo aberto em \mathfrak{R} , e seja um ponto a tal que $a \in I$. Seja $w(t)$ uma função contínua em I . A função $w(t)$ tem um ponto de inflexão em $(a, w(a))$, se existe um $\epsilon > 0$ tal que o gráfico da função tem a concavidade voltada para cima (para baixo) em $]a - \epsilon, a[$ e concavidade voltada para baixo (para cima) em $]a, a + \epsilon[$.

TABELA 5.3 – Modelos matemáticos não-lineares

Função	Ponto de Inflexão	Depreciação no Ponto de Inflexão
Exponencial Negativa	-	-
Monomolecular	-	-
Mitcherlich	-	-
Gompertz	$a = \frac{\ln \beta_1}{\beta_2}$	$w(a) = \frac{\beta_0}{e}$
Logístico	$a = \frac{\ln \beta_1}{\beta_0}$	$w(a) = \frac{\beta_0}{2}$
Von Bertalanffy	$a = \frac{1}{\beta_2} \ln(\beta_1/[1 - \beta_3])$	$w(a) = \frac{1}{\beta_3^{1 - \beta_3}} \beta_0$
Richards	$a = \frac{\ln(\beta_1/\beta_3)}{\beta_2}$	$w(a) = \beta_0 [1 - \beta_3]^{\beta_3}$
Weibull	$a = [\frac{\beta_3 - 1}{\beta_2 \beta_3}]^{1/\beta_3}$	$w(a) = \beta_0 - \beta_1 [\exp(\frac{\beta_3 - 1}{\beta_3})]$

bull permitem calcular o ponto de inflexão, enquanto os modelos Exponencial Negativo, Monomolecular e Mitcherlich não permitem obter pontos de inflexão.

5.4 Dados Utilizados

O banco de dados utilizado neste capítulo foi apresentado em maiores detalhes na Subseção 2.4, onde mostramos que existem mais de 600 modelos distintos de carros. Para não inflacionar a quantidade de modelos estimada foi preciso adotar algum critério de seleção de veículos representativos. Para isto, foi utilizada a classificação de veículos por categoria de design: Conversível, Cupê, Hatchback, Minivan, Perua, Picape, Sedã e Utilitário, selecionando os modelos mais frequentes em cada categoria. Os veículos resultantes encontram-se na Tabela 5.4.

Os modelos Discovery-Land Rover (R\$169.211), Hilux-Toyota (R\$106.122) e SLK-Mercedes-Benz (R\$146.789) são os veículos mais caros analisados, o que provavelmente irá impactar na depreciação máxima atingida por estes veículos e no ponto de inflexão (Tabela 5.4).

TABELA 5.4 – Modelos de Crescimento Não-Linear: Logistic

Modelo Representativo	Marca	Categoria	Preço Médio(R\$)	Frequência
A6	Audi	Perua	104.576	67
Carnival	Kia	Minivan	69.344	46
Corolla	Toyota	Sedã	44.525	1.395
Discovery	Land Rover	Utilitário	169.211	582
Hilux	Toyota	Picape	106.122	2.531
Palio	Fiat	Hatchback	26.144	3.470
Passat	Volkswagen	Cupê	65.076	446
SLK	Mercedes-Benz	Conversível	146.789	185

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

Os Gráficos 5.4 e 5.5 mostram a evolução temporal dos preços dos veículos: Audi-A6 (Perua), Kia-Carnival (Minivan), Toyota-Corolla (Sedã), Land Rover-Discovery (Utilitário), Toyota-Hilux (Picape), Fiat-Palio (Hatchback), Volkswagen-Passat (Cupê) e Mercedes-Benz-SLK (Conversível). Vemos que o processo de desvalorização dos veículos segue uma tendência não linear.

5.5 Resultados

Prieto *et al.* (2015) apontam a idade como um dos principais determinantes dos preços de veículos usados. As demais variáveis apontadas são quilometragem, potência do motor, tipo de motor (diesel ou gasolina), segmento de mercado, cor e demais itens opcionais, tais como: pintura metálica, freio ABS, *Cruise Control*⁸, computador de bordo e ar condicionado. Nesta Subseção, a variável idade será a única variável incluída nos modelos não lineares para explicar a evolução da depreciação de preços dos veículos.

As abordagens econometria que utilizam várias variáveis explicativas geralmente utilizam modelos lineares. A utilização de um modelo não linear utilizando várias variáveis

⁸ *Cruise Control*: Sistema de programação de velocidade, onde uma vez que a velocidade programada é atingida, pode-se retirar o pé do acelerador e a velocidade se manterá constante. Isto possibilita um maior conforto da condução em uma viagem.

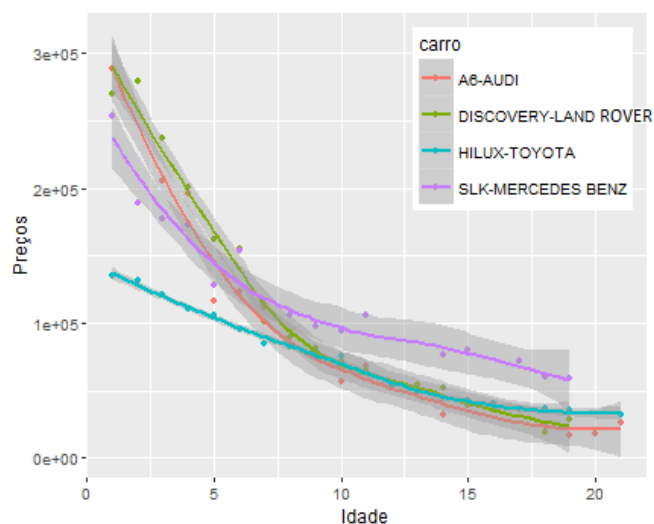


FIGURA 5.4 – Evolução dos Preços do Audi-A6 (Perua), Land Rover-Discovery (Utilitário), Toyota-Hilux (Picape) e Mercedes-Benz-SLK (Conversível)

explicativas tornaria o processo de estimação excessivamente complexo e computacionalmente impraticável. Tanto os modelos de crescimento não-lineares considerados na Subseção 5.3.1 como também os modelos contábeis da Subseção 5.2.3 contém uma única variável explicativa. Outra justificativa para modelagem econométrica da depreciação considerando somente uma variável explicativa encontra-se no fato de estarmos utilizando agregação de preços médios de veículos por ano, ou seja, não estamos modelando diretamente uma seccional dos preços para cada um dos carros, tal como feito no capítulo 4. Desta forma, teríamos dificuldade em construir agregações para as variáveis que representam a existência de itens de conforto e segurança, tais como Ar Condicionado, Freio ABS, Computador de Bordo, entre outras.

No presente modelo, a variável dependente, chamada de depreciação, é definida como sendo a diferença entre o valor do veículo no tempo t menos o valor máximo do veículo no tempo inicial t_0 normalizada pelo valor máximo: $depre_t = \left[\frac{p_{t_0} - p_t}{p_{t_0}} \right]$, $depre_t \in [0, 1]$ (FENG *et al.*, 2013). Os pontos de inflexão serão calculados utilizando os parâmetros estimados dos modelos não-lineares, conforme Tabela 5.3. Os modelos Negative Exponencial, Monomolecular e Mitcherlich não possuem ponto de inflexão, sendo apresentados visando avaliar a qualidade do ajuste dos modelos não lineares para cada um dos carros (Tabela D.5). Os modelos Gompertz, Logistic, VonBertalanffy, Richards e Weibull possuem ponto

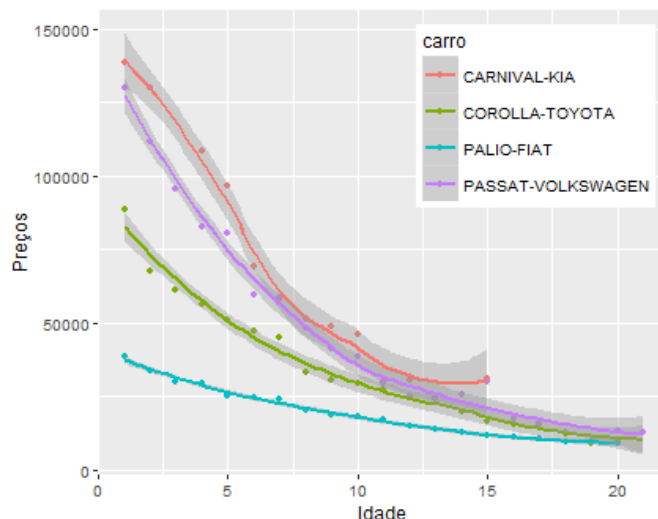


FIGURA 5.5 – Evolução dos Preços dos Veículos Kia-Carnival (Minivan), Toyota-Corolla (Sedã), Fiat-Palio (Perua) e Volkswagen-Passat (Cupê)

de inflexão, os quais foram apresentados anteriormente na Tabela 5.3.

Vamos apresentar os resultados estimados para os modelos Logístico e Richards (Tabelas D.6 e D.7). As Tabelas no Apêndice D.3, contém os parâmetros estimados, coeficientes de determinação (R^2) e os pontos de inflexão (“ a ”) para os modelos não apresentados neste capítulo.

Os pontos de inflexão foram obtidos segundo as expressões da Tabela 5.3. A estimativa de depreciação máxima dos veículos (β_0) difere consideravelmente dos resultados dos modelos sem ponto de inflexão (Tabela D.5), e a coluna $w(a)$ mostra o valor da variável dependente no ponto de inflexão, ou seja, qual a depreciação sofrida pelo veículo antes do ponto de inflexão.

O modelo de crescimento logístico possui ponto de inflexão ocorrendo 50% do valor da depreciação total: $w(a) = \left[\frac{\beta_0}{2} \right]$ (Tabela 5.3), possuindo depreciação no ponto de inflexão ($w(a)$) relativamente mais elevada que os demais modelos. Observamos que os valores para a variável dependente no ponto de inflexão variam entre 34,7% e 43%. Os modelos Fiat-Palio e Toyota-Corolla apresentaram desvalorização anual antes do ponto de inflexão ($w(a)/a$) de 20%, enquanto o modelo Kia Carnival apresentou a menor desvalorização anual antes do ponto de inflexão (9%).

TABELA 5.5 – Modelos de Crescimento Não-Linear: Logistic

Carro		β_0	β_1	β_2	a	$w(a)$	$w(a)/a$	R^2
A6	Parâm.	0,86	8,48	0,45	2,5	43,0%	0,17	97,4
Audi	Est. t	52,4*	7,6*	13,4*				
Carnival	Parâm.	0,69	15,04	0,71	3,9	34,7%	0,09	98,8%
Kia	Est. t	96,9*	5,4*	12,0*				
Corolla	Parâm.	0,84	5,92	0,31	2,1	42,1%	0,20	98,4%
Toyota	Est. t	42,3*	13,7*	17,5*				
Discovery	Parâm.	0,86	10,65	0,50	2,8	42,8%	0,16	97,4%
Land Rover	Est. t	46,0*	6,2*	11,6*				
Hilux	Parâm.	0,75	9,59	0,33	3,0	37,4%	0,12	99,4%
Toyota	Est. t	34,5*	17,0*	22,1*				
Palio	Parâm.	0,76	4,41	0,27	2,0	38,1%	0,20	99,6%
Fiat	Est. t	46,8*	33,1*	27,9*				
Passat	Parâm.	0,86	7,20	0,39	2,3	43,0%	0,19	98,2%
Volkswagen	Est. t	52,8*	11,1*	14,3*				
SLK	Parâm.	0,65	5,94	0,34	2,7	32,7%	0,12	95,4%
Mercedes-Benz	Est. t	28,0*	7,1*	9,7*				

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “*ns*” Não possui significância estatística.

Segundo o modelo de Richard, os veículos Toyota-Corolla e Volkswagen-Passat possuem depreciação máxima de 92%, enquanto os veículos Kia-Carnival e Toyota-Hilux possuem depreciação máxima de 71% e 78%, respectivamente, e os veículos Fiat-Palio e SLK-Mercedes Benz possuem depreciação máxima de 92% e 83%, respectivamente (Tabela 5.6).

O Toyota-Hilux possui a menor taxa de depreciação anual ($w(a)/a$) antes do ponto de inflexão (16%), indicado ser uma opção de veículo zero quilômetro para consumidores que pretendem revender seu veículo antes dos 5 anos de idade. A taxa de depreciação anual antes do ponto de inflexão para o automóvel Volkswagen-Passat ($w(a)/a = 47\%$) indicando ser uma opção de compra ruim para consumidores de veículos zero quilômetro que pretendem revender seus carros antes do primeiro ano de idade.

Os parâmetros β_1 e β_2 não apresentaram significância estatística para os veículos

TABELA 5.6 – Modelos de Crescimento Não-Linear: Richards

Carro		β_0	β_1	β_2	β_3	a	$w(a)$	$w(a)/a$	R^2
A6	Parâm.	0,91	-1,17	0,20	-0,94	1,1	49%	0,43	99,6%
Audi	Est. t	73,8*	-15,5*	8,5*	-4,6*				
Carnival	Parâm.	0,71	0,03	0,49	0,01	3,0	71%	0,24	99,0%
Kia	Est. t	104,8*	0,0 ^{ns}	4,4*	0,0 ^{ns}				
Corolla	Parâm.	0,92	-1,01	0,14	-0,87	1,1	53%	0,50	99,6%
Toyota	Est. t	45,7*	-10,9*	6,2*	-3,3*				
Discovery	Parâm.	0,93	-1,32	0,18	-1,22	0,4	35%	0,80	99,7%
Land Rover	Est. t	82,6*	-29,4*	9,2*	-5,7*				
Hilux	Parâm.	0,78	-0,28	0,21	-0,10	4,9	78%	0,16	99,7%
Toyota	Est. t	38,9*	-0,8 ^{ns}	9,6*	-0,7 ^{ns}				
Palio	Parâm.	0,80	-0,31	0,17	-0,18	3,1	78%	0,25	99,8%
Fiat	Est. t	70,6*	-1,3*	10,8*	-1,1*				
Passat	Parâm.	0,92	-1,10	0,17	-0,92	1,1	51%	0,47	99,8%
Volkswagen	Est. t	90,0*	-24,8 ^{ns}	12,8*	-7,2 ^{ns}				
SLK	Parâm.	0,83	-1,08	0,1	-1,48	-4,2	-	-	99,4%
Mercedes-Benz	Est. t	12,1*	-60*	3,5*	-8,3*				

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “ns” Não possui significância estatística.

Kia-Carnival, Land Rover-Discovery, Toyota-Hilux e Volkswagen-Passat. Entretanto, o parâmetro representando a depreciação máxima (β_0) apresentou significância estatística para todos os carros. O Kia-Carnival possui a maior taxa de mudança média da variável dependente entre dois períodos consecutivos ($\beta_2 = 0,49$), enquanto o modelo Mercedes-Benz-SLK possui a menor taxa de mudança média da variável dependente entre dois períodos consecutivos ($\beta_2 = 0,1$).

Para diferentes veículos, o modelo de Richards apresentou pontos de inflexão variando entre 0,49 e 4,9. Por exemplo, o veículo Audi-A6 possui ponto de inflexão igual 1,1 ano e taxa de depreciação no ponto de inflexão de 49%, ou seja, praticamente a metade da depreciação do veículo ocorre no primeiro ano. O modelo de Richards também produziu ponto de inflexão negativo para o veículo Mercedes-Benz-SLK. Isto se deve à razão entre os parâmetros estimados para β_1 e β_3 ter apresentado valor menor que 1, conseqüentemente

o logaritmo $a = \ln[-\beta_1/\beta_3]$ possui valor negativo.

Os Gráficos 5.7 e 5.8 mostram que existe proximidade entre os ajustes dos modelos Logístico e Richards. Especialmente para os veículos Toyota-Corolla (Sedã), Volkswagen-Passat (Cupê) e Audi-A6 (Perua), as curvas dos valores observados e preditos são muito próximos, e os valores estimados para a depreciação máxima encontram-se bem próximos.

Para o modelo Kia-Carnival (Minivan), os modelos não-lineares Logístico e Richards encontram-se ligeiramente acima da curva de depreciação nos período antes do ponto de inflexão, e encontram-se ligeiramente acima da curva de depreciação observada no período após o ponto de inflexão. Ao mesmo tempo, a depreciação máxima estimada encontra-se ligeiramente abaixo da depreciação observada (Gráficos 5.7 e 5.8). Entretanto, de forma geral, os modelos Richards e Logístico, analisados nesta seção, capturam adequadamente o processo de depreciação dos veículos analisados.

Os modelos de crescimento exponencial mostraram-se importantes ferramentas quantitativas para análise do processo de depreciação. Os pontos de inflexão mostram que veículos de diferentes categorias possuem comportamentos bem distintos no seu processo de depreciação. As tabelas e gráficos contendo os resultados estimados para os modelos: encontram-se no Apêndice D.3.

5.6 Conclusões

No presente capítulo analisamos os modelos de crescimento exponencial aplicados à depreciação de preços dos veículos representativos das categorias: Perua, Minivan, Sedã, Utilitário, Picape, Hatchback, Cupê e Conversível. Observamos que os modelos de crescimento exponencial permitem caracterizar o padrão de perda de preços, revelando interessantes aspectos da curva de depreciação, tais como o ponto de inflexão e a estimativa de depreciação máxima dos veículos.

De forma geral, os valores observados e preditos apresentam proximidade, indicando razoável qualidade dos modelos ajustados. Os gráficos de depreciação de preços e os

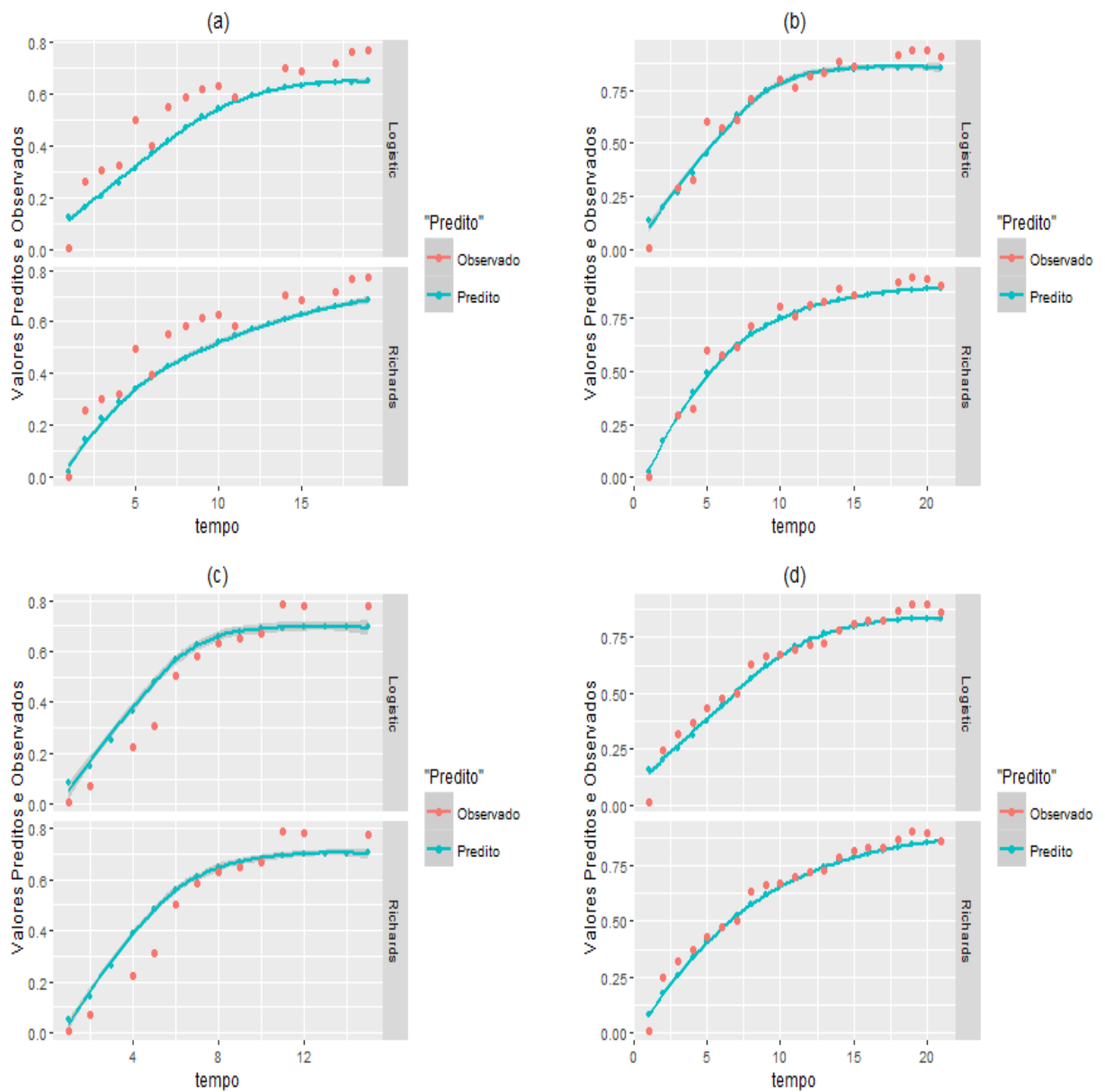
valores dos coeficientes de determinação (Tabela D.9 - Apêndice D.3) confirmam que, de fato, o processo de desvalorização dos veículos segue uma tendência não linear, conforme havia sido apontado pela literatura de bens duráveis.

Considerando os resultados do modelo logístico, o ponto de inflexão ($w(a)$) do veículo Audi-A5 foi de 2,5 anos, enquanto o Audi-Carnival possui ponto de inflexão de 3,9 anos e o Mercedes-Benz-SLK possui ponto de inflexão em 2,7 anos (Tabela 5.5). A desvalorização anual antes do ponto de inflexão apresentou grande variabilidade por categoria de veículo, assim como a depreciação máxima estimada.

Optamos por complementar a análise com um indicador de depreciação anual antes do ponto de inflexão ($w(a)/a$). O veículo Toyota-Corolla, apesar de possuir ponto de inflexão no primeiro ano de vida, possui desvalorização anual (50%) maior que o veículo Kia-Carnival (24%), que possui ponto de inflexão em 3 anos.

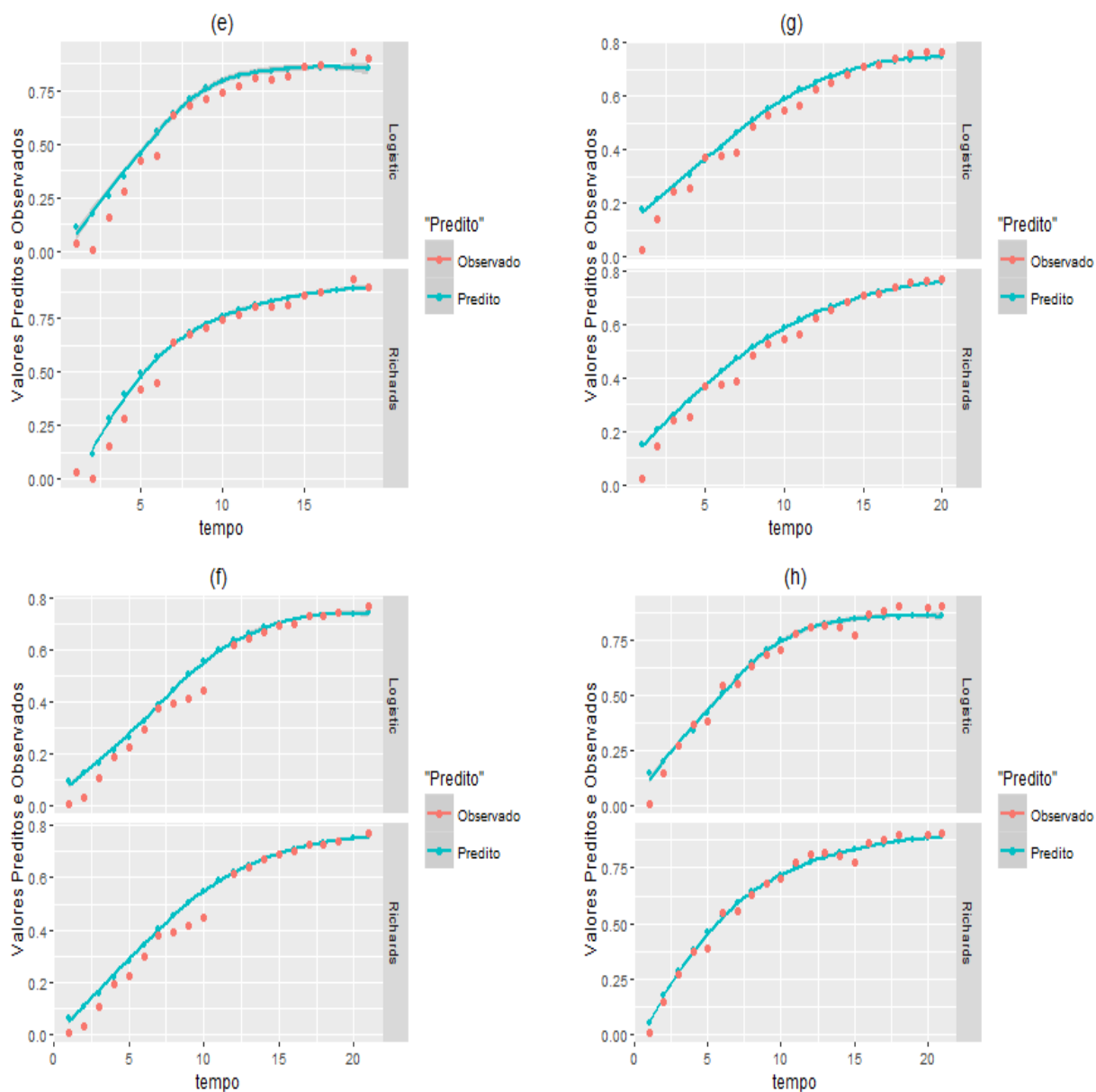
Os resultados sugerem que a decisão entre adquirir um veículo novo ou usado depende da categoria escolhida e do tempo em que se pretende manter a posse desses veículos.

TABELA 5.7 – Gráficos dos Valores Estimados e Observados



FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.
OBS: Valores Preditos (Linhas) e Observados (Pontos) para os veículos: Mercedes-Benz-SLK (a), Audi-A6 (b), Kia-Carnival (c) e Toyota-Corolla (d).

TABELA 5.8 – Gráficos dos Valores Estimados e Observados



FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.
OBS: Valores Preditos (Linhas) e Observados (Pontos) para os veículos: Land Rover-Discovery (e), Toyota-Hilux (f), Fiat-Palio (g) e Volkswagen-Passat (h).

6 Conclusão

O advento do *Big Data* trouxe contribuições importantes na Medicina, Biologia, Indústria e Marketing. Entretanto, não se encontram muitas aplicações em pesquisa econômica, especialmente em um momento onde os cientistas econômicos são desafiados a adentrar nos aspectos computacionais de análise de dados de *Big Data* (VARIAN, 2014). Nesta tese nós contribuímos para a literatura demonstrando como algumas técnicas de *Big Data* podem contribuir em pesquisa econômica. Mais particularmente as técnicas de extração de dados na internet, que permitem reunir informações relevantes economicamente, e a análise de cluster utilizando algoritmo não-supervisionado, que permitem a emergência de padrões espontâneos dos dados.

Escolhemos analisar o mercado secundário de carros devido à sua importância econômica, tanto pela ampliação do espectro de bens disponíveis para os consumidores, quanto pela aglomeração de diversos agentes econômicos, tais como concessionárias, seguradoras, locadoras, financeiras e fabricantes de peças de reposição (BETTS; TARAN, 2006).

As contribuições do Capítulo 2 incluem um debate sobre o paradigma do *Big Data* em pesquisa econômica e a apresentação dos detalhes metodológicos da técnica de *Web Scraping*. No Brasil, 75% dos anúncios de carros na internet são de veículos usados. Este resultado encontra correspondência na literatura, sendo que nos EUA as negociações de carros usados representam entre 66% (STOLYAROV, 2002) e 75% (GAVAZZA *et al.*, 2014) de todas as transações de automóveis.

A segmentação do mercado de carros usados, realizada no Capítulo 3, revelou segmentos de mercado que merecem atenção dos agentes atuantes nesse mercado, tais como os

segmentos de carros encalhados (cluster 5) e sucatas (cluster 2). Considerando os critérios de quilometragem e preços, os segmentos de seminovos de luxo (cluster 1), carros luxuosos de uso diário (cluster 3) e carros de final de semana (cluster 6) possuem potencial de canibalizar o mercado de carros zero quilometro (Tabela 3.4).

De forma feral, as segmentações baseadas em design e tamanho, normalmente utilizadas no mercado de carros zero quilômetro, não capturam adequadamente (Tabela 3.2) os segmentos de mercado de usados, o qual está mais relacionado por variáveis como quilometragem, idade, potência e tempo de anúncio.

Contribuímos para a literatura no Capítulo 4 verificando se os efeitos da comunalidade sobrevivem no mercado secundário. Para isto, percorremos as literaturas de canibalização, comunalidade, preços hedônicos e modelos hierárquicos e especificamos um “modelo hierárquico hedônico”, que captura os efeitos da comunalidade utilizando a marca dos veículos como efeito aleatório. O coeficiente de correlação intraclasse do modelo de componentes aleatórios ($MHH0$) foi de 20,9% (Tabela 4.3), indicando que a estimação de um modelo hierárquico é mais adequado que um modelo OLS. Atributos de fábrica, como potência dos carros e origem da montadora, explicam melhor os efeitos aleatórios do que os atributos adquiridos no mercado secundário, como a quilometragem, o que corrobora a percepção de que os efeitos da comunalidade sobrevivem no mercado secundário. As estimativas do modelo hierárquico hedônico com dois níveis (marca e modelo) comprovam que os efeitos aleatórios da marca capturam a percepção dos consumidores a respeito da comunalidade dos carros

Os coeficientes de correlação intraclasse, para o modelo de componentes aleatórios com dois níveis, foram 14,1% para marca e 49% para o modelo (Tabela C.1 - Apêndice C.1). Também nos modelos hierárquicos hedônicos com dois níveis os atributos de fábrica explicam melhor os efeitos aleatórios que os atributos de potência.

Contribuímos para a literatura no Capítulo 5 realizando uma modelagem da depreciação de valor utilizando uma classe de modelos de crescimento exponencial. Os parâmetros estimados dos modelos de crescimento exponencial permitem calcular o ponto de inflexão,

que é o momento do tempo onde ocorre a desaceleração da curva de depreciação de valor dos carros.

Segundo modelo de Richards, O veículo Audi-A6 possui ponto de inflexão ocorrendo no primeiro ano, momento em que experimenta a depreciação de aproximadamente a metade do seu valor (Tabela 5.6). Os valores preditos (Gráficos 5.7 e 5.8) mostram a proximidade dos ajustes dos modelos Logístico e de Richards, especialmente para o Toyota-Corolla, Volkswagen-Passat e Audi-A6 (Perua).

De forma geral, o ponto de inflexão e a estimativa de depreciação variam bastante em função do tipo de veículo. Veículos mais caros tendem experimentar menores depreciações antes do ponto de inflexão em relação aos veículos mais baratos. Por exemplo, o veículo Mercedes Benz SLK possui depreciação no ponto de inflexão de 32,7%, enquanto o Fiat-Palio possui depreciação estimada no ponto de inflexão de 38,1% (Tabela 5.5).

A evolução da curva de depreciação antes do ponto de inflexão sugere que consumidores de veículos zero quilômetro irão absorver grande parte da perda de preço no momento da revenda. Consumidores de veículos usados, com idade maior que o ponto de inflexão, não experimentarão grandes perdas de investimento no momento da revenda.

Referências Bibliográficas

ABEELE, P. V.; GIJSBRECHTS, E.; VANHUELE, M. Specification and empirical evaluation of a cluster-asymmetry market share model. **Journal of Political Economy**, v. 7, n. 4, p. 223–247, 1990. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0167811690900014>>. Acesso em: 03 fev. 2016.

ABRAHAM, R.; ZIKIYE, M. W.; HARRINGTON, C. A theoretical model of competitive equilibria in the new car market. **Theoretical Economics Letters**, v. 5, n. 2, p. 196–211, apr 2015. Disponível em: <<http://www.scirp.org/journal/PaperInformation.aspx?PaperID=55175>>. Acesso em: 14 nov. 2015.

ADDA, J.; COOPER, R. Balladurette and juppette: A discrete analysis of scrapping subsidies. **Journal of Political Economy**, v. 108, n. 4, p. 223–247, ago. 2000. Disponível em: <<http://www.jstor.org/stable/10.1086/316096>>. Acesso em: 02 dez. 2015.

AKERLOF, G. The market for “lemons” quality uncertainty and the market mechanism. **Quarterly Journal of Economics**, v. 84, n. 3, p. 488–500, fev. 1970. Disponível em: <qje.oxfordjournals.org/content/84/3/488.abstract>. Acesso em: 18 dez. 2015.

ALBERINI, A.; HARRINGTON, W.; MCCONNELL, V. Fleet turnover and old car scrap policies. In: RESOURCES FOR THE FUTURE, Washington, DC. **Discussion Paper**. 1998. Disponível em: <<http://ageconsearch.umn.edu/handle/10897>>.

ANDRIKOPOULOS, A.; MARKELLOS, R. Dynamic interaction between markets for leasing and selling automobiles. **Journal of Banking and Finance**, v. 50, n. 6, p. 260–270, jan 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378426614000466>>. Acesso em: 26 jan. 2015.

ANGLIN, P. M.; GENCAI, R. Semiparametric estimation of a hedonic price function. **Journal of Applied Econometrics**, v. 11, n. 6, p. 633–648, nov 1996. Disponível em: <[http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1099-1255\(199611\)11:6%3C633::AID-JAE414%3E3.0.CO;2-T/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1255(199611)11:6%3C633::AID-JAE414%3E3.0.CO;2-T/abstract)>. Acesso em: 10 dez. 2015.

ANNADURAI, G.; BABU, S. R.; SRINIVASAMOORTHY, V. R. Development of mathematical models (logistic, gompertz and richards models) describing the growth pattern of *Pseudomonas Putida* (nicm 2174). **Bioprocess Engineering**, v. 23, n. 6, p. 607–612, dec 2000. Disponível em: <<http://link.springer.com/article/10.1007/s004490000209>>. Acesso em: 09 dez. 2015.

ARABIE, P.; CARROLL, J. D.; DESARBO, W.; WIND, J. Overlapping clustering: A new method for product positioning. **Journal of Marketing Research**, v. 18, n. 3, p. 310–317, aug 1981. Disponível em: <<http://www.jstor.org/stable/3150972>>. Acesso em: 16 nov. 2015.

ARGUEA, N. M.; HSIAO, C. Econometric issues of estimating hedonic price functions: With an application to the u.s. market for automobile. **Journal of Econometrics**, v. 56, n. 1, p. 243–267, mar. 1993. Disponível em: <<http://www.sciencedirect.com/science/article/pii/030440769390108H>>. Acesso em: 02 dez. 2015.

BAJARI, P.; FRUEHWIRTH, J. C.; KIM, K.; TIMMINS, C. A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution. **The American Economic Review**, v. 102, n. 5, p. 1898–1926, ago. 2012. Disponível em: <www.jstor.org/stable/41724609>. Acesso em: 03 fev. 2016.

BARNES, G.; LANGWORTHY, P. The per-mile costs of operating automobiles and trucks. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1864, n. 1, p. 71–77, ago. 2003. Disponível em: <<http://trrjournalonline.trb.org/doi/abs/10.3141/1864-10>>. Acesso em: 03 fev. 2016.

BARROS, D. C.; PEDRO, L. S. **As mudanças estruturais do setor automotivo, os impactos da crise e as perspectivas para o Brasil**. Rio de Janeiro, 2011. BNDES Setorial. Acesso em: 12 mar. 2015.

BELL, B.; ENE, M.; SMILEY, W.; SCHOENEBERGER, J. A multilevel model primer using sas proc mixed. In: SAS INSTITUTE, 2013, USA. **SAS Global Forum 2013**. San Francisco, CA: SAS Institute, 2013. p. 433–2013. Disponível em: <<http://support.sas.com/resources/papers/proceedings13/433-2013.pdf>>.

BERRY, S.; LEVINSOHN, J.; PAKES, A. Differentiated products demand systems from a combination of micro and macro data: The new car market. **Journal of Political Economy**, v. 112, n. 1, p. 68–103, nov 2004. Disponível em: <<http://www.journals.uchicago.edu/doi/10.1086/379939>>. Acesso em: 03 fev. 2016.

BETTS, S.; TARAN, Z. Prospect theory and perceptions of quality: Non-linear effects of quality comparisons on price in the used car market. **Academy of Marketing Studies Journal**, v. 10, n. 2, p. 1–57, dez. 2006. Disponível em: <<http://search.proquest.com/openview/b2ebf3648e3d083e9859043246bf13c2/1?pq-origsite=gscholar>>. Acesso em: 02 dez. 2015.

BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. 1rd. ed. Mumbai, India: O'Reilly, 2013.

BOLLIER, D. **The Promise and Peril of Big Data**. Washington, DC, 2010. Communications and Society Program. Disponível em: <<http://www.aspeninstitute.org/publications/promise-peril-big-data>>. Acesso em: 12 mar. 2015.

BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon: A decade in internet time: the dynamics of the internet and society. **Information, Communication and Society**, v. 15, n. 5, p. 662–679, maio 2012. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>>. Acesso em: 03 fev. 2016.

BRONNENBERG, B.; DUBE, J.-P.; GENTZKOW, M. The evolution of brand preferences: Evidence from consumer migration. **American Economic Review**, v. 102, n. 6, p. 2472–2508, out. 2012. Disponível em: <<http://www.aeaweb.org/articles.php?doi=10.1257/aer.102.6.2472>>. Acesso em: 03 fev. 2016.

BUSSE, M.; KNITTEL, C.; ZETTELMEYER, F. Are consumers myopic? evidence from new and used car purchases. **American Economic Review**, v. 103, n. 1, p. 220–256,

fev. 2012. Disponível em: <<http://www.aeaweb.org/articles.php?doi=10.1257/aer.103.1-.220>>. Acesso em: 30 nov. 2015.

BUSSE, M.; SILVA-RISSO, J. “one discriminatory rent” or “double jeopardy”: Multicomponent negotiation for new car purchases. **American Economic Review**, v. 100, n. 2, p. 470–474, may 2010. Disponível em: <<http://www.jstor.org/stable/27805041>>. Acesso em: 12 mar. 2015.

CASTRO, B. de; BARROS, D. C.; VAZ, L. F. H. **Além da Engenharia**: Panorama do capital nacional na indústria automotiva brasileira e insights para uma política pública rumo ao desenvolvimento de tecnologia automotiva no brasil. Rio de Janeiro, 2014. BNDES Setorial. Disponível em: <<http://web.bndes.gov.br/bib/jspui/handle/1408/3106>>. Acesso em: 12 mar. 2015.

CHAN, C. C. H. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. **Expert Systems with Applications**, v. 34, n. 4, p. 101–115, maio 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417407001686>>. Acesso em: 21 nov. 2015.

CHERON, E.; KLEINSCHMIDT, E. A review of industrial market segmentation research and a proposal for an integrated segmentation framework. **European Marketing Academy**, v. 2, n. 2, p. 101–115, jun. 1985. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0167811685900278x>>. Acesso em: 21 mar. 2015.

COHEN, A. J.; HARCOURT, G. C. Whatever happened to the cambridge capital theory controversies. **Journal of Economic Perspectives**, v. 17, n. 1, p. 199–214, 2003. Disponível em: <<http://ideas.repec.org/a/aea/jecper/v17y2003i1p199-214.html>>. Acesso em: 30 nov. 2015.

CRATTY, L. M.; SAHUTSKE, M. J. **Modeling and developing a commonality strategy in the automotive industry**. Dissertação (Mestrado) — Massachusetts Institute of Technology, System Design and Management Program, 77 Massachusetts Ave, Cambridge, MA 02139, jan. 2003. [Http://hdl.handle.net/1721.1/91795](http://hdl.handle.net/1721.1/91795).

CUKIER, K. **Data, Data Everywhere**. Washington, DC, 2010. Special report on managing information. Disponível em: <<http://www.economist.com/node/15557443>>. Acesso em: 12 mar. 2015.

DARGAY, J.; HANLY, M. Volatility of car ownership, commuting mode and time in the uk. **Transportation Research Part A: Policy and Practice**, v. 41, n. 10, p. 934–948, dez. 2007. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0965856407000407>>. Acesso em: 05 fev. 2016.

DATTA, Y. Market segmentation - an integrated framework. **Long Range Planning**, v. 29, n. 6, p. 797–811, dez. 1996. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0007681384900776>>. Acesso em: 12 mar. 2015.

DAVENPORT, T. H.; DYCHE, J. **Big Data in Big Companies**. Portland, Oregon, 2013. White Paper. Disponível em: <<http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf>>. Acesso em: 12 mar. 2015.

DAVIS, L. W.; KAHN, M. E. International trade in used vehicles: The environmental consequences of nafta. **American Economic Journal: Economic Policy**, v. 2, n. 4, p. 58–82, nov. 2010. Disponível em: <<http://www.aeaweb.org/articles.php?doi=10.1257/pol.2.4.58>>. Acesso em: 21 mar. 2015.

DEMIDENKO, E. **Mixed Models: Theory and Applications with R**. 2nd. ed. Hoboken New Jersey: John Wiley and Sons, 2013.

DESAI, P. Quality segmentation in spatial markets: When does cannibalization affect product line design? **Marketing Science**, v. 20, n. 3, p. 265–283, jun. 2001. Acesso em: 21 mar. 2015.

DINIS, P.; SILVA, F. M. da. The passenger car in portugal: Features of functional modules. **Procedia Manufacturing**, v. 3, n. 1, p. 6369–6375, jan 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2351978915003145>>. Acesso em: 24 dez. 2015.

EINAV, L.; LEVIN, J. D. The data revolution and economic analysis. **Innovation Policy and the Economy**, v. 14, n. 1, p. 1–40, maio 2014. Disponível em: <<http://web.stanford.edu/~jdlevin/Papers/BigData.pdf>>. Acesso em: 28 nov. 2015.

ENGERS, M.; HARTMANN, M.; STERN, S. Annual miles drive used car prices. **Journal of Applied Econometrics**, v. 24, n. 1, p. 1–33, jan. 2009. Disponível em: <<http://www.jstor.org/stable/40206260>>. Acesso em: 28 nov. 2015.

ERAT, S.; KAVADIAS, S. Introduction of new technologies to competing industrial customers. **Management Science**, v. 52, n. 11, p. 1675–1688, nov. 2006. Disponível em: <<http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1060.0561>>. Acesso em: 28 nov. 2015.

ESTEBAN, S.; SHUM, M. Durable-goods oligopoly with secondary markets: The case of automobiles. **The RAND Journal of Economics**, v. 38, n. 2, p. 332–354, jun. 2007. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1756-2171.2007.tb00071.x/abstract>>. Acesso em: 21 jan. 2016.

FEKEDULEGN, D. B.; SIURTAİN, M. M.; COLBERT, J. Parameter estimation of non-linear growth models in forestry. **Silva Fennica**, v. 33, n. 4, p. 327–336, maio 1999. Disponível em: <<http://www.silvafennica.fi/article/653>>. Acesso em: 28 dez. 2015.

FENG, Y.; FULLERTON, D.; GAN, L. Vehicle choices, miles driven and pollution policies. **Journal of Regulatory Economics**, v. 44, n. 1, p. 4–29, maio 2013. Disponível em: <<http://link.springer.com/article/10.1007%2Fs11149-013-9221-z>>. Acesso em: 28 nov. 2015.

FOX, J.; WEISBERG, S. Nonlinear regression and nonlinear least squares in r. In: MINNESOTA, S. W. U. of (Ed.). **An R Companion to Applied Regression**. [S.l.]: Urban Institute Press, 2011. v. 2, cap. Appendix, p. 472.

GARFINKEL, S.; COX, D. Finding and archiving the internet footprint. In: ARTS AND HUMANITIES RESEARCH COUNCIL, 1., 2009, UK. **First Digital Lives Research Conference: Personal Digital Archives for the 21st Century**. London, 2009. p. 9–11. Disponível em: <<http://www.simson.net/clips/academic/2009.BL.InternetFootprint.pdf>>.

GAVAZZA, A.; LIZZERI, A.; ROKETSKIY, N. A quantitative analysis of the used-car market. **American Economic Review**, v. 104, n. 11, p. 3668–3700, nov. 2014. Acesso em: 15 dez. 2015.

GOLDBERG, P. K.; VERBOVEN, F. The evolution of price dispersion in the european car market. **The Review of Economic Studies**, v. 68, n. 4, p. 811–848, out. 2001. Disponível em: <<http://www.jstor.org/stable/2695910>>. Acesso em: 01 jan. 2016.

GOLDSTEIN, H. **Multilevel Statistical Models**. 1rd. ed. West Sussex, United Kingdom: John Wiley and Sons, 2011.

GORDON, R. J. **The Measurement of Durable Goods Prices**. 1rd. ed. London, UK: The University of Chicago Press, 1990.

GOSHU, A. T.; KOYA, P. R. Derivation of inflection points of nonlinear regression curves: Implications to statistics. **American Journal of Theoretical and Applied Statistics**, v. 2, n. 6, p. 268–272, jan. 2013. Disponível em: <<http://article.sciencepublishinggroup.com/pdf/10.11648.j.ajtas.20130206.25.pdf>>. Acesso em: 12 abr. 2015.

GRIMM, K. Non-linear growth models in mplus and sas. **Structural Equation Modeling**, v. 16, n. 4, p. 676–701, out. 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3717396/>>. Acesso em: 12 abr. 2015.

GRUENSPECHT, H. K. Differentiated regulation: The case of auto emissions standards. **The American Economic Review**, v. 72, n. 2, p. 328–331, maio 1982. Disponível em: <<http://www.jstor.org/stable/1802352>>. Acesso em: 12 abr. 2015.

HACKELING, G. **Mastering Machine Learning with scikit-learn: Apply effective learning algorithms to real-world problems using scikit-learn**. 1rd. ed. Birmingham, UK: Packt Publishing Ltd, 2014.

HEERDE, H. J. van; SRINIVASAN, S.; DEKIMPE, M. G. Sibling rivalry: Estimating cannibalization rates for innovations. **GfK Marketing Intelligence Review**, v. 4, n. 2, p. 32–41, nov. 2012. Disponível em: <<http://www.degruyter.com/view/j/gfkmir.2012.4.issue-2/gfkmir-2014-0033/gfkmir-2014-0033.xml>>. Acesso em: 10 fev. 2016.

HEESE, H. S.; SWAMINATHAN, J. Product line design with component commonality and cost-reduction effort. **Manufacturing and Service Operations Management**, v. 8, n. 2, p. 206–219, abr. 2006. Disponível em: <<http://pubsonline.informs.org/doi/abs/10.1287/msom.1060.0103>>. Acesso em: 12 dez. 2015.

HILBERT, M.; LOPEZ, P. The world's technological capacity to store, communicate, and compute information. **Science**, v. 332, n. 60, p. 60–65, abr. 2011. Disponível em: <<http://www.uvm.edu/~pdodds/files/papers/others/2011/hilbert2011a.pdf>>. Acesso em: 12 mar. 2015.

HOX, J. **Multilevel Analysis Techniques and Applications**. 2nd. ed. New York, NY: Routledge, 2010.

HRUSCHKA, H. Market definition and segmentation using fuzzy clustering methods. **International Journal of Research in Marketing**, v. 3, n. 2, p. 117–134, dez. 1986. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0167811686900157>>. Acesso em: 22 nov. 2015.

HUANG, S.; YANG, Y.; ANDERSON, K. A theory of finitely durable goods monopoly with used goods market and transaction costs. **Management Science**, v. 47, n. 11, p. 1515–1532, nov 2001. Acesso em: 18 dez. 2015.

HULTEN, C. R.; WYKOFF, F. C. The measurement of economic depreciation. In: HULTEN, C. R. (Ed.). **Depreciation Inflation and The Taxation of Income from Capital**. [S.l.]: Urban Institute Press, 1981. v. 1, cap. 1, p. 119–152.

JANS, R.; DEGRAEVE, Z.; SCHEPENSC, L. Analysis of an industrial component commonality problem. **European Journal of Operational Research**, v. 186, n. 2, p. 801–811, abr. 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0377221707000914>>. Acesso em: 25 dez. 2015.

KAGAWAA, S.; HUBACEKB, K.; NANSAIC, K.; KATAOKAA, M.; MANAGID, S.; SUHE, S.; KUDOHF, Y. Better cars or older cars? assessing co2 emission reduction potential of passenger vehicle replacement programs. v. 23, n. 6, p. 1807–1818, dez. 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0959378013001295>>. Acesso em: 28 mar. 2015.

KAHM, M.; HASENBRINK, G.; FRATE, H. L.; LUDWIG, J.; KSCHISCHO, M. Grofit - fitting biological growth curves with r. **Journal of Statistical Software**, v. 33, n. 7, p. 1–27, fev. 2010. Disponível em: <<http://www.jstatsoft.org/article/view/v033i07/v33i07.pdf>>. Acesso em: 28 mar. 2015.

KAHN, J. A. Gasoline prices and the used automobile market a rational expectations asset price approach. **The Quarterly Journal of Economics**, v. 101, n. 2, p. 323–340, maio 1986. Disponível em: <<http://www.jstor.org/stable/1891118>>. Acesso em: 28 mar. 2015.

KIM, J.-C. The market for “lemons” reconsidered: A model of the used car market with asymmetric information. **The American Economic Review**, v. 75, n. 4, p. 836–843, set. 2000. Disponível em: <<http://www.jstor.org/stable/1821360>>. Acesso em: 01 jan. 2016.

KIM, K.; CHHAJEDB, D. Commonality in product design: Cost saving and cannibalization. **European Journal of Operational Research**, v. 125, n. 3, p. 602–621, set. 2000. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0377221799002714>>. Acesso em: 12 mar. 2015.

KITCHIN, R. Big data, new epistemologies and paradigm shifts. **Big Data and Society**, v. 1, n. 1, p. 1–12, jun. 2014. Disponível em: <<http://bds.sagepub.com/content/1/1/2053951714528481.full.pdf+html>>. Acesso em: 12 mar. 2015.

KLEINBERG, J.; PAPADIMITRIOU, C.; RAGHAVAN, P. A microeconomic view of data mining. **Data Mining and Knowledge Discovery**, v. 2, n. 4, p. 311–324, out. 1998. Disponível em: <<http://link.springer.com/article/10.1023%2FA%3A1009726428407>>. Acesso em: 22 jan. 2016.

LANTZ, B. **Machine Learning with R**. 2nd. ed. Birmingham, UK: Packt Publishing Ltd, 2015.

LAUGHLIN, J.; TAYLOR, C. An approach to industrial market segmentation. **Industrial Marketing Management**, v. 20, n. 2, p. 127–136, maio 1991. Disponível em: <<http://www.sciencedirect.com/science/article/pii/001985019190031A>>. Acesso em: 22 nov. 2015.

LEESON, P.; SOBEL, R. Costly price discrimination. **Economics Letters**, v. 99, n. 1, p. 206–208, abr. 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016517650700256X>>. Acesso em: 20 nov. 2015.

LINDSTROM, M.; BATES, D. Nonlinear mixed effects models for repeated measures data. **Biometrics**, v. 46, n. 3, p. 673–687, aug 1990. Disponível em: <<http://www.jstor.org/stable/2532087>>. Acesso em: 02 dez. 2015.

LUTZ, M. Pricing in segmented markets, arbitrage barriers, and the law of one price: Evidence from the european car market. **Review of International Economics**, v. 12,

n. 3, p. 456–475, aug 2004. Disponível em: <[http://onlinelibrary.wiley.com/doi/10.1111-j.1467-9396.2004.00461.x/abstract](http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9396.2004.00461.x/abstract)>. Acesso em: 01 jan. 2015.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. **Big Data: The next frontier for innovation, competition and productivity**. [S.l.], 2011. Report McKinsey Global Institute. Acesso em: 12 mar. 2015.

MAZUR, A.; STANIEC, I. Classification of car-buying clients: K-means and neural network approach. **International Advances in Economic Research**, v. 6, n. 4, p. 794–794, nov 2000. Disponível em: <<http://link.springer.com/article/10.1007%2F02295389>>. Acesso em: 01 fev. 2015.

MEIER, S.; SPRENGER, C. Present-biased preferences and credit card borrowing. **American Economic Journal: Applied Economics**, v. 2, n. 1, p. 193–210, jan. 2010. Disponível em: <<http://www.aeaweb.org/articles.php?doi=10.1257/app.2.1.193>>. Acesso em: 30 nov. 2015.

MOORE, G. Cramming more components onto integrated circuits. **Electronics**, v. 38, n. 8, p. 114–117, abr. 1965. Acesso em: 12 mar. 2015.

MOORTHY, K. S.; PNG, I. P. L. Market segmentation, cannibalization and the timing of product introductions of product introductions. **Management Science archive**, v. 38, n. 3, p. 345–359, mar. 1992. Disponível em: <<http://www.jstor.org/stable/2632479>>. Acesso em: 24 nov. 2015.

MURRAY, J.; SARANTIS, N. Quality- user cost- forward-looking behavior and the demand for cars in the uk. **Journal of Economics and Business**, v. 51, n. 3, p. 237–258, mar. 1999. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0148619599000041>>. Acesso em: 14 jan. 2016.

OHTA, M.; GRILICHES, Z. Automobile prices and quality did the gasoline price increases change consumer tastes in the us. **Journal of Business and Economic Statistics**, v. 4, n. 2, p. 187–198, abr. 1986. Disponível em: <<http://www.jstor.org/stable/1391318>>. Acesso em: 24 nov. 2015.

OPLAND, L. **Size Classification of Passenger Cars**: Pre-study on how to size classify passenger cars by inventorying the existing classification models. Goteborg, Swe-

den, 2007. Master's Thesis in the Master Degree Programme Civil Engineering. Disponível em: <<http://publications.lib.chalmers.se/records/fulltext/44868.pdf>>. Acesso em: 25 nov. 2015.

PARRY, I.; WALLS, M.; HARRINGTON, W. Automobile externalities and policies. **Journal of Economic Literature**, v. 45, n. 2, p. 373–399, jun. 2007. Disponível em: <<http://www.aeaweb.org/articles.php?doi=10.1257/jel.45.2.373>>. Acesso em: 24 nov. 2015.

PLANK, R. E. A critical review of industrial market segmentation. **Industrial Marketing Management**, v. 14, n. 2, p. 79–91, mar. 1985. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0019850185900458>>. Acesso em: 24 nov. 2015.

PRIETO, M.; CAEMMERER, B.; BALTAS, G. Using a hedonic price model to test prospect theory assertions the asymmetrical and nonlinear effect of reliability on used car prices. **Journal of Retailing and Consumer Services**, v. 14, n. 2, p. 206–212, jan. 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0969698914001210>>. Acesso em: 26 jan. 2015.

QI, L.; SAWHILL, J. How durable should durable products be made under different scenarios of technological advance? **International Journal of Production Economics**, v. 156, n. 2, p. 75–82, out. 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925527314001777>>. Acesso em: 15 jan. 2015.

REIS, H. J.; SILVA, J. S. Hedonic prices indexes for new passenger cars in portugal 1997-2001. **Economic Modelling**, v. 23, n. 6, p. 890–908, dez. 2006. Disponível em: <<http://www.sciencedirect.com/science/article/ii/S0264999306000447>>. Acesso em: 12 mar. 2015.

RODRIGUES, J. M. **Evolução da Frota de Automóveis e Motos no Brasil: 2001-2012**. Rio de Janeiro, 2013. Observatório das Metrôpoles. Acesso em: <http://www.observatoriodasmetrolopes.net/download>.

ROSEN, S. Hedonic prices and implicit markets product differentiation in pure competition. **Journal of Political Economy**, v. 82, n. 1, p. 34–55, jan. 1974. Disponível em: <<http://www.jstor.org/stable/1830899>>. Acesso em: 24 dez. 2015.

RUEBECK, C. Model exit in a vertically differentiated market: Interfirm competition versus intrafirm cannibalization in the computer hard disk drive industry. **Review of Industrial Organization**, v. 26, n. 1, p. 27–59, fev. 2005. Disponível em: <<http://link.springer.com/article/10.1007%2Fs11151-004-6576-4>>. Acesso em: 07 fev. 2016.

SALLEE, J. M. The taxation of fuel economy. **Tax Policy and the Economy**, v. 25, n. 1, p. 1–38, sep 2011. Disponível em: <<http://www.jstor.org/stable/10.1086/658379>>. Acesso em: 19 dez. 2015.

SALLEE, J. M. The taxation of fuel economy. **NBER Working Paper**, v. 21441, n. 1, p. 1–37, jul. 2015. Disponível em: <<http://www.nber.org/papers/w21441>>. Acesso em: 19 dez. 2015.

SALVADOR, F.; FORZA, C.; RUNGTUSANATHAMA, M. Modularity, product variety, production volume, and component sourcing: theorizing beyond generic prescriptions. **Journal of Operations Management**, v. 20, n. 5, p. 549–575, set. 2002. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S027269630200027X>>. Acesso em: 07 fev. 2016.

SAWHILL, J. W. **Three Essays on Strategic Considerations for Product Development**. Dissertação (Mestrado) — University of California, Berkeley, 7 2010. [Http://escholarship.org/uc/item/44f1m78x](http://escholarship.org/uc/item/44f1m78x).

SCHIRALDI, P. Automobile replacement: A dynamic structural approach. **The RAND Journal of Economics**, v. 42, n. 2, p. 266–291, maio 2011. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1756-2171.2011.00133.x/abstract>>. Acesso em: 03 dez. 2015.

SHAPIRO, B.; BONOMA, T. V. How to segment industrial markets. **Harvard Business Review**, v. 62, n. 1, p. 103–110, mai 1984. Disponível em: <<http://hbr.org/1984/05/how-to-segment-industrial-markets>>. Acesso em: 19 nov. 2015.

SILVER, M.; HERAVI, S. The measurement of quality-adjusted price changes. **National Bureau of Economic Research**, v. 110, n. 26, p. 277–316, jan. 2003. Disponível em: <<http://www.nber.org/chapters/c9739.pdf>>. Acesso em: 03 dez. 2015.

- SIMKIN, S. D. L. Implementation problems in industrial market segmentation. **Industrial Marketing Management**, v. 23, n. 1, p. 55–63, 1994. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0019850194900272>>. Acesso em: 21 mar. 2015.
- SNIJDERS, T.; BOSKER, R. **Multilevel Analysis An Introduction to Basic and Advanced Multilevel Modeling**. 2nd. ed. London, United Kingdom: SAGE, 2012.
- SOUZA, G. da Silva e. **Introdução aos Modelos de Regressão Linear e não-Linear**. 1rd. ed. Brasília, DF: Embrapa, 1998.
- SRINIVASAN, S. R.; RAMAKRISHNAN, S.; GRASMAN, S. Identifying the effects of cannibalization on the product portfolio. **Marketing Intelligence and Planning**, v. 23, n. 4, p. 359–371, fev. 2005. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/02634500510603465>>. Acesso em: 10 fev. 2016.
- STOLYAROV, D. Turnover of used durables in a stationary equilibrium: Are older goods traded more? **Journal of Political Economy**, v. 110, n. 6, p. 1390–1413, dez. 2002. Disponível em: <<http://www.jstor.org/stable/10.1086/343745>>. Acesso em: 01 nov. 2015.
- SUDARSHAN, D.; WINTER, F. Strategic segmentation of industrial markets. **Journal of Business and Industrial Marketing**, v. 13, n. 1, p. 8–21, 1998. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/08858629810206160>>. Acesso em: 19 nov. 2015.
- SULLIVAN, M. How brand names affect the demand for twin automobiles. **Journal of Marketing Research**, v. 35, n. 2, p. 154–165, may 1998. Disponível em: <<http://www.jstor.org/stable/3151844>>. Acesso em: 24 nov. 2015.
- VARIAN, H. R. Big data: New tricks for econometrics. **Journal of Economic Perspectives**, v. 28, n. 2, p. 3–28, 2014. Disponível em: <<http://www.aeaweb.org/articles.php?doi=10.1257/jep.28.2.3>>. Acesso em: 12 mar. 2015.
- VERBOVEN, F. Product line rivalry and market segmentation with an application to automobile optional engine pricing. **The Journal of Industrial Economics**, v. 47, n. 4, p. 399–425, dec 1999. Disponível em: <<http://www.jstor.org/stable/117539>>. Acesso em: 24 nov. 2015.

VERBOVEN, F. Quality based price discrimination and tax incidence the market for gasoline and diesel cars in europe. **RAND Journal of Economics**, v. 33, n. 2, p. 275–297, dec 2002. Acesso em: 30 nov. 2015.

WALDMAN, M. Durable goods pricing when quality matters. **The Journal of Business**, v. 69, n. 4, p. 489–510, out. 1996. Disponível em: <<http://www.jstor.org/stable/2353406>>. Acesso em: 12 mar. 2015.

WANG, C. H. Outlier identification and market segmentation using kernel based clustering techniques. **Expert Systems with Applications**, v. 36, n. 2, p. 3744–3750, mar. 1996. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417408001498>>. Acesso em: 19 nov. 2015.

WAZED, M. A.; AHMED, S.; NUKMAN, Y. Commonality in manufacturing resources planning - issues and models: a review. **European Journal of Industrial Engineering**, v. 4, n. 2, p. 167–188, jan. 2010. Disponível em: <<http://www.inderscience.com/info/inarticle.php?artid=31076>>. Acesso em: 08 fev. 2016.

WEE, B. V.; MOLL, H.; DIRKS, J. Environmental impact of scrapping old cars. **Transportation Research Part D Transport and Environment**, v. 5, n. 2, p. 137–143, mar. 2000. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1361920999000309>>. Acesso em: 17 dez. 2015.

WOLTMAN, H.; FELDSTAIN, A.; MACKAY, C.; ROCCHI, M. An introduction to hierarchical linear modeling. **Tutorials in Quantitative Methods for Psychology**, v. 8, n. 1, p. 52–69, jul. 2012. Disponível em: <<http://www.tqmp.org/Content/vol08-1/p052/p052.pdf>>. Acesso em: 21 jan. 2016.

WYKOFF, F. Obsolescence vs. deterioration with embodied technological change. In: INFLATION PRODUCTIVITY EMPLOYMENT AND EARNINGS RESEARCH, 1., 2004, Canada. **SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity**. Vancouver, British Columbia, 2004. p. 1–35.

YANKELOVICH, D.; MEER, D. Rediscovering market segmentation. **Harvard Business Review**, v. 84, n. 2, p. 122–131, out. 2006. Disponível em: <<http://hbr.org/2006/02/rediscovering-market-segmentation>>. Acesso em: 18 nov. 2015.

ZHAO, H.; JAGPALB, S. The effect of secondhand markets on the firms dynamic pricing and new product introduction strategies. **International Journal of Research in Marketing**, v. 23, n. 3, p. 295–307, sep 2006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167811606000486>>. Acesso em: 19 nov. 2015.

Apêndice A - Capítulo 2

A.1 Análise Descritiva da Base de Dados

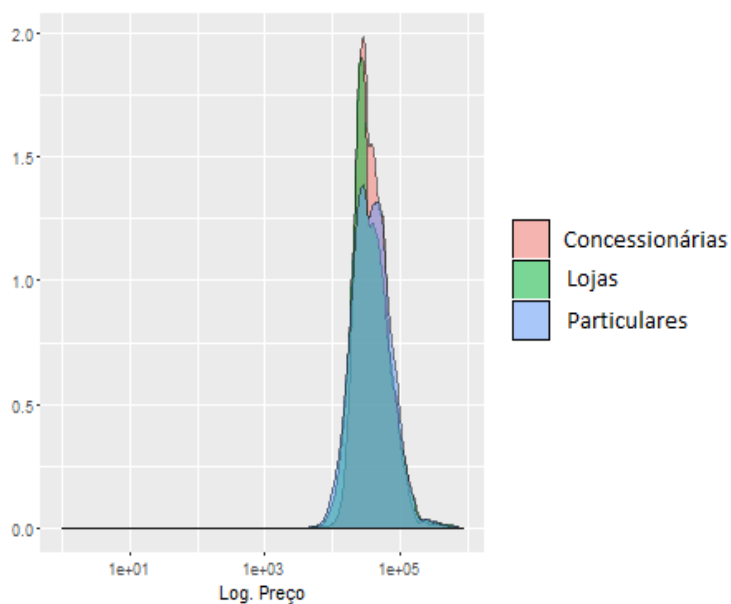


FIGURA A.1 – Histograma do Logaritmo dos Preços por Tipo de Vendedor

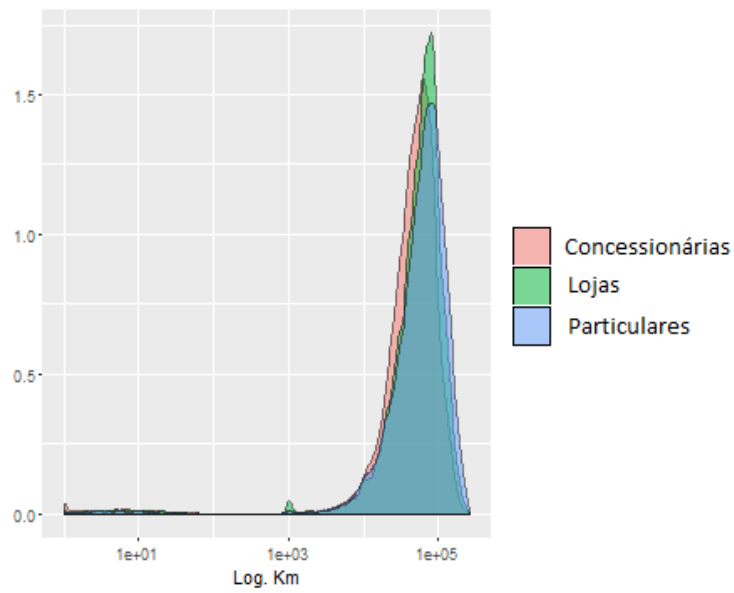


FIGURA A.2 – Histograma dos Quilômetros Rodados por Tipo de Vendedor

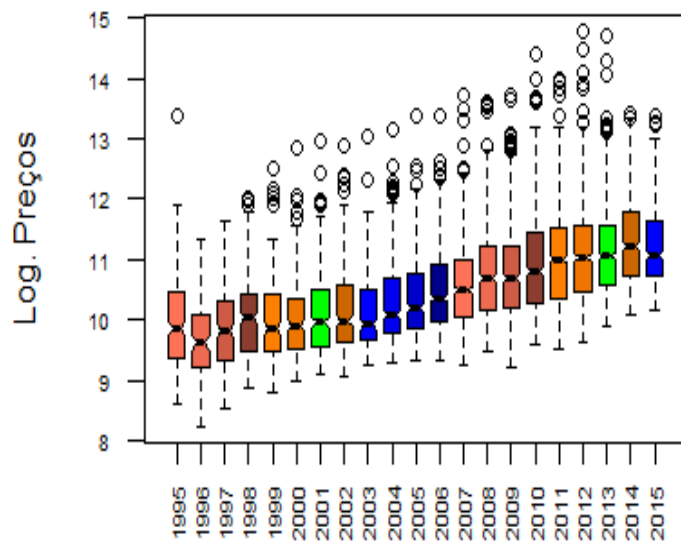


FIGURA A.3 – Box-Plot dos Preços dos Veículos Usados por Ano

TABELA A.1 – Quantidade de Carros, Preços Médios, Quilômetros Rodados e Coeficiente de Variação das Principais Marcas

Marca	Frequência (%)	Preço (R\$)		Quilometragem	
		Média	Coef. Variação	Média	Coef. Variação
CHEVROLET	16.834 (16%)	34.802	56,2	74.324	101,6
VOLKSWAGEN	15.075 (14%)	35.418	62,7	74.195	118,4
FIAT	13.580 (13%)	28.311	37,9	67.109	104,1
FORD	10.591 (10%)	37.194	60,1	71.604	100,7
PEUGEOT	7.648 (7%)	28.732	43,9	66.611	84,1
HONDA	6.297 (6%)	44.179	39,3	77.230	68,7
TOYOTA	4.674 (4%)	67.164	51,5	89.680	79,1
RENAULT	4.584 (4%)	28.193	40,1	64.343	94,2
MITSUBISHI	3.985 (4%)	63.394	36,4	79.900	88,1
CITROEN	3.008 (3%)	30.567	35,6	72.260	105,6
HYUNDAI	2.796 (3%)	46.979	24,9	62.321	95,2
MERCEDES-BENZ	2.674 (3%)	99.161	66,3	60.103	94,6
LAND ROVER	2.356 (2%)	139.582	47,6	60.499	79,8
BMW	2.115 (2%)	100.192	58,9	55.312	85,5
AUDI	1.838 (2%)	86.002	61,2	59.049	128,3
KIA	1.694 (2%)	54.261	38,5	64.670	76,0
NISSAN	1.127 (1%)	49.100	50,6	73.373	81,2
PORSCHE	719 (1%)	284.340	49,0	30.128	73,3
VOLVO	679 (1%)	80.114	34,7	58.785	54,0
DODGE	590 (1%)	82.587	51,1	69.491	111,3
DEMAIS MARCAS	4021 (4%)	90.927	169,8	58.694	108,9
TOTAL	106.885	47.711	108,1	70.345	99,9

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

TABELA A.2 – Relação de variáveis disponíveis na Base de Dados

Modelo	Modelo dos veículos	Conteúdo do Campo
Codigo	Identificador do anúncio	-
Zero	Identificador de carros zero	-
Marca	Denominação do fabricante	Ex: Gol, Tucson, Soul...
Coleta	Data de postagem na internet	Ex: 01/05/2015
Anúncio	Tempo de publicação do anúncio	Ex: 10 Dias
Preco	Preços dos veículos	Ex: R\$ 55.000,00
Km	Quilômetros rodados	Ex: 35.000
Categoria	Categoria de Carro	Ex: Sedã, Cupê, Minivan...
Cidade	Cidade de anúncio do carro	Ex: Goiânia, Santos, Vitória...
TipoA	Categoria do anunciante	Loja, Concessionária, Particular
UF	Unidade da federação	Ex: GO, SP, ES...
Ano	Ano do carro	Ex: 1978,...,2015
Idade	2016 - Ano do carro	-
Potência	Potência do motor	1.0, 1.4, 1.6, 2.0
Valvulas	Quantidade de válvulas do motor	2, 4, 8, 16, Mais que 16
Combustivel	Tipo de combustível	Ex: Gasolina, Álcool, Flex
Cor	Cor da pintura do veículo	Ex: Branco, Prata, Vermelho
ABS	Possui sistema de freios ABS	0, 1
AirBag	Possui <i>airbag</i>	0, 1
Alarme	Possui alarme	0, 1
Arcondicionado	Possui ar condicionado	0, 1
Arquente	Possui sistema ar quente	0, 1
RegulaBancos	Possui regulagem de bancos	0, 1
ConroleVelo	Possui controle de Velocidade	0, 1
BancoCouro	Possui banco de couro	0, 1
Desembacador	Possui desembacador traseiro	0, 1
Direcao hid	Possui direção hidráulica	0, 1
Estacionamento	Possui sensor de estacionamento	0, 1
CDMP3	Possui som com CD ou MP3	0, 1
Computadorbordo	Possui Computador de Bordo	0, 1
RodasLiga	Possui rodas de liga leve	0, 1
Retrovisor	Possui regulagem de retrovisor	0, 1
Tração	Possui tração nas quatro rodas	0, 1
Trava	Possui trava elétrica	0, 1
Vidro	Possui vidro elétrico	0, 1
Volante	Possui regulagem de volante	0, 1
Porta	Número de portas	Ex: 2, 4, 5

TABELA A.3 – Principais Modelos de Cada Montadora na Base de Dados

Marca	Modelo	Freq.	Preço (R\$)		Quilometragem		Idade Média
			Média	Coef. Var.	Média	Coef. Var.	
Audi	A3	616	54.172,7	58,5	83.031	127,8	7,2
Bmw	320i	382	84.138,3	34,4	49.999	114,5	4,4
Chevrolet	Cruze	1829	58.267,9	11,5	40.475	67,3	2,3
Citroen	C3	1248	27.023,6	30,2	65.914	118,8	4,9
DODGE	Journey	365	68.418,4	32,2	62.058	143,4	4,1
Fiat	Palio	3192	23.217,0	36,5	79.836	100,2	6,2
Ford	Fiesta	2853	25.221,1	34,1	64.747	109,8	5,1
Honda	Civic	2664	41.192,6	34,6	83.485	70,2	6,5
Hyundai	Tucson	649	40.817,0	18,4	72.444	38,9	5,7
Kia	Sportage	484	61.516,0	37,5	63.818	82,8	4,6
Land Rover	Freelander	416	84.746,5	39,3	71.754	78,9	5,3
Mercedes-benz	C 180	411	83.923,8	23,2	39.937	68,8	3,5
Mitsubishi	Pajero	678	47.279,1	23,8	64.253	60,6	4,7
Nissan	Frontier	392	69.530,8	26,2	94.932	87,7	5,1
Peugeot	207	3007	23.853,7	12,9	61.968	81,1	4,1
Porsche	Cayenne	256	218.084,3	45,9	45.611	53,1	5,2
Renault	Sandero	1666	28.161,9	19,2	50.426	107,8	3,0
Toyota	Corolla	1497	41.047,0	39,3	91.050	69,9	6,9
Volkswagen	Gol	3361	20.836,7	35,0	83.360	119,8	7,3
Volvo	XC60	385	90.935,8	19,6	58.764	45,5	3,9

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

Box-Plot dos Preços dos Veículos Usados por Potência e Tipo de

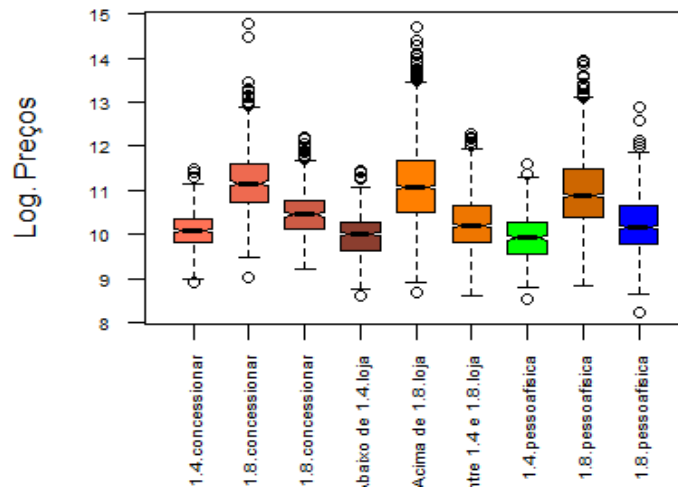


FIGURA A.4 – Box-Plot do Preços por Potência e Tipo de Vendedor

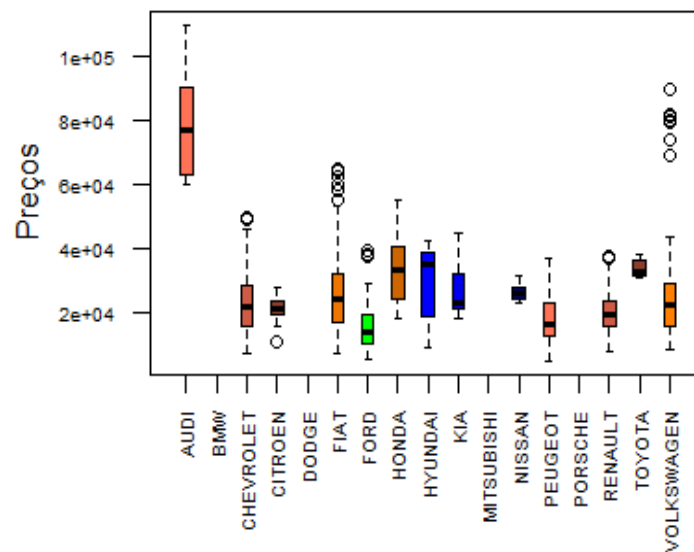


FIGURA A.5 – Box-Plot do Preço por Marca e Potência Abaixo de 1.4

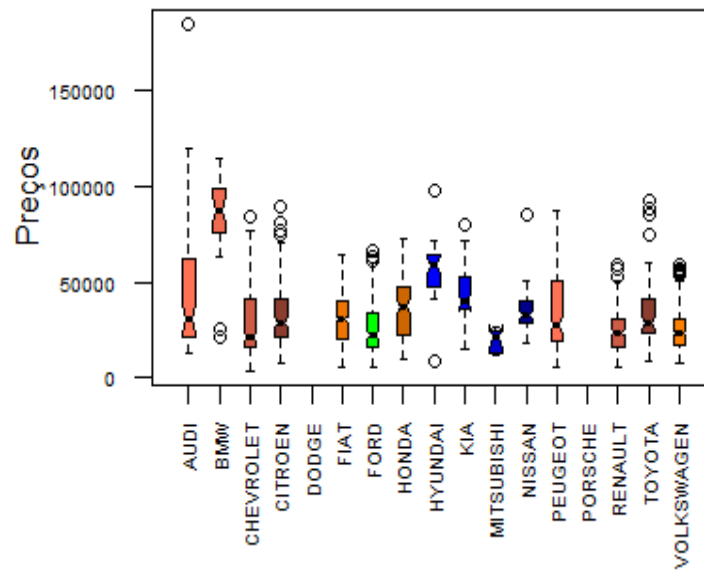


FIGURA A.6 – Box-Plot do Preço por Marca e Potência Entre 1.4 e 1.6

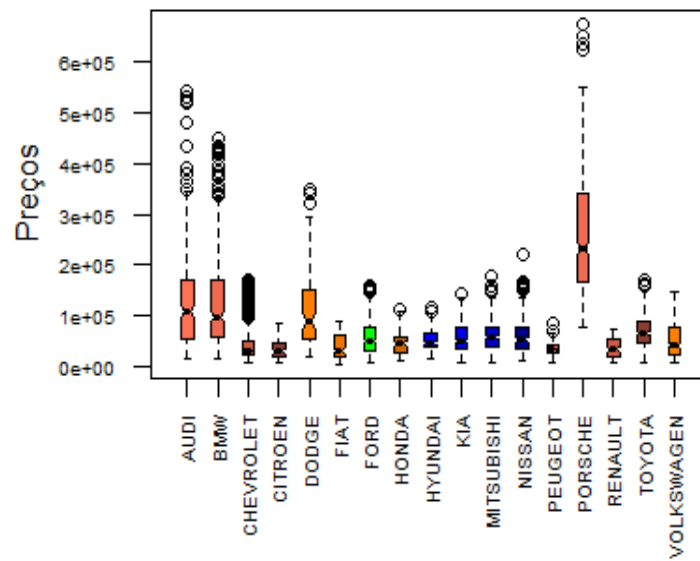


FIGURA A.7 – Box-Plot do Preço por Marca e Potência Acima de 2.0

Apêndice B - Capítulo 3

B.1 Tabelas da Análise de Cluster

TABELA B.1 – Características dos Itens de Conforto e Segurança nos Clusters

Características	1	2	3	4	5	6	7
Air Bag	91%	34%	79%	76%	32%	60%	21%
Alarme	87%	66%	74%	85%	69%	56%	47%
Ar Condicionado	97%	72%	81%	95%	74%	66%	65%
CD e/ou MP3	79%	47%	67%	74%	57%	51%	31%
Computador de Bordo	71%	17%	71%	56%	23%	56%	10%
Controle de Velocidade	40%	8%	60%	30%	11%	33%	1%
Banco de Couro	59%	29%	77%	57%	20%	60%	5%
Desembaçador Traseiro	81%	60%	61%	76%	56%	32%	47%
Direção Hidráulica	96%	75%	78%	95%	75%	60%	68%
Sensor de Estacionamento	51%	7%	64%	35%	13%	48%	3%
Freio ABS	90%	28%	79%	73%	30%	62%	19%
Rodas de Liga Leve	84%	48%	74%	81%	39%	53%	19%
Regulagem de Banco	67%	33%	66%	61%	37%	48%	15%
Retrovisor Elétrico	85%	43%	74%	80%	33%	53%	18%
Tração nas Quatro Rodas	28%	6%	62%	22%	8%	40%	0%
Trava Elétrica	96%	80%	78%	95%	82%	59%	71%
Vidro Elétrico	97%	76%	78%	95%	78%	58%	69%
Regulagem de Volante	72%	33%	67%	65%	35%	47%	13%

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

TABELA B.2 – Correlações por Clusters e Tipo de Vendedor

Cluster	Vendedor	Freq.	Valor Total	Correlação com Preço			
				Km	Potência	Itens	Idade
1	Concessionárias	3.373	226.501.003	-8,1%	50,1%	17,3%	-11,4%
1	Lojas	18.595	1.274.273.137	-12,5%	50,5%	20,8%	-10,6%
1	Particulares	8.644	583.144.250	-20,3%	43,7%	18,3%	-19,3%
2	Concessionárias	654	15.399.830	16,7%	79,9%	39,3%	-14,2%
2	Lojas	9.107	214.777.114	19,8%	75,0%	37,1%	-19,6%
2	Particulares	4.169	105.852.429	8,3%	69,0%	29,7%	-13,0%
3	Concessionárias	179	48.640.140	-41,0%	0,5%	-22,8%	-41,0%
3	Lojas	1.800	476.038.879	-45,2%	-21,4%	-9,0%	-15,5%
3	Particulares	642	149.263.596	-44,9%	-32,5%	-12,5%	-38,0%
4	Concessionárias	2.281	119.610.573	-26,6%	63,3%	27,3%	-36,1%
4	Lojas	15.876	806.618.154	-4,7%	62,3%	26,2%	-30,3%
4	Particulares	7.384	351.961.200	-8,7%	58,2%	21,9%	-28,7%
5	Concessionárias	64	1.950.920	-15,4%	69,7%	62,4%	-39,5%
5	Lojas	472	18.461.892	-14,2%	75,5%	57,0%	-33,2%
5	Particulares	16	445.290	-53,6%	61,7%	91,4%	-77,6%
6	Concessionárias	-	-	-	-	-	-
6	Lojas	66	73.798.90	-51,3%	-24,2%	-8,4%	-75,1%
6	Particulares	7	7.087.00	-2,2%	74,7%	87,1%	-8,5%
7	Concessionárias	6.285	203.774.910	-22,9%	64,5%	-0,1%	-34,7%
7	Lojas	21.608	641.008.541	-21,9%	63,3%	-9,1%	-30,2%
7	Particulares	4.001	119.978.610	-27,0%	63,9%	-16,0%	-32,7%
Total		105.223	5.438.586.368	-28,2%	57,9%	27,8%	-23,5%

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

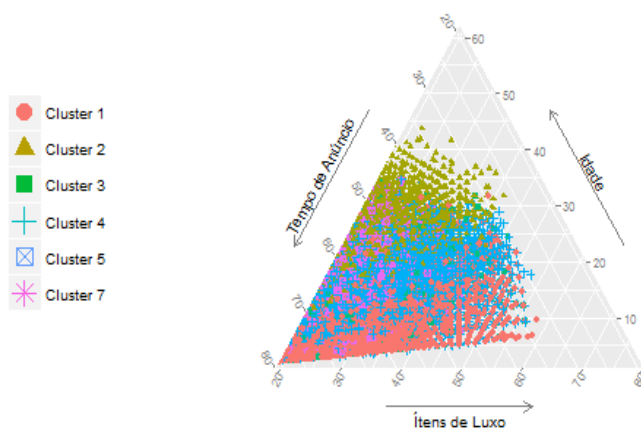


FIGURA B.1 – Ternary Plot do Tempo de Anúncio, Idade e Itens de Luxo por Cluster

Apêndice C - Capítulo 4

C.1 Resultados dos Modelos Hierárquicos Hedônicos

TABELA C.1 – Elasticidades dos Itens Opcionais para MHH5 com Dois Níveis

Explicativas	Parâmetro	Elast.	Explicativas	Parâmetro	Elast.
Número de Portas	0,007 9,7***	-	Controle Velocidade	0,045 13,7***	8,2%
Combustível: Álcool	0,174 9,9***	19,0%	Bancos de Couro	0,079 28,4***	8,2%
Combustível: Diesel	0,303 58,5***	35,4%	Desembaçador Traseiro	-0,048 -19,1***	-4,7%
Combustível: Flex	-0,173 -7,4***	-15,9%	Direção Hidráulica	0,008 2,0*	0,8%
Combustível: Gasolina	-	-	Sensor Estacionamento	0,046 16,1***	4,7%
Anúncio: Concessionária	0,056 15,4***	5,7%	Freios ABS	0,041 12,8***	4,3%
Anúncio: Loja	0,033 14,3***	3,4%	Rodas de Liga Leve	0,071 25,1***	7,4%
Anúncio: Particular	- -	-	Regulagem Banco	-0,019 -6,8***	-1,8%
<i>Air Bag</i>	0,013 4,03*	1,3%	Regulagem Retrovisor	0,032 10,7***	3,2%
Alarme	-0,002 -0,60 ^{ns}	-0,2%	Tração nas 4 Rodas	0,050 14,2***	5,1%
Ar Condicionado	0,033 8,8***	3,4%	Trava Elétrica	-0,042 -9,0***	-4,1%
CD-MP3	-0,00 -0,1 ^{ns}	-0,0%	Vidro Elétrico	0,001 0,1 ^{ns}	0,1%
Computador de Bordo	0,024 8,7***	2,5%	Regulagem Volante	-0,011 -4,0***	-1,1%

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “ns” Não possui significância estatística.

TABELA C.2 – Modelos de Preços Hedônicos Hierárquicos com Três Níveis

Explicativas	MHH0	MHH1	MHH2	MHH3	MHH4	MHH5
Intercepto	10,9	11,8	11,6	11,6	11,1	11,0
	215,1***	371,4***	398,8***	409,3***	317,0***	360,2***
Potência	-	0,177	0,175	0,139	0,671	0,216
	-	40,9***	41,6***	16,9***	40,5***	70,3***
ln(Km)	-	-0,123	-0,123	-0,119	-0,031	-0,033
	-	-107,6***	-110,4***	-112,6***	-35,4***	-36,6***
Tempo de Anúncio	-	-	-	-	0,0001	0,0002
	-	-	-	-	10,3***	9,9***
Idade do Carro	-	-	-	-	-0,079	-0,077
	-	-	-	-	-241,0***	-217,7***
Válvulas	-	-	-	-	0,005	0,004
	-	-	-	-	18,9***	15,4***
Europa	-	-	0,310	0,494	0,442	0,453
	-	-	3,6***	6,1 ^{ns}	8,8 ^{ns}	11,2***
Estima. de Covariância						
Intercepto (Marca)	0,065	0,019	0,006	0,006	0,018	0,011
	9,4***	13,8***	7,4***	9,9***	15,4***	17,9***
Intercepto (Modelo)	0,227	0,056	0,053	0,009	0,006	0,006
	20,9***	29,6***	30,9***	22,2***	21,0***	23,7***
Potência	-	-	-	0,014	0,062	0,044
	-	-	-	25,6***	20,4***	20,7***
Europa	-	-	0,077	0,078	-	-
	-	-	7,4 ^{ns}	8,5***	-	-
Resíduo	0,171	0,203	0,147	0,131	0,084	0,077
	209,9***	210,5***	209,4***	208,0***	209,5***	208,9***
ICC(Marca)	14,1%	8,1%	2,9%	2,4%	11,2%	8,6%
ICC(Modelo)	49,0%	24,5%	25,8%	3,6%	4,0%	4,2%
ICC(Potência)	-	-	-	55,3%	39,0%	33,3%
ICC(Europa)	-	-	37,7%	32,8%	-	-
AIC	66.550,8	52.316,6	50.244,2	44.587,0	-20.811,7	-24.061,5
BIC	66.555,7	52.321,5	50.250,8	44.595,2	-20.818,2	-24.054,9

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “*ns*” Não possui significância estatística.

TABELA C.3 – Elasticidades dos Itens Opcionais para MHH5 com Três Níveis

Explicativas	Parâmetro	Elast.	Explicativas	Parâmetro	Elast.
Número de Portas	0,012 5,1***	-	Controle de Velocidade	0,022 7,0***	2,2%
Combustível: Álcool	0,178 10,5***	19,5%	Bancos de Couro	0,028 10,1***	2,8%
Combustível: Diesel	0,296 43,3***	34,5%	Desembaçador Traseiro	-0,014 -5,6***	-1,4%
Combustível: Flex	-0,172 -7,9***	-15,8%	Direção Hidráulica	-0,005 -1,4 ^{ns}	-0,5%
Combustível: Gasolina	-	0,0%	Sensor de Estacionamento	0,032 11,7***	3,2%
Anúncio: Concessionária	0,041 11,9***	4,2%	Freios ABS	0,012 3,8***	1,2%
Anúncio: Loja	0,029 12,3***	2,9%	Rodas de Liga Leve	0,033 11,8***	3,3%
Anúncio: Particular	-	0,0%	Regulagem Banco	-0,009 -3,4***	-0,9%
<i>Air Bag</i>	0,000 0,04 ^{ns}	0,0%	Regulagem Retrovisor	0,004 1,4 ^{ns}	0,4%
Alarme	0,006 2,4*	0,6%	Tração nas 4 Rodas	0,023 7,0***	2,4%
Ar Condicionado	0,017 4,6***	1,7%	Trava Elétrica	-0,009 -2,1*	-0,9%
CD-MP3	-0,003 -1,2 ^{ns}	-0,3%	Vidro Elétrico	0,005 1,2 ^{ns}	0,5%
Computador de Bordo	0,018 7,0***	1,9%	Regulagem Volante	-0,017 -6,3***	-1,7%

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “ns” Não possui significância estatística.

Apêndice D - Capítulo 5

D.1 Derivadas Parciais dos Modelos de Crescimento Exponencial

Nas Tabelas [D.1](#), [D.2](#), [D.3](#) e [D.4](#), são apresentadas as derivadas parciais em relação a cada um dos parâmetros β_0 , β_1 , β_2 e β_3 dos modelos de crescimento exponencial. O procedimento *NLIN* do *SAS* requer a especificação das derivadas parciais em relação aos parâmetros a serem estimados, as quais devem ser inseridas na sintaxe de programação. O procedimento *MODEL* do *SAS* e a função *NLS* do **R** não necessitam da especificação das derivadas parciais (FOX; WEISBERG, 2011).

TABELA D.1 – Derivadas Parciais em relação à β_0

Caso	$\partial w(t)/\partial \beta_0$
Exponencial Negativa	$1 - \exp(-\beta_2 t)$
Monomolecular	$1 - \beta_1 \exp(-\beta_2 t)$
Mitcherlich	1
Gompertz	$\exp(-\beta_1 \exp(-\beta_2 t))$
Logística	$1/[1 + \beta_1 \exp(-\beta_2 t)]$
Von Bertalanffy	$[\beta_0^{(1-\beta_3)} - \beta_1 \exp(-\beta_2 t)] / \{\beta_0^{\beta_3}\}$
Richards	$(1 + \beta_1 \exp(-\beta_2 t))^{\frac{-1}{\beta_3}}$
Weibull	1

TABELA D.2 – Derivadas Parciais em relação à β_1

Função	$\partial w(t)/\partial \beta_1$
Exponencial Negativa	
Monomolecular	$-\beta_0 \exp(-\beta_2 t)$
Mitcherlich	$-\beta_2^t$
Gompertz	$-\beta_0 [\exp(-\beta_1 \exp(-\beta_2 t))] (\exp(-\beta_2 t))$
Logística	$\frac{-\beta_0 \exp(-\beta_2 t)}{[1 + \beta_1 \exp(-t \beta_2)]^2}$
Von Bertalanffy	$\frac{-\exp(-\beta_2 t)}{1 - \beta_3} (\beta_0^{(1-\beta_3)} - \beta_1 \exp(-\beta_2 t)) \frac{1}{1 - \beta_3}^{-1}$
Richards	$\frac{-\beta_0}{\beta_3} (1 + \beta_1 \exp(-\beta_2 t))^{\frac{-1}{\beta_3} - 1} (\exp(-\beta_2 t))$
Weibull	$-\exp(-\beta_2 t^{\beta_3})$

TABELA D.3 – Derivadas Parciais em relação à β_2

Função	$\partial w(t)/\partial \beta_2$
Exponencial Negativa	$-\beta_0 t \exp(-\beta_2 t)$
Monomolecular	$\beta_0 \beta_1 t \exp(-\beta_2 t)$
Mitcherlich	$-\beta_1 t \beta_2^{t-1}$
Gompertz	$\beta_0 \beta_1 t \exp(-\beta_1 \exp(-\beta_2 t)) (\exp(-\beta_2 t))$
Logística	$\frac{\beta_0 \beta_1 t}{[1 + \beta_1] [\exp(-\beta_2 t)]^2 (\exp -\beta_2 t)}$
Von Bertalanffy	$\frac{(\beta_1 t)}{(1 - \beta_3)} (\exp(-\beta_2 t)) [\beta_0^{1-\beta_3} - \beta_1 \exp(-\beta_2 t)]$
Richards	$\frac{\beta_0 \beta_1 t}{\beta_3} (1 + \beta_1 \exp(-\beta_2 t)) \frac{\beta_3}{1 - \beta_3} (\exp(-\beta_2 t))$
Weibull	$\exp(-\beta_2 t^{\beta_3}) \beta_1 t^{\beta_3}$

TABELA D.4 – Derivadas Parciais em relação à β_3

Função	$\partial w(t)/\partial \beta_3$
Exponencial Negativa	
Monomolecular	-
Mitcherlich	-
Gompertz	-
Logística	-
Von Bertalanffy	$\left[\exp\left(\frac{1}{1-\beta_3}\right) \ln(\eta)^{\frac{\beta_3}{1-\beta_3}} \right] / [1 - \beta_3] \times \left\{ \frac{\ln(\eta)}{1 - \beta_3} - \frac{\ln(\beta_0) [\beta_0^{(1-\beta_3)}]}{\eta} \right\}$
Richards	$\beta_0 (1 + \beta_1 \exp(-\beta_2 t))^{\frac{-1}{\beta_3}} \ln(1 + \beta_1 \exp(-\beta_2 t)) \beta_3^{-2}$
Weibull	$\exp(-\beta_2 t^{\beta_3}) \beta_1 \beta_2 \ln(t) t^{\beta_3}$

OBS: $\eta = \left[\beta_0^{(1-\beta_3)} - \beta_1 \exp(-\beta_2 t) \right]$

D.2 Derivação dos Pontos de Inflexão

Considerando notação do modelo de Weibull em [Goshu e Koya \(2013\)](#), o ponto de inflexão a é dado por:

$$\begin{aligned}
 w(t) &= \beta_0 \exp(-\beta_1 \exp(-\beta_2 t)) \\
 f(t) &= A \exp(-B \exp(-\kappa t)) \\
 B &= \ln \left[\frac{A}{A_0} \right] \\
 B &= \beta_1, \quad A = \beta_0, \quad \kappa = \beta_2 \\
 A &= \beta_0 \\
 \kappa &= \beta_2 \\
 a &= \frac{1}{\kappa} \ln \left[\ln \left[\frac{A}{A_0} \right] \right] = \frac{1}{\beta_2} \ln \beta_1
 \end{aligned}$$

Obtemos o ponto de inflexão derivando duas vezes as equações de crescimento e igualando a segunda derivada a zero. Considerando o modelo de crescimento logístico generalizado segundo notação de [Goshu e Koya \(2013\)](#) e derivando duas vezes em relação à t :

$$\begin{aligned}
f(t) &= A_L + (A - A_L)[1 - B \exp(-k[t - \mu])]^m \\
B &= \left[1 - \left(\frac{A_\mu - A_L}{A - A_L} \right)^{1/m} \right] \\
f'(t) &= mk[f(t) - A_L]^{(1-\frac{1}{m})} ([A - A_L]^{1/m} - [f(t) - A_L]^{1/m}) \\
f''(t) &= m k \left[\left(1 - \frac{1}{m} \right) \left[\frac{A - A_L}{f(t) - A_L} \right]^{1/m} - 1 \right] f'(t) \\
f''(t) &= 0 \quad \Rightarrow \quad 0 = \left(1 - \frac{1}{m} \right) \left[\frac{A - A_L}{f(t) - A_L} \right] - 1 \\
f(t) &= A_L + (A - A_L) \left[1 - \frac{1}{m} \right]^m \\
\frac{1}{m} &= B \exp[-k(t - \mu)] \\
\frac{1}{m B} &= \exp[-k(t - \mu)] \\
k(t - \mu) &= \ln \left[m \left(1 - \left[\frac{A_\mu - A_L}{A - A_L} \right]^{1/m} \right) \right] \\
a &= \frac{1}{k} \ln \left(m \left[1 - \frac{A_\mu - A_L}{A - A_L} \right]^{1/m} \right) + \mu
\end{aligned}$$

Fazendo as substituições: $f(t) = w(t)$, $A_L = 0$, $\mu = 0$, $A = \alpha$, $B = -\beta$ e $m = -1$, temos o caso particular do modelo de crescimento logístico apresentado em [Fekedulegn et al. \(1999\)](#).

Considerando o modelo de crescimento de Richards conforme notação de [Goshu e Koya](#)

(2013), e derivando duas vezes em relação a t :

$$\begin{aligned}
 f(t) &= A [1 - B \exp(-k t)]^n \\
 B &= \left[1 - \left(\frac{A_0}{A} \right)^{1/n} \right] \\
 f'(t) &= n k [f(t)]^{(1-\frac{1}{n})} (A^{1/n} - f(t)^{1/n}) \\
 f''(t) &= k f'(t) [f(t)]^{\frac{-1}{n}} ([n-1] A^{1/n} - n f^{1/n}) \\
 f''(t) &= 0 \quad \Rightarrow \quad [n-1] A^{1/n} - n f^{1/n} = 0 \\
 f(t) &= \left[\frac{n-1}{n} \right]^n A \\
 \left[\frac{n-1}{n} \right]^n A &= A [1 - B \exp(-k t)]^n \quad \Rightarrow \quad \frac{n-1}{n} = 1 - B \exp(-k t) \\
 \frac{1}{n} &= B \exp(-k t) \\
 k t &= \ln(B n) \\
 k t &= \ln(B n) \\
 a &= \frac{1}{k} \ln B n
 \end{aligned}$$

Fazendo as substituições: $f(t) = w(t)$, $A = \alpha$, $B = -\beta$ e $-n = \frac{1}{m}$, temos o caso particular do modelo de Richards, conforme a notação apresentada em [Fekedulegn et al. \(1999\)](#).

Considerando o modelo de crescimento exponencial Weibull, e derivando duas vezes em relação a t :

$$\begin{aligned}
w(t) &= \beta_0 - \beta_1 \exp(-\beta_2 t^{\beta_3}) \\
w'(t) &= -\beta_1 \exp(-\beta_2 t^{\beta_3})[-\beta_2 \beta_3 t^{\beta_3-1}] \\
&= (\beta_1 \beta_2 \beta_3) [(t^{\beta_3-1}) \exp(-\beta_2 t^{\beta_3})] \\
w''(t) &= \beta_1 \beta_2 \beta_3 (t^{\beta_3-1}) [\exp(-\beta_2 t^{\beta_3})] [-\beta_2 \beta_3 t^{\beta_3-1}] \\
&\quad + [(\beta_3 - 1)t^{\beta_3-2}] \exp(-\beta_2 t^{\beta_3}) \\
w'''(t) &= 0 \\
0 &= (t^{\beta_3-1}) [-\beta_2 \beta_3 t^{-1}] + [(\beta_3 - 1)t^{-2}] \\
\beta_3 - 1 &= \beta_2 \beta_3 t^{\beta_3} \\
a &= \left[\frac{\beta_3 - 1}{\beta_2 \beta_3} \right]^{\left(\frac{1}{\beta_3}\right)}
\end{aligned}$$

D.3 Estimativas dos Parâmetros dos Modelos

Os modelos Exponencial Negativa, Monomolecular e Mitcherlich apresentam valores altos para os coeficientes de explicação e parâmetro estimados significantes. O parâmetro β_0 mostra uma estimativa para a desvalorização máxima dos veículos.

Vemos que o modelo SLK-Mercedes-Benz possui desvalorização máxima de 77% segundo o modelo Exponencial Negativa. Segundo os modelos Monomolecular e Mitcherlich, a desvalorização máxima estimada do modelo SLK-Mercedes-Benz é de 73%.

O modelo Carnival-Kia possui desvalorização máxima estimada variando entre 74% e 81% dependendo do modelo. Considerando o modelo *Exponencial Negativa* ajustado para o veículo Discovery, a desvalorização máxima de um veículo atinge o valor de 100%, entretanto, segundo os modelos Monomolecular e Mitcherlich a desvalorização máxima é de 93% (Tabela D.5).

A Tabela D.6 mostra os parâmetros estimados e o ponto de inflexão para todos os veículos analisados, considerando o modelo de crescimento Gompertz. O parâmetro β_0 mostra que a desvalorização máxima dos veículos A6, Carnival, Corolla, Discovery, Hilux, Palio, Passat e SLK é, respectivamente, 87%, 71%, 99,1%, 86%, 87%, 78%, 79% e 88%.

Considerando o veículo SLK-Mercedes-Benz, o parâmetro estimado para a depreciação

TABELA D.5 – Modelos de Crescimento Não-Linear: Exponencial Negativa, Monomolecular e Mitcherlich

Carro		Exp. Neg.		Monomolecular			Mitcherlich		
		β_0	β_2	β_0	β_1	β_2	β_0	β_1	β_2
A6	Parâm.	0,97	0,14	0,92	1,18	0,19	0,92	1,08	0,83
Audi	Est. t	28,4*	14,9*	102*	40,1*	22,5*	102*	49,7*	121*
	R^2	98,0%		99,6%			99,6%		
Carnival	Parâm.	0,81	0,17	0,74	1,29	0,26	0,74	0,95	0,77
Kia	Est. t	14,9*	6,8*	42,0*	21,5*	10,9*	42*	26,6*	42,2*
	R^2	94,5%		97,9%			97,9%		
Corolla	Parâm.	0,96	0,11	0,93	1,04	0,13	0,93	0,97	0,88
Toyota	Est. t	64,9*	21,0*	68,5*	44,1*	17,3*	68*	56,7*	139*
	R^2	99,4%		99,6%			99,6%		
Discovery	Parâm.	1,01	0,13	0,93	1,22	0,19	0,93	1,13	0,83
Land Rover	Est. t	17,7*	10,1*	79,3*	79,1*	26,6*	79*	96,0*	141*
	R^2	97,2%		99,6%			99,6%		
Hilux	Parâm.	0,99	0,08	0,89	1,09	0,10	0,89	0,97	0,90
Toyota	Est. t	20,0*	11,2*	33,0*	47,5*	14,4*	33*	41,3*	143*
	R^2	98,5%		99,2%			99,2%		
Palio	Parâm.	0,83	0,12	0,87	0,94	0,10	0,87	0,82	0,90
Fiat	Est. t	46,0*	22,6*	58,2*	61,7*	18,6*	58*	68,2*	178*
	R^2	99,3%		99,6%			99,6%		
Passat	Parâm.	0,97	0,13	0,93	1,12	0,16	0,93	1,04	0,85
Volkswagen	Est. t	41,2*	19,5*	132*	53,5*	32,5*	133*	66,7*	203*
	R^2	99,0%		99,8%			99,8%		
SLK	Parâm.	0,77	0,11	0,73	1,08	0,14	0,73	0,78	0,87
Mercedes-Benz	Est. t	20,6*	11,8*	38,7*	36,6*	14,1*	38*	40,5*	102*
	R^2	98,7%		99,0%			97,2%		

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “ns” Não possui significância estatística.

máxima, segundo o modelo Gompertz, é de 0,67. O ponto de inflexão é de 3,8 anos, momento de tempo onde a depreciação é de 24,8% e a depreciação anual é de 6,6%. Vemos que o ponto de inflexão do veículo SLK é menor que o ponto de inflexão dos veículos Hilux-Toyota (5,1 anos) e Corolla-Toyota (4,0 anos), ou seja, o preço do SLK começa a se estabilizar mais rapidamente que os preços do Hilux e Corolla. O veículo Audi-A6 é o que apresenta a maior desvalorização anual antes do ponto de inflexão (9,2%).

O modelo de Von Bertalanffy produz pontos de inflexão menores que os pontos de inflexão dos demais modelos. Para os veículos Audi-A6, Toyota-Corolla e Volkswagen-Passat, os pontos de inflexão foram menores que 1 (Tabela D.7). Nota-se que para estes

TABELA D.6 – Modelos de Crescimento Não-Linear: Gompertz

Carro		β_0	β_1	β_2	a	$w(a)$	$w(a)/a$	R^2
A6	Parâm.	0,87	3,07	0,32	3,5	32,2%	9,2%	98,8%
Audi	Est. t	73,1*	18,2*	19,5*				
Carnival	Parâm.	0,71	4,26	0,49	3,0	26,0%	8,7%	99,1%
Kia	Est. t	119*	6,1*	10,9*				
Corolla	Parâm.	0,86	2,41	0,22	4,0	31,8%	8,0%	99,2%
Toyota	Est. t	67,5*	31,4*	23,4*				
Discovery	Parâm.	0,87	3,51	0,35	3,6	32,1%	8,9%	98,9%
Land Rover	Est. t	66,0*	14,7*	16,8*				
Hilux	Parâm.	0,78	3,07	0,22	5,1	28,6%	5,6%	99,7%
Toyota	Est. t	48,7*	39,7*	27,7*				
Palio	Parâm.	0,79	1,97	0,19	3,6	29,1%	8,1%	99,8%
Fiat	Est. t	125*	79,3*	44,3*				
Passat	Parâm.	0,88	2,74	0,27	3,7	32,3%	8,8%	99,2%
Volkswagen	Est. t	69,1*	29,4*	19,0*				
SLK	Parâm.	0,67	2,48	0,24	3,8	24,8%	6,6%	97,2%
Mercedes-Benz	Est. t	32,7*	14,3*	11,6*				

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “*ns*” Não possui significância estatística.

mesmos modelos de carros o parâmetro β_0 não possui significância estatística (Tabela D.7).

Para o modelo Mercedes-Benz-SLK, o parâmetro estimado para β_3 é negativo, desta forma: $\{\ln(\beta_1/[1 - \beta_3]) < 1\}$ e o ponto de inflexão possui valor negativo.

A expressão $\ln(\frac{\beta_1}{1-\beta_3})$ não possui valor definido quando $\beta_1 = 0$ (Tabela 5.3). Desta forma, notamos que todos os pontos de inflexão calculados para o modelo de Von Bertalanffy possuem limitações de análise. Isto pode ser devido ao problema de estimação de máximos locais no lugar de máximos globais.

Os Gráficos D.10 e D.11 mostram os valores observados e preditos para os veículos Audi-A6 (Perua), Kia -Carnival (Minivan), Toyota-Corolla (Sedã), Land Rover-Discovery (Utilitário), Toyota-Hilux (Picape), Fiat-Palio (Hatchback), Volkswagen-Passat (Cupê) e Mercedes-Benz-SLK (Conversível) para os modelos de crescimento não-linear: Negative Exponential, Monomolecular, Mitcherlich, Gompertz, Von Bertalanffy e Weibull. Vemos de forma geral que os modelos apresentam bom ajuste.

TABELA D.7 – Modelos de Crescimento Não-Linear: Von-Bertalanffy

Carro		β_0	β_1	β_2	β_3	a	$w(a)$	$w(a)/a$	R^2
A6	Parâm.	0,91	1,08	0,20	0,06	0,7	4,6%	0,07	99,6%
Audi	Est. t	73,8*	18,6*	8,45*	0,30 ^{ns}				
Carnival	Parâm.	0,71	-0,03	0,49	1,01	3,0	26,0%	0,09	99,0%
Kia	Est. t	105*	-0,01 ^{ns}	4,41*	1,97***				
Corolla	Parâm.	0,92	0,94	0,14	0,13	0,5	8,7%	0,17	99,5%
Toyota	Est. t	45,7*	11,8*	6,18*	0,49 ^{ns}				
Discovery	Parâm.	0,91	1,13	0,22	0,17	1,4	10,8%	0,08	99,7%
Land Rover	Est. t	88,8*	47,8*	14,4*	1,88***				
Hilux	Parâm.	0,78	0,27	0,21	0,90	4,8	27,3%	0,06	99,7%
Toyota	Est. t	38,9*	0,82 ^{ns}	9,59*	6,2*				
Palio	Parâm.	0,80	0,30	0,17	0,82	2,9	26,5%	0,09	99,8%
Fiat	Est. t	70,6*	1,36 ^{ns}	10,8*	4,92*				
Passat	Parâm.	0,92	1,02	0,17	0,08	0,6	6,0%	0,10	99,8%
Volkswagen	Est. t	90,0*	27,4*	12,8*	0,65 ^{ns}				
SLK	Parâm.	0,83	0,82	0,08	-0,48	-7,8	-	-	99,4%
Mercedes-Benz	Est. t	12,1*	11,98*	3,54*	-2,7**				

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “ns” Não possui significância estatística.

TABELA D.8 – Modelos de Crescimento Não-Linear: Weibull

Carro		β_0	β_1	β_2	β_3	a	$w(a)$	$w(a)/a$	R^2
A6	Parâm.	0,02	-1,04	4,50	-1,09	2,19	17,7%	0,08	99,54%
Audi	Est. t	0,9 ^{ns}	-18,3*	7,1*	-11,4*				
Carnival	Parâm.	0,05	-0,72	7,69	-1,71	2,52	20,2%	0,08	98,38%
Kia	Est. t	1,4 ^{ns}	-12,4*	2,6*	-6,7*				
Corolla	Parâm.	0,07	-1,21	4,35	-0,78	2,29	19,1%	0,08	99,34%
Toyota	Est. t	1,8***	-7,9*	4,9*	-6,0*				
Discovery	Parâm.	0,02	-1,03	5,27	-1,21	2,41	18,1%	0,08	99,84%
Land Rover	Est. t	0,1 ^{ns}	-4,6*	1,8***	-5,7*				
Hilux	Parâm.	0,06	-1,04	7,05	-0,97	3,61	19,5%	0,05	99,17%
Toyota	Est. t	3,5*	-10,4*	5,6*	-8,1*				
Palio	Parâm.	0,14	-1,02	5,25	-0,79	2,89	25,1%	0,09	99,48%
Fiat	Est. t	9,6*	-8,9*	6,9*	-7,4*				
Passat	Parâm.	0,04	-1,09	4,38	-0,96	2,22	18,5%	0,08	99,70%
Volkswagen	Est. t	2,2**	-20,6*	9,3*	-13,8*				
SLK	Parâm.	0,00	-1,14	3,32	-0,63	1,49	8,6%	0,06	99,45%
Mercedes-Benz	Est. t	0,0 ^{ns}	-10,0*	21,6*	-8,8**				

FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

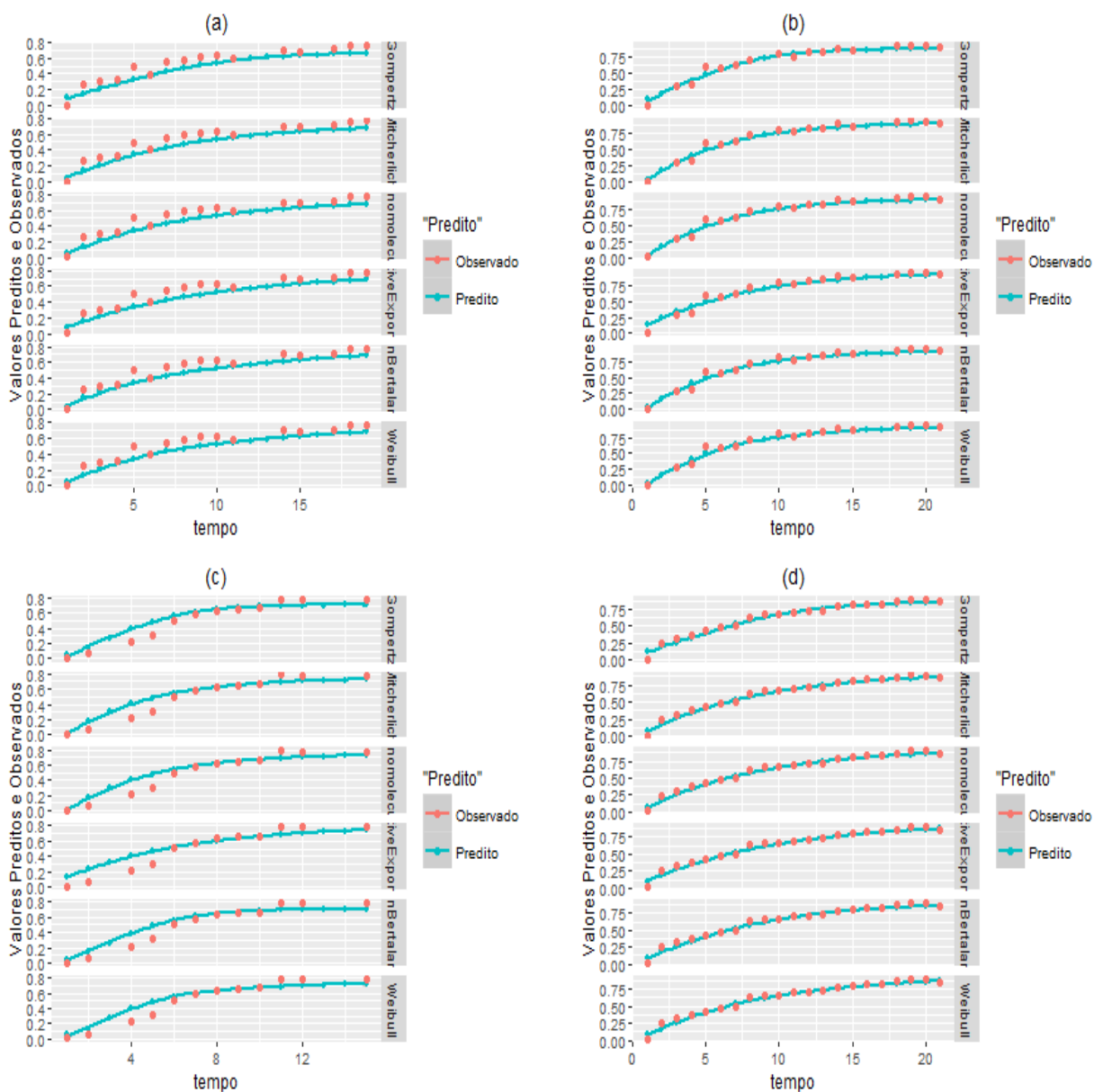
OBS: * Significância estatística de 1%; ** Significância estatística de 5%; *** Significância estatística de 10%; e “ns” Não possui significância estatística.

TABELA D.9 – Critério de Informação de Akaike e Coeficiente de Determinação

Veículo		A6	Carnival	Corolla	Discovery	Hilux	Palio	Passat	SLK
Exp. Neg.	R^2	98,0%	94,5%	99,4%	97,2%	98,5%	99,3%	99,0%	98,7%
	AIC	-122	-73	-149	-102	-131	-144	-136	-125,9
Monomolecular	R^2	99,6%	97,9%	99,6%	99,6%	99,2%	99,6%	99,8%	99,0%
	AIC	-143	-77	-144	-125	-133	-145	-160	-120,9
Mitcherlich	R^2	99,6%	97,9%	99,6%	99,6%	99,2%	99,6%	99,8%	99,0%
	AIC	-143	-77	-144	-125	-133	-145	-160	-120,9
Gompertz	R^2	98,8%	99,1%	99,2%	98,9%	99,7%	99,8%	99,2%	97,2%
	AIC	-122	-87	-133	-108	-151	-156	-132	-104,2
Logistic	R^2	97,4%	98,8%	98,4%	97,4%	99,4%	99,6%	98,2%	95,4%
	AIC	-109	-83	-120	-94	-139	-144	-116	-96,2
Von Bertalanffy	R^2	99,6%	99,0%	99,6%	99,7%	99,7%	99,8%	99,8%	99,4%
	AIC	-132	-76	-133	-118	-140	-145	-149	-117,8
Richards	R^2	99,6%	99,0%	99,6%	99,7%	99,7%	99,8%	99,8%	99,4%
	AIC	-132	-76	-133	-116	-140	-145	-149	-117,8
Weibull	R^2	99,5%	98,4%	99,3%	99,8%	99,2%	99,5%	99,7%	99,5%
	AIC	-130	-71	-126	-128	-123	-128	-138	-130,2

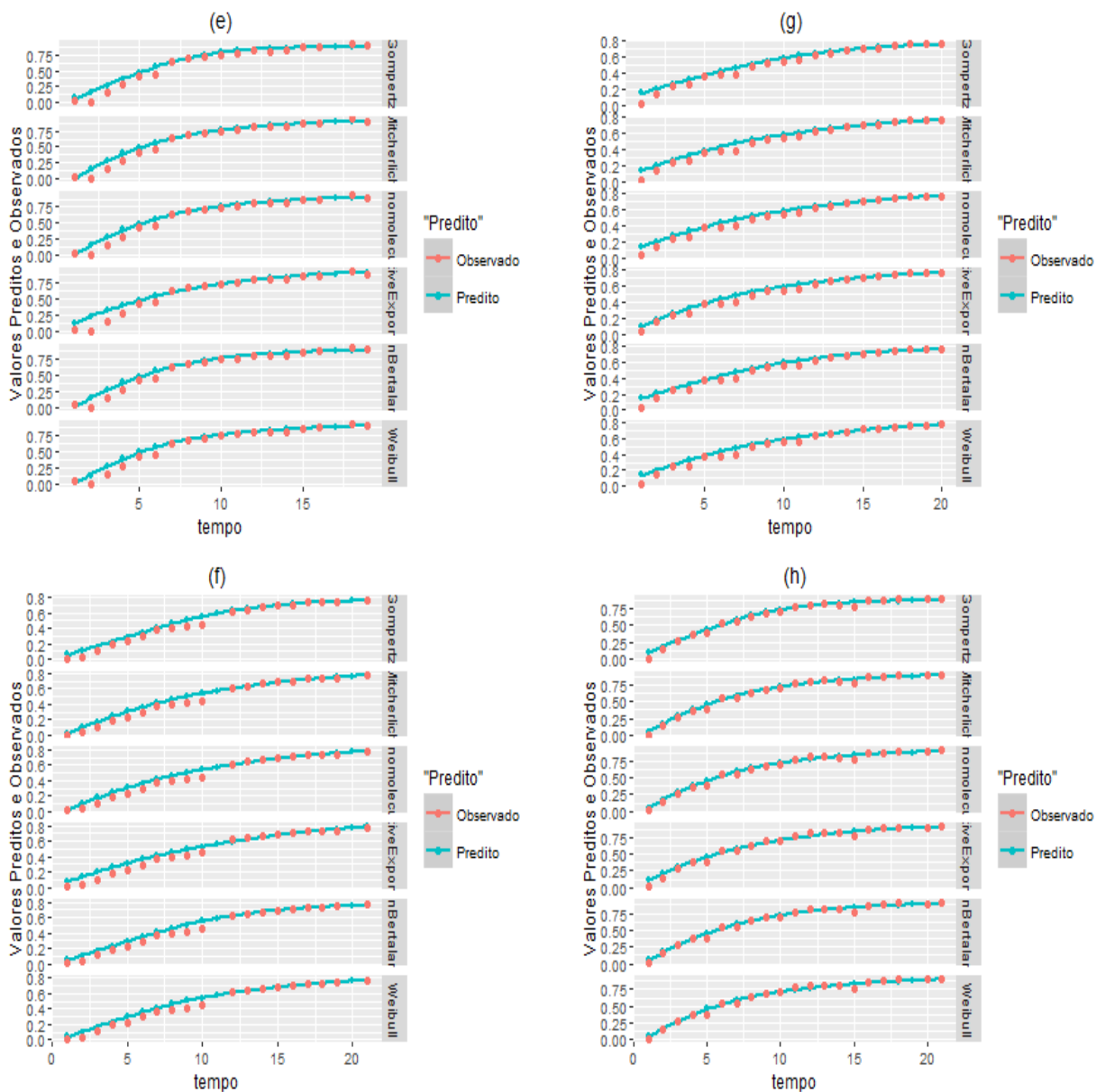
FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.

TABELA D.10 – Gráficos dos Valores Estimados e Observados



FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.
 OBS: Valores Preditos (Linhas) e Observados (Pontos) para os veículos: Mercedes-Benz-SLK (a), Audi-A6 (b), Kia-Carnival (c) e Toyota-Corolla (d).

TABELA D.11 – Gráficos dos Valores Estimados e Observados



FONTE: Base de Dados de Veículos Obtida por *Web Scraping* da Internet Utilizando *Scrapy-Python*.
 OBS: Valores Preditos (Linhas) e Observados (Pontos) para os veículos: Land Rover-Discovery (e),
 Toyota-Hilux (f), Fiat-Palio (g) e Volkswagen-Passat (g).

Anexo A - Rotinas

A.1 Rotinas R: Segmentação de Mercado

```
library(stargazer)
library(doBy)
library(ggplot2)
library(sas7bdat)
usados <- read.sas7bdat("C:/Users/patrick/.../usados.sas7bdat")
usados2 <- usados[c("preco", "preco2", "km", "km2", "idade", "idade2", "kimage", "anuncio", "itens",
  "potencia")]
usados2 <- na.omit(usados2) #listwise deletion of missing
mydata <- scale(usados2) #standardize variables

# Determine number of clusters: NumeroClusters
wss <- (nrow(mydata) - 1) * sum(apply(mydata, 2, var))
for (i in 2:20) wss[i] <- sum(kmeans(mydata, centers=i)$withinss)
plot(1:20, wss, type="b", xlab="Numero de Clusters", ylab="Soma de Quadrados", col="blue")

#Rodada inicial
results <- kmeans(mydata, 7, iter=150)
usados$cluster <- results$cluster
inicial <- summaryBy(preco+preco2+km+km2+idade+idade2+anuncio+itens+kimage+potencia ~
  cluster, data=usados, FUN=c(mean), na.rm=TRUE)
inicial <- inicial[c(-1)]
inicial <- scale(inicial) # standardize variables

#Rodada Definitiva
results <- kmeans(mydata, 7, iter=150, centers=inicial)
usados$cluster <- results$cluster
results$size
results$withinss
results$totss
results$tot.withinss
results$ifault
```



```

usados$usados[usados$tipoa==1]<-"Concessionarias"
usados$usados[usados$tipoa==2]<-"Lojas"
usados$usados[usados$tipoa==3]<-"Particulares"

#Distribuicao do Log KM por Cluster
sp1<-qplot(km,data=rdata,geom="density",fill=xtipoa,alpha=I(.5),log=("x"),ylab="Densidade",
           ,xlab="Log Quilometros")
sp1 + facet_grid( cluster ~ .)

#Distribuicao do Log dos çPreos por Cluster
sp2<-qplot(preco,data=rdata,geom="density",fill=xtipoa,alpha=I(.4),log=("x"),ylab="Densidade",
           ,xlab="Log Precos")
sp2 + facet_grid( cluster ~ .)

#Distribuicao da Idade dos Carros por Cluster
sp3<-qplot(idade,data=rdata,geom="density",fill=xtipoa,alpha=I(.4),ylab="Densidade",xlab="Idade")
sp3 + facet_grid( cluster ~ .)

#Distribuicao da Potencia por Cluster
sp4<-qplot(potencia,data=rdata,geom="density",fill=xtipoa,alpha=I(.4),xlab="Potencia",
           ,ylab="Densidade")
sp4 + facet_grid( cluster ~ .)

#Distribuicao dos Itens de Luxo Carros por Cluster
sp5<-qplot(itens,data=rdata,geom="density",fill=xtipoa,alpha=I(.4),ylab="Densidade",xlab="Itens de Luxo e Seguranca")
sp5+facet_grid( cluster ~ .)

#Distribuicao do Tempo de úAnncio por Cluster
sp6<-qplot(anuncio,data=rdata,geom="density",fill=xtipoa,alpha=I(.4),log=("x"),ylab="Densidade",
           ,"Itens de Luxo e Seguranca")
sp6+facet_grid( cluster ~ .)

library(ggtern)
library(sas7bdat)

rdata <- read.sas7bdat("C:/Users/.../Anlise no R/rdata.sas7bdat")
rdata$lkm <-log(rdata$km,base=exp(1))
rdata$lpreco<-log(rdata$preco,base=exp(1))
rdata$lpot <-log(rdata$potencia,base=exp(1))
rdata$lage <-log(rdata$idade,base=exp(1))

rdata$xcluster[rdata$cluster==1]<-"Cluster 1"
rdata$xcluster[rdata$cluster==2]<-"Cluster 2"
rdata$xcluster[rdata$cluster==3]<-"Cluster 3"
rdata$xcluster[rdata$cluster==4]<-"Cluster 4"
rdata$xcluster[rdata$cluster==5]<-"Cluster 5"

```

```

rdata$xccluster[rdata$cluster==6]<-"Cluster 6"
rdata$xccluster[rdata$cluster==7]<-"Cluster 7"

#TernaryIdade
plot<-ggtern(data=rdata, aes(x=lkm, y=lpreco, z=idade))+geom_point(aes(shape=xccluster, color=
  xccluster), size=1.5)+ggtitle("Distribuicao dos Clusters")+theme(legend.position=c
  (-0.1,0.9), legend.justification=c(0,0.75), legend.box.just='left')
plot+theme_showarrows()+tern_limits(T=0.5,L=0.5,R=0.4)+labs(fill="Clusters", xarrow="Ln(Km
  )", yarrow="Ln(çPreo)", zarrow="Idade")

#TernaryPotencia
plot<-ggtern(data=rdata, aes(x=lkm, y=lpreco, z=potencia))+geom_point(aes(shape=xccluster,
  color=xccluster), size=1.5)+ggtitle("Distribuicao dos Clusters")+theme(legend.position=c
  (-0.1,0.9), legend.justification=c(0,0.75), legend.box.just='left')
plot+theme_showarrows()+tern_limits(T=0.65,L=0.6, R=0.3)+labs(fill="Clusters", xarrow="Log
  (Km)", yarrow="Log(çPreo)", zarrow="Potencia")

#Ternaryitens
plot<-ggtern(data=rdata, aes(x=lkm, y=lpreco, z=itens)) + geom_point(aes(shape=xccluster,
  color=xccluster),
  size=1.5)+ggtitle("Distribuicao dos
  Clusters") + theme(legend.position=c
  (-0.1,0.9), legend.justification=c
  (0,0.75), legend.box.just='left')
plot+theme_showarrows()+tern_limits(T=0.5,L=0.5,R=0.5)+labs(fill="Clusters", xarrow="Log(
  Km)", yarrow="Log(çPreo)", zarrow="Itens de Luxo")

#Ternary: Tempo de Anuncio, Idade, Itens de Luxo
plot<-ggtern(data=rdata, aes(x=anuncio, y=idade, z=itens)) + geom_point(aes(shape=xccluster,
  color=xccluster), size=1.5)+ggtitle("Distribuicao dos Clusters")+theme(legend.position=c
  (-0.1,0.9), legend.justification=c(0,0.75), legend.box.just='left')
plot+theme_showarrows()+tern_limits(T=0.62,L=0.8,R=0.8)+labs(fill="Clusters", xarrow="
  Tempo de Anuncio", yarrow="Idade", zarrow="Itens de Luxo")

```

A.2 Rotinas SAS: Modelo Hedônico Hierárquico

```

proc means data=casas.brasil noprint nway;
class marca;
output out=saida(drop=_type_) max(idade)=wtime;
run;

data saida(drop=FREQ);set saida;
if marca in ('CHEVROLET', 'VOLKSWAGEN', 'FORD', 'FIAT', 'CHRYSLER', 'TROLLER', 'WILLYS',
  CADILLAC') then wEUA=1;else wEUA=0;
if marca in ('PEUGEOT', 'HONDA', 'RENAULT', 'CITROEN', 'MERCEDES-B', 'LAND ROVER', 'BMW', 'AUDI',
  'FERRARI', 'PORSCHE', 'VOLVO', 'DODGE', 'MINI', 'JAGUAR', 'SMART', 'IVECO', 'ALFA ROMEO',
  PUMA') then wEuropa=1;else wEuropa=0;

```

```
if marca in ('KIA', 'TOYOTA', 'MITSUBISHI', 'HONDA', 'HYUNDAI', 'SSANGYONG', 'JAC', 'NISSAN', 'SUZUKI', 'SUBARU', 'JEEP', 'CHERY', 'LIFAN', 'INFINITI') then wAsia=1; else wAsia=0;
```

```
/*1-EUA e Brasil 2-Europa 3-Asia*/
```

```
wshare = _freq_ / 131650;
wtime = (wtime - 29.4) / 410.2;
run;
```

```
/*—— Cap4: Modelos Hedonicos Hierarquicos ——*/
```

```
proc sort data=casas.brasil; by marca; run;
proc sort data=saida ; by marca; run;
```

```
data brasil; merge casas.brasil(in=a) saida(in=b);
by marca;
if a and b;
if 1990 < ano < 2016 and zero = "usado";
```

```
if freq_cod=2 and coleta='junho' then delete;
```

```
if marca in ('CHEVROLET', 'VOLKSWAGEN', 'FIAT', 'FORD', 'PEUGEOT', 'HONDA', 'TOYOTA', 'RENAULT', 'MITSUBISHI', 'CITROEN', 'HYUNDAI', 'MERCEDES-B', 'MERCEDES-BENZ', 'LAND ROVER', 'BMW', 'AUDI', 'KIA', 'NISSAN', 'PORSCHER', 'VOLVO', 'JEEP', 'DODGE', 'CHRYSLER', 'MINI', 'SUZUKI', 'SUBARU', 'TROLLER', 'SSANGYONG', 'CHERY', 'JAGUAR', 'JAC', 'LIFAN', 'FERRARI', 'SMART', 'IVECO', 'WILLYS', 'ALFA ROMEO', 'INFINITI', 'PUMA', 'CADILLAC');
```

```
run;
```

```
/*MHH0*/
```

```
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2;
model lprecox = /solution;
random intercept / sub=marca2 type=vc;
run;
```

```
/*MHH1*/
```

```
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2;
model lprecox = potencia lkm /solution;
random intercept / sub=marca2 type=vc;
run;
```

```
/*MHH2*/
```

```
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2;
model lprecox = potencia lkm wEuropa wtime /solution;
random intercept wEuropa wtime / sub=marca2 type=vc;
run;
```

```
/*MHH3*/
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2;
model lprecox =potencia lkm wEuropa wTime /solution;
random intercept potencia lkm wEuropa wTime / sub=marca2 type=vc;
run;

/*MHH4*/
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2;
model lprecox =potencia lkm anuncio idade valv wEuropa /solution;
random intercept potencia wEuropa / sub=marca2 type=vc;
run;

/*MHH5*/
ods output SolutionF=param;
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 combx tipoa;
model lprecox =potencia lkm anuncio idade valv por wEuropa combx tipoa dairbag dalarme
      darcondic dcdmp3 dcomputa dcontrvel dcouro ddesemba ddirhidr destacio dfreioabs
      dligaleve dregbanco dretrovis dtracao dtravaele dvidroele dvolante /solution;
random intercept potencia wEuropa / sub=marca2 type=vc;
run;

/*—— Modelos Hierarquico com 3 Niveis ——*/

/*MHH0*/
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 modelo;
model lprecox =/solution;
random intercept / sub=marca2 type=vc;
random intercept / sub=modelo(marca2) type=vc;
run;

/*MHH1*/
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 modelo;
model lprecox =potencia lkm/solution;
random intercept / sub=marca2 type=vc;
random intercept / sub=modelo(marca2) type=vc;
run;

/*MHH2*/
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 modelo;
model lprecox =potencia lkm wEuropa /solution;
random intercept wEuropa / sub=marca2 type=vc;
random intercept / sub=modelo(marca2) type=vc;
```

```
run;
```

```
/*MHH3*/
```

```
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 modelo;
model lprecox =potencia lkm wEuropa /solution;
random intercept wEuropa / sub=marca2 type=vc;
random intercept potencia / sub=modelo(marca2) type=vc;
run;
```

```
/*MHH4*/
```

```
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 modelo;
model lprecox =lpot lkm anuncio idade valv wEuropa /solution;
random intercept / sub=marca2 type=vc;
random intercept lpot / sub=modelo(marca2) type=vc;
run;
```

```
/*MHH5*/
```

```
ods output SolutionF=param;
proc mixed data=brasil covtest noclprint method=MIVQUE0;
class marca2 modelo combx tipoa;
model lprecox =potencia lkm anuncio idade valv por wEuropa combx tipoa dairbag dalarme
darcondic dcdmp3 dcomputa dcontrvel dcouro ddesemba ddirhidr destacio dfreioabs
dligaleve dregbanco dretrovis dtracao dtravaele dvidroele dvolante /solution;
random intercept / sub=marca2 type=vc;
random intercept lpot / sub=modelo(marca2) type=vc;
run;
```

A.3 Rotinas SAS: Modelos de Crescimento Exponencial

```
proc sql;
create table saida as
select distinct(ano) as ano,modelo,marca,avg(preco) as preco from carros.Brasil
where modelo in ('SLK','PASSAT','PALIO','CARNIVAL','A6','HILUX','COROLLA','DISCOVERY')
and 1994<ano<2016
group by ano,modelo,marca order by modelo,marca;
quit;
```

```
proc sql;
create table max as select distinct(carro) as carro,max(preco) as maxpreco from saida
group by carro order by carro ;
quit;
```

```
proc sort data=saida;by carro;run;
data saida;merge saida(in=a) max;
```

```
if a;
by carro;
depre_pad = (maxpreco-preco+1000)/(maxpreco+1000);
anox      = 2016-ano;
run;
```

```
Title Negative Exponential;
ods output ResidSummary=Resultados1 ParameterEstimates=Parametros1;
proc model data=saiday maxiter=1000 outparms=param1;
by carro;
parms beta0 beta1 beta2;
depre_pad = beta0*(1-exp(-beta2*anox));
fit depre_pad start=(beta0 0.7 beta1 1 beta2 0.8) / fiml out=resid1(rename=depre_pad=
    NegativeExponential) outall ;
run;quit;
```

```
Title Monomolecular;
ods output ResidSummary=Resultados2 ParameterEstimates=Parametros2;
proc model data=saiday maxiter=1000 outparms=param2;
by carro;
parms beta0 beta1 beta2;
depre_pad = beta0*(1-beta1*exp(-beta2*anox));
fit depre_pad start=(beta0 0.7 beta1 1 beta2 0.8) / fiml out=resid2(rename=depre_pad=
    Monomolecular) outall ;
run;quit;
```

```
Title Mitcherlich;
ods output ResidSummary=Resultados3 ParameterEstimates=Parametros3;
proc model data=saiday maxiter=1000;
by carro;
parms beta0 beta1 beta2;
depre_pad = beta0-beta1*(beta2**anox);
fit depre_pad start=(beta0 0.7 beta1 1 beta2 0.8) / fiml out=resid3(rename=depre_pad=
    Mitcherlich) outall ;
run;quit;
```

```
Title Gompertz;
ods output ResidSummary=Resultados4 ParameterEstimates=Parametros4;
proc model data=saiday maxiter=1000;
by carro;
parms beta0 beta1 beta2;
depre_pad = beta0*exp((-beta1)*exp(-beta2*anox));
fit depre_pad start=(beta0 0.7 beta1 1.8 beta2 0.2) / fiml out=resid4(rename=depre_pad=
    Gompertz) outall ;
run;quit;
```

```
proc transpose data=Parametros4 out=Parametros4;by carro;
id Parameter;
var Estimate tValue;
run;
```

```
data Parametros4;set Parametros4;
if _NAME_='Estimate' then do;
ponto =(1/beta2)*log(beta1);
taxa =beta0/exp(1);
indic =taxa/ponto ;
end;
if _NAME_='tValue' then carro='';
run;
```

```
Title Logistic;
ods output ResidSummary=Resultados5 ParameterEstimates=Parametros5;
proc model data=saiday maxiter=1000;
by carro;
parms beta0 beta1 beta2;
depre_pad = beta0 / ( 1 + beta1*exp(-beta2*anox) );
fit depre_pad start=(beta0 0.7 beta1 1.8 beta2 0.2) / fiml out=resid5(rename=depre_pad=
Logistic) outall ;
quit;
```

```
proc transpose data=Parametros5 out=Parametros5;by carro;
id Parameter;
var Estimate tValue;
run;
```

```
data Parametros5;set Parametros5;
if _NAME_='Estimate' then do;
ponto =(1/beta0)*log(beta1);
taxa =beta0/2;
indic =taxa/ponto ;
eligi = (1/beta2)*log( (1-beta0)/(beta0*beta0) );
end;
if _NAME_='tValue' then carro='';
run;
```

```
Title Von Bertalanffy;
ods output ResidSummary=Resultados6 ParameterEstimates=Parametros6;
proc model data=saiday maxiter=2000;
by carro;
parms beta0 beta1 beta2 beta3;
bounds beta0>0;
nego0=((beta0**(1-beta3))-(beta1*exp(-beta2*anox)));
```

```

nego1=(1/(1-beta3));
depre_pad = nego0**nego1;
fit depre_pad start=(beta0 0.80 beta1 0.80 beta2 0.10 beta3 0.20) / fiml out=resid6(rename
    =depre_pad=VonBertalanffy) outall ;
run; quit;

proc transpose data=Parametros6 out=Parametros6; by carro;
id Parameter;
var Estimate tValue;
run;

data Parametros6; set Parametros6;
if _NAME_='Estimate' then do;
ponto =(1/beta2)*log(beta1/(1-beta3));
temp2 =beta1/(1-beta3);
temp3 =log(beta1/(1-beta3));
taxa =(beta3**(1/(1-beta3)))*beta0;
indic =taxa/ponto ;
end;
if _NAME_='tValue' then carro='';
run;

Title Richards;
ods output ResidSummary=Resultados7 ParameterEstimates=Parametros7;
proc model data=saiday maxiter=2000;
by carro;
bounds beta0>0;
parms beta0 beta1 beta2 beta3;
depre_pad = beta0/((1 + beta1*exp(-beta2*anox))**(1/beta3));
fit depre_pad start=(beta0 0.92 beta1 -1.34 beta2 0.21 beta3 -1.00) / fiml out=resid7
    (rename=depre_pad=Richards) outall ;
run; quit;

proc transpose data=Parametros7 out=Parametros7; by carro;
id Parameter;
var Estimate tValue;
run;

data Parametros7; set Parametros7;
if _NAME_='Estimate' then do;
ponto = log(beta1/beta3)/beta2;
taxa = beta0*((1-beta3)**beta3);
indic = taxa/ponto ;
end;
if _NAME_='tValue' then carro='';
run;

```



```
Title Weibull;
ods output ResidSummary=Resultados8 ParameterEstimates=Parametros8;
proc model data=saiday maxiter=1000;
by carro;
parms beta0 beta1 beta2 beta3;
bounds beta0>0;
depre_pad = beta0 - beta1*exp(-beta2* (anox**beta3) );
fit depre_pad start=(beta0 0.8 beta1 -1 beta2 0.7 beta3 -1) / fiml out=resid8(rename=
    depre_pad=Weibull) outall ;
run;quit;

proc transpose data=Parametros8 out=Parametros8;by carro;
id Parameter;
var Estimate tValue;
run;

data Parametros8;set Parametros8;
if _NAME_='Estimate' then do;
ponto =((beta3-1)/(beta2*beta3)**(1/beta3);
taxa = beta0-beta1*exp((1-beta3)/beta3);
indic = taxa/ponto ;
end;
if _NAME_='tValue' then carro='';
run;
```

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO TD	2. DATA 29 de fevereiro de 2016	3. DOCUMENTO Nº FACE/UnB/TD-999/2016	4. Nº DE PÁGINAS 161
5. TÍTULO E SUBTÍTULO: Uma Análise Empírica do Mercado de Revenda de Carros Usando Raspagem de Dados da Internet			
6. AUTOR(ES): Patrick Franco Alves			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Departamento de Economia - Universidade de Brasília - UnB			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: Big Data, Preços Hedônicos, Carros Usados, Comunalidade			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Big Data, Preços Hedônicos, Carros Usados, Comunalidade			
10. APRESENTAÇÃO: <input checked="" type="checkbox"/> Nacional <input type="checkbox"/> Internacional UnB, Brasília. Curso de Doutorado. Programa de Pós-Graduação em Economia. Área de Economia Aplicada. Orientador: Prof. Dr. Daniel Oliveira Cajueiro. Defesa em 29/02/2016. Publicada em 03/03/2016.			
11. RESUMO: Estudamos o mercado secundário de automóveis utilizando anúncios de venda de carros na internet. O mercado secundário de carros é importante por introduzir uma série de substitutos parciais para os carros novos. Apesar de sua relevância, existem relativamente poucos estudos sobre o mercado de carros usados. Contribuímos para a literatura ilustrando como as técnicas de <i>Big Data</i> podem contribuir para pesquisa econômica. Inspirados pelo paradigma do <i>Big Data</i> , nós construímos uma base de dados utilizando <i>Python</i> e <i>Web Scraping</i> . A segmentação de carros usados, utilizando um algoritmo não-supervisionado, revela clusters que possuem baixa correlação com os segmentos de carros zero quilômetro. Sabe-se que as montadoras projetam carros utilizando comunalidade para obter economias de custos. Usando modelos de comunalidade, preços hedônicos e modelos hierárquicos, verificamos que os efeitos da comunalidade sobrevivem no mercado secundário de carros. Por último, caracterizamos a perda de preço dos veículos utilizando modelos de crescimento exponencial, onde o ponto de inflexão determina o momento de desaceleração da curva de depreciação. Consumidores de carros zero irão absorver grande perda de preços no momento da revenda, enquanto consumidores de veículos com idade maior que o ponto de inflexão não terão grandes perdas de preço.			
12. GRAU DE SIGILO: <input checked="" type="checkbox"/> OSTENSIVO <input type="checkbox"/> RESERVADO <input type="checkbox"/> CONFIDENCIAL <input type="checkbox"/> SECRETO			