UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E
CONTABILIDADE DE RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA
ÁREA: ECONOMIA APLICADA

CAIO MATTEÚCCI DE ANDRADE LOPES

Estimando a aversão ao risco no mercado de seguros de automóveis

ORIENTADORA: PROFA. DRA. ELAINE TOLDO PAZELLO

RIBEIRÃO PRETO
2015

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Dante Pinheiro Martinelli
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Renato Leite Marcondes
Chefe do Departamento de Economia

CAIO MATTEÚCCI DE ANDRADE LOPES

Estimando a aversão ao risco no mercado de seguros de automóveis

Dissertação apresentada ao Programa de Pós-Graduação em Economia - Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Orietadora: Profa. Dra. Elaine Toldo Pazello

Ribeirão Preto - SP

2015

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTA TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Lopes, Caio Matteúcci de Andrade

Estimando a aversão ao risco no mercado de seguros de automóveis.

Ribeirão Preto, 2015.

56 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo.

Orientador: Pazello, Elaine Toldo

1. Aversão ao risco 2. Mercado de seguros 3. Inferência bayesiana

FOLHA DE APROVAÇÃO

Nome: LOPES, Caio Matteúcci de Andrade

Título: Estimando a aversão ao risco no mercado de seguros de automóveis

Dissertação apresentada ao Programa de Pós-Graduação em Economia - Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Aprovada em: _____

BANCA EXAMINADORA

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

AGRADECIMENTOS

Durante a realização deste trabalho tive o apoio de muitas pessoas. Primeiramente, gostaria de agradecer meus pais, Carlos e Rosana, meu irmão Vinícius, minha cunhada Cláudia e meu sobrinho Bento pela companhia e pelos bons momentos vividos durante esse período.

Gostaria de agradecer aos professores da FEARP, em especial, Bruno Ledo, Elaine Pazello, Cláudio Lucinda e Márcio Laurini pelos conselhos e sugestões. Sou grato também aos colegas de mestrado que, de alguma forma, colaboraram para a conclusão desse trabalho.

Por fim, gostaria de agradecer a Renata pelas risadas, carinhos, sugestões, revisões ortográficas, e todos os momentos felizes que partilhamos até hoje.

Sumário

	Sumário	7
	Lista de tabelas	9
	Lista de ilustrações	10
1	Introdução	13
2	Revisão Bibliográfica	16
3	Objetivos e Motivação	18
4	Dados	20
4.1	Características Individuais	21
4.2	Prêmios e Franquias	21
5	Metodologia	24
5.1	Abordagem teórica	24
5.2	Modelo Econométrico	26
6	Inferência Bayesiana	29
6.1	Introdução à Inferência Bayesiana	29
6.2	MCMC - Monte Carlo via Cadeia de Markov	31
6.2.1	Cadeia de Markov e Matriz de Transição	31
6.2.2	Distribuição de Probabilidade	32
6.2.3	Distribuição Estacionária e Convergência	33
6.2.4	Definições	34
6.3	Algoritmo MCMC	35
6.3.1	Amostragem de <i>Gibbs</i>	36
6.3.2	Algoritmo <i>Metropolis-Hastings</i>	37
6.3.3	<i>Sliced Sampler</i>	38
7	Resultados	40
7.1	Estimação - Forma Reduzida	40
7.2	Estimação - Inferência Bayesiana	41
8	Conclusão	46
	Referências	48
A	DESCRIÇÃO DA AMOSTRAGEM DE <i>GIBBS</i> (GS)	50
B	ESTATÍSTICAS DESCRITIVAS DAS <i>DUMMIES</i> DE CEP	53

C	RESULTADO DAS <i>DUMMIES</i> DE CEP - ESTIMAÇÃO BAYESIANA	56
---	---	----

Lista de tabelas

Tabela 1	– Estatísticas Descritivas - Covariadas	22
Tabela 2	– Estatísticas descritivas - Prêmio, Franquia e Acionamento	22
Tabela 3	– Ausência de Heterogeneidade no Risco	41
Tabela 4	– Resultados - Estimação Bayesiana	43
Tabela 5	– Estimativas de Aversão ao risco	44
Tabela 6	– Estatísticas Descritivas das <i>dummies</i> de CEP	53
Tabela 7	– Resultado das <i>dummies</i> de CEP - Estimação Bayesiana	56

Lista de ilustrações

Figura 1 – Estimativa da Distribuição Conjunta - Risco e Aversão	45
--	----

Resumo

LOPES, C. M. A. **Estimando a aversão ao risco no mercado de seguros de automóveis**. 2015. 56 p. Dissertação (Mestrado) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2015.

O objetivo deste trabalho é estimar a distribuição conjunta do risco e da aversão ao risco no mercado de seguros de automóveis. Para tal, será utilizado o modelo estrutural proposto por Cohen e Einav (2007), que permite identificar esta distribuição à partir das coberturas escolhidas pelos segurados e dos sinistros declarados. Na metodologia empírica, utilizamos o método de Monte Carlo via Cadeias de Markov (MCMC). A base de dados utilizada se refere à apólices de seguros transacionadas na região metropolitana de São Paulo, apenas para a seguradora com maior participação neste mercado. Os resultados obtidos indicam que os coeficientes de aversão ao risco absoluto apresentam média baixa, mediana ainda menor e elevada heterogeneidade não observada. Observou-se também uma correlação negativa entre o risco e a aversão ao risco.

Palavras Chave: 1. Aversão ao risco 2. Mercado de seguros 3. Inferência bayesiana

Abstract

LOPES, C. M. A. **Estimating Risk Preferences From Auto Insurance Market**. 2015. 56 p. Dissertation (Master Degree). Graduate School of Economics, Business and Accounting, University of São Paulo, Ribeirão Preto, 2015.

This study aims to estimate the distribution of risk aversion from the car insurance market. For this, the method proposed by Cohen e Einav (2007) model that allows unobserved risk is used. The data refer to the metropolitan area of São Paulo with an analysis restricted to only one insurer. The methodology will be the Gibbs sampling which enables increased data risk of latent variables and risk aversion. The results indicate a small mean level of absolute risk aversion and even lower median, featuring high dispersion coefficients.

Key words: 1. Risk preferences 2. Insurance market 3. Bayesian inference

1 Introdução

As decisões sob incerteza estão presentes em diversas áreas da economia. Os agentes podem reagir de diferentes formas frente às situações de risco, pois possuem preferências heterogêneas em relação ao mesmo. Neste contexto, um estudo empírico, em determinado mercado, a fim de estimar tais preferências torna-se muito relevante. Um mercado de seguros em que os agentes procuram compartilhar seus riscos é o ponto de partida ideal para estudar a aversão ao risco dos indivíduos. Este trabalho pretende estimar a aversão ao risco e o risco de uma população a partir de dados sobre o mercado de seguro de automóveis em São Paulo.

Existe, porém, no mercado de seguros fortes evidências acerca da presença de assimetria de informação, ou seja, é provável que o segurado possua algum tipo de informação privada, não observável pelas seguradoras. Em relação à assimetria, pode-se dividi-la em dois contextos: antes da assinatura do contrato e durante a vigência do mesmo. O primeiro caso, conhecido como seleção adversa consiste na existência de informação privada por alguma das partes (contratante ou contratado) desde antes do acordo contratual. O segundo refere-se a possibilidade que os indivíduos têm de influenciar o comprimento do contrato a partir de suas ações durante a vigência do mesmo.

Segundo Chiappori e Salanié (2000), a diferença entre seleção adversa e risco moral é mais uma questão de interpretação que estrutura. Na presença de seleção adversa, a escolha de um contrato é determinada pelas preferências dos indivíduos (por exemplo, aversão ao risco), que podem estar relacionadas com o risco. Nesse caso, os agentes levam em conta não só suas preferências, mas também seu risco, na hora de escolher uma cobertura. Já com risco moral, o indivíduo que escolheu um contrato com maior cobertura tem menos incentivo a prevenir-se contra acidente. Apesar de, em ambos os casos, existir a correlação entre a escolha da cobertura e o risco de acidente, a relação de causalidade é invertida. Portanto, diferentemente da seleção adversa, em caso de risco moral os agentes não escolhem determinada cobertura por causa do baixo risco, mas sim, o risco é baixo devido aos incentivos que o contrato escolhido proporciona.

O estudo citado acima mostrou que, muitas vezes, os modelos de assimetria de informação no mercado de seguros não possuem sustentação empírica. Isto é, os dados não mostram correlação entre a escolha da cobertura e o risco de acidente. Com isso, a escolha de um modelo teórico que permita (ou não) assimetria de informação - seleção adversa e/ou risco moral - mostra-se pouco trivial. Frente a tais dificuldades metodológicas, inferir sobre a aversão ao risco a partir de dados do mercado de seguros pode ser, por um lado, atraente, já que a causa primária da existência das seguradoras é a aversão ao risco, por outro, complexo, devido às possíveis assimetrias deste mercado.

Diante disso, o presente trabalho visa aplicar o modelo proposto por Cohen e Einav (2007), a fim de estimar a distribuição do coeficiente de aversão ao risco absoluto por meio do mercado de seguros de automóveis. A abordagem desenvolvida por esses autores consiste em um modelo estrutural com heterogeneidade não observada tanto no risco quanto na aversão ao risco (duas fontes de assimetria de informação), o que permite a presença de seleção adversa. Os autores utilizaram dados da principal seguradora de automóveis de Israel em sua análise. Similarmente, utilizaremos dados da principal seguradora brasileira, que foram disponibilizados pela Superintendência de Seguros Privado (SUSEP), para obter os resultados desejados.

Em linhas gerais, os resultados obtidos neste estudo assemelham-se ao encontrados por Cohen e Einav (2007). A média da distribuição do coeficiente de aversão ao risco absoluto estimado é baixa e a mediana muito próxima de zero. Encontrou-se elevada heterogeneidade não observada na equação do risco e, principalmente, na da aversão ao risco, o que corrobora um importante resultado também encontrado por Cohen e Einav (2007). Além disso, os resultados sugerem correlação negativa não só entre o risco e a aversão estimadas, mas também entre os componentes não observáveis das duas equações. As mulheres se mostraram mais avessas ao risco em relação aos homens enquanto a faixa etária mais avançada revelou-se a mais avessa ao risco.

O texto está organizado em 8 seções, além desta primeira, introdutória. Na segunda seção, será apresentada uma revisão bibliográfica composta por duas partes. A primeira, refere-se à alguns trabalhos que modelaram mercados de seguros com assimetrias de informação, enquanto, a segunda, reúne os principais estudos acerca da aversão ao risco.

A terceira seção traz os objetivos desta pesquisa, bem como sua relevância. O conceito de coeficiente de aversão ao risco absoluto também é desenvolvido nesta seção. A quarta, consiste na seleção amostral e análise descritiva dos dados. Na seção 5, são desenvolvidos os modelos teórico e empírico. A sexta seção trata de alguns conceitos importantes de Inferência Bayesiana aplicados neste trabalho. Na sétima seção, são apresentados os resultados e interpretação da estimação do modelo. Por fim, a última seção apresenta as conclusões encontradas neste trabalho.

2 Revisão Bibliográfica

A literatura sobre mercado de seguros procura modelar as assimetrias de informação de diferentes formas. O modelo de seleção adversa proposto por Landsberger e Meilijson (1999), permite, assim como o presente estudo, heterogeneidade no risco dos indivíduos. Os autores apresentaram uma abordagem unificada na qual os agentes podem diferir tanto em relação à utilidade quanto em relação ao risco. Mais próximo ao modelo deste trabalho, o estudo de Smart (2000) analisa o equilíbrio de um mercado de seguros competitivo em que os segurados diferem tanto na aversão ao risco quanto na probabilidade de acidente.

Ainda no contexto de assimetria de informação, é impossível não citar o trabalho de Chiappori e Salanié (2000), em que os autores encontraram evidência empírica de um *puzzle*. Apesar do modelo de Rothschild e Stiglitz (1976) mostrar que um mercado de seguros em equilíbrio é caracterizado pela correlação positiva entre risco e cobertura, os autores não encontraram, por meio de um teste de dependência condicional, correlação entre estas variáveis.

Muitos trabalhos procuraram explicar os resultados obtidos por Chiappori e Salanié (2000) como Chiappori et al. (2006) e Cutler, Finkelstein e McGarry (2008). O primeiro concluiu que a aversão ao risco afeta não só a escolha da cobertura, mas também os incentivos referentes às medidas de cautela durante a vigência. Os autores argumentam também que a aversão ao risco é intrínseca ao indivíduo e, portanto, não pode ser diretamente observada pelas seguradoras. Por fim, segundo eles, para indivíduos heterogêneos tanto no risco quanto na preferência sobre o risco, a relação entre cobertura e risco pode ter qualquer sinal.

Já Cutler, Finkelstein e McGarry (2008), após analisarem cinco mercados de seguros distintos, concluíram que as preferências heterogêneas sobre o risco - e seu impacto no acionamento e escolha do nível de cobertura - podem ajudar a explicar os diferentes padrões de seleção observados em diversos mercados de seguros.

Nesse sentido, o modelo de Cohen e Einav (2007) que será aplicado neste estudo se mostra muito interessante, pois, apesar de não permitir que o risco dependa diretamente

da escolha da cobertura (risco moral), admite que a escolha do contrato é explicada pelo par risco e aversão ao risco, que possuem fatores não observados pela seguradora. Com isso, os autores fazem uso deste arcabouço estrutural para estimar a distribuição do coeficiente de aversão ao risco absoluto, empregando o método de Monte Carlo via Cadeias de Markov (MCMC).

Na literatura, muitas pesquisas evidenciam, e até estimam, a presença de aversão ao risco por meio da análise individual em diferentes situações. Smith e Walker (1993) utilizaram dados sobre leilão de primeiro preço para mostrar que os indivíduos tendem a ser mais avessos ao risco. Já Jullien e Salanié (2000) analisaram o caso de apostadores em pistas de corrida a fim de estimar as preferências sob incerteza, por meio de um modelo multinomial.

Saha (1997), Chetty (2006) e Paravisini, Rappoport e Ravina (2010) também estimaram a aversão ao risco dos agentes voltando-se para as decisões de produção das firmas, oferta de trabalho e mercado financeiro, respectivamente. Além disso, Paravisini, Rappoport e Ravina (2010) ressaltam a importância da heterogeneidade não observada na estimação (não viesada) das preferências sobre o risco. Tema este, também abordado por Barseghyan et al. (2013), por meio de um modelo estrutural muito próximo ao de Cohen e Einav (2007).

Ainda no contexto de aversão, outra questão relevante remete à estabilidade das preferências sobre o risco frente a diferentes mercados. Barseghyan, Prince e Teitelbaum (2011) propuseram um teste para verificar a estabilidade das preferências a partir de uma única base de dados com informações sobre seguros de automóveis e casas. Os autores encontraram que a hipótese de estabilidade das preferências sobre o risco foi rejeitada. Einav et al. (2010) concluíram que a existência de um único fator responsável pelas escolhas em ambientes de incerteza está sistematicamente relacionado às características individuais, isto é, a estabilidade das preferências é maior para os indivíduos mais ricos e mais velhos.

3 Objetivos e Motivação

O objetivo central deste trabalho é estimar a distribuição do coeficiente de aversão ao risco absoluto a partir de dados sobre o mercado de seguros de automóveis em São Paulo. Para isso, será utilizado um modelo estrutural com seleção adversa. Pretende-se analisar como o nível de aversão ao risco está relacionado às características observáveis dos segurados. Com isso, poderemos responder perguntas interessantes, como:

- Em qual faixa etária o indivíduo é mais avesso ao risco?
- Como o gênero influencia as diferenças de aversão ao risco?
- Como a heterogeneidade não observada do risco e aversão ao risco se comportam?

Diante desse contexto, deve-se esclarecer qual a importância de estimar a aversão ao risco dos agentes. Diversos modelos procuram explicar as relações econômicas mas, para isso, é preciso definir, muitas vezes, as relações de preferências dos agentes. No caso do coeficiente de aversão ao risco, o interesse está em descobrir quais são as preferências dos indivíduos em ambientes de incerteza (loterias), portanto estimar o coeficiente de aversão ao risco é o passo inicial para interpretar, explicar e prever modelos econômicos.

A existência do mercado de seguros de automóveis justifica-se, principalmente, pelo risco de acidentes, desta forma, o agente incorreria em uma perda de riqueza devido à depreciação de seu automóvel. Contudo, o agente pode estar disposto a pagar uma quantia - de acordo com suas preferências - para não se expor à situação de incerteza. Formalmente, a diferença entre a utilidade do valor esperado de uma loteria e a utilidade esperada da mesma reflete de algum modo a postura do agente diante do risco.

Suponha que os elementos do conjunto de resultado sejam valores monetários, ou seja, $A = \{w_1, \dots, w_2, \dots, w_3\}$, sendo $w_i \geq 0$ a riqueza associada ao i -ésimo resultado. Seja $u(\cdot)$ uma função utilidade VMN de um indivíduo representando sua relação de preferência.

Define-se utilidade esperada da loteria e utilidade do valor esperado da loteria, l , respectivamente por:

$$u(l) = \sum_n^{i=1} p_i u(w_i)$$

$$u(E(l)) = u\left(\sum_{i=1}^n p_i w_i\right)$$

Então, para uma loteria simples $l = (p_1 w_1, \dots, p_n w_n)$, o indivíduo pode apresentar:

- Aversão ao risco em l se $u(E(l)) > u(l)$.
- Neutralidade ao risco em l se $u(E(l)) = u(l)$.
- Propensão ao risco em l se $u(E(l)) < u(l)$.

Portanto, o comportamento do tomador de decisão perante o risco está diretamente ligado à curvatura da função utilidade VMN. Uma forma de avaliar esta curvatura é através de Medida Arrow-Pratt de aversão absoluta ao risco, $r(w) = \frac{-u''(w)}{u'(w)}$, tal que para $r(w) = 0$ o indivíduo apresenta neutralidade ao risco.¹

¹ A metodologia empregada nesse trabalho não admite que os indivíduos sejam propensos ao risco.

4 Dados

A base de dados usada neste trabalho foi disponibilizada pela Superintendência de Seguros Privados, Susep. Criada em 1966 como uma Autarquia vinculada ao Ministério da Fazenda, este órgão é responsável pelo controle e fiscalização dos mercados de seguro, previdência privada aberta, capitalização e resseguro.

Neste estudo, os dados são referentes ao período de 01 de janeiro à 31 de junho de 2011 e estão de acordo com o regimento estabelecido na circular nº. 360 de fevereiro de 2008. A base original possui diversas informações que não possuem finalidade para o propósito deste trabalho e, portanto, foram desconsideradas. Primeiramente, restringimos nossa análise a região metropolitana de São Paulo e a seguradora Porto Seguro. A primeira representa a região com o maior número de apólices de seguro de automóveis, enquanto a segunda é a empresa com maior participação no mercado de seguro de automóveis.

Considerou-se apenas os veículos de passeio - nacionais ou importados - cujos contratos possuíam cobertura abrangente². Foi mantida na base apenas as apólices referentes à carros de pessoas físicas. Originalmente, a base apresenta três tipos de franquia: reduzida, normal e majorada. Em virtude da metodologia empregada neste trabalho, a franquia do tipo majorada, que representa menos de 1% da amostra, foi considerada como normal. Segundo, Cohen e Einav (2007), esta abordagem não causa nenhum tipo de viés, já que, de acordo com o modelo estrutural desenvolvido, o indivíduo que escolheu a franquia majorada optaria pela franquia normal em detrimento à reduzida. Por fim, foram desconsideradas as apólices que sofreram endosso, ou seja, qualquer alteração durante sua vigência e as apólices coletivas.

² Cobertura Abrangente é a mais utilizada no seguro auto, oferecendo cobertura sobre: colisão, incêndio e roubo. É também ampla, pois contempla os casos de perda parcial e perda total do veículo.

4.1 Características Individuais

Basicamente a base de dados pode ser divididas em três segmentos. Primeiramente, as covariadas, que representam as características observáveis dos segurados, que são: gênero, faixa etária, importância segurada, ano do modelo do veículo. As variáveis referentes ao par prêmio-franquia estão atreladas, a princípio, a três tipos de cobertura: reduzida, normal e majorada. Finalmente, em caso de sinistro, sabe-se o número de vezes que o seguro foi acionado por contrato.

A tabela 1 apresenta informações sobre as características dos segurados. Estas estão divididas em três grupos: demográficas, características do carro e endereço postal. As variáveis demográficas limitam-se à idade³ e gênero.⁴ Nessa amostra, a média de idade dos segurados é de 41 anos, em que a maioria é do sexo masculino. Dentre as características do carro, existem três *dummies*, que se referem ao ano do modelo do carro (antes de 2008, entre 2008 e 2010, após 2010), informações sobre a importância segurada do casco e potência do motor. Além das variáveis já mencionadas, o modelo inclui *dummies* com os três primeiros dígitos do Código de Endereçamento Postal (CEP) de utilização do veículo. Esta variável visa captar os efeitos da região de circulação do veículo sobre o risco de acidente. Como existem, para este caso, 77 *dummies*, a tabela com a análise descritiva não será reportada aqui.⁵

4.2 Prêmios e Franquias

Após conhecer o vetor de características individuais, x_i , a seguradora oferece um *menu* com três opções de contrato. Dentre elas, a franquia normal, ou regular, é a mais escolhida entre os segurados e, relativamente, similar entre as seguradoras. A franquia reduzida corresponde à 50% da franquia normal, enquanto a franquia majorada equivale ao dobro da regular. Dessa forma, sabe-se o valor da franquia para os três tipos de contrato. O prêmio

³ Na estimação só foram incluídas *dummies* que identificam 6 coortes de idade.

⁴ A variável gênero assume valor 1 para homens e 0 para mulheres.

⁵ A análise descritiva dos Códigos de Endereçamento Postal encontram-se no apêndice B.

Tabela 1: Estatísticas Descritivas - Covariadas

Variável^a		Média	D.P.	Min.	Max.
Demográficas:	Gênero	0,571	0,495	0	1
	Idade	41,243	13,371	18	100
	Idade 18-25	0,1	0,299	0	1
	Idade 26-35	0,305	0,46	0	1
	Idade 36-45	0,236	0,424	0	1
	Idade 46-55	0,195	0,396	0	1
	Idade 56-65	0,116	0,32	0	1
	Idade > 66	0,05	0,217	0	1
Características do carro:	IS Casco ^b	31781,739	15512,206	4805	463723
	Modelo Ano < 2008	0,313	0,464	0	1
	Modelo Ano 2008-2010	0,363	0,481	0	1
	Modelo Ano > 2010	0,324	0,468	0	1
	Potência Motor	1,338	0,394	1	6,100

^aAs informações sobre o código postal não foram reportadas.

^bValor referente à importância Segurada do Casco

Nota: 114870 observações.

Fonte: elaboração própria.

de cada franquia varia entre os indivíduos de acordo com uma função determinística, $p_i = f(x_i)$. Como, não se tem acesso ao algoritmo de precificação de contratos praticado pela seguradora, realizou-se 20 cotações via internet variando características do carro e do indivíduo. O resultado mostrou que escolher a franquia reduzida em relação à normal aumenta o prêmio em 21%. Além disso, a análise é focada nas franquias normal e reduzida, de modo que as franquias majoradas foram consideradas normais.

Tabela 2: Estatísticas descritivas - Prêmio, Franquia e Acionamento

Variável		Média	D.P.	Min.	Max.	N
Franquia	Reduzida	1012,478	343,572	200,5	14356	112630
	Normal	2024,957	687,144	401	28712	112630
Prêmio	Reduzido	1451,586	724,38	60,5	26225,539	112630
	Normal	1199,658	598,662	50	21674	112630
$\Delta p / \Delta d$		0,259	0,151	0,01	3,927	112630
Escolha	F. Reduzida	0,139	0,346	0	1	112630
	F. Normal	0,861	0,346	0	1	112630
Acionamento	Ambos	0,037	0,194	0	2	112630
	Reduzida	0,044	0,208	0	2	15704
	Normal	0,036	0,191	0	2	96926
Acionamento por Semestre	Ambos	0,056	0,678	0	181	112630
	Reduzida	0,063	0,412	0	18,1	15704
	Normal	0,055	0,711	0	181	96926

Fonte: elaboração própria.

Com isso, pode-se calcular os valores do *menu* - prêmio e franquia - para ambos os cenários (franquia reduzida e normal). A análise descritiva referente à estes dados podem ser observadas na tabela 2. Vale ressaltar que a variável *Acionamento por semestre* representa o número de vezes que o seguro foi acionado dividido pela exposição do segurado. Isto é, se a apólice permaneceu ativa durante todo o período analisado (primeiro semestre de 2011), a exposição foi de 1. Da mesma maneira, se o contrato perdurou apenas um dia no semestre, a exposição foi de 1/181⁶. A partir da avaliação desta variável, observa-se que a taxa semestral de acionamento do seguro, em relação a todos os indivíduos, foi de 0,056. Contudo, quando focamos em apenas um tipo de franquia, os dados mostram que o acionamento por semestre foi um pouco maior para a franquia reduzida. Em média, os indivíduos que escolheram franquia reduzida tiveram taxas de acionamentos maiores (0,063) em relação aos que escolheram franquia normal (0,055).

⁶ O valor 181 remete ao número de dias do primeiro semestre de 2011

5 Metodologia

5.1 Abordagem teórica

O modelo teórico desenvolvido por Cohen e Einav (2007) baseia-se na ideia do agente indiferente entre dois contratos. Estes correspondem a um par prêmio e franquia, de tal forma que (p_i^h, d_i^h) e (p_i^l, d_i^l) representam os contratos para o indivíduo i de franquia normal e reduzida, respectivamente. Além disso, seja w_i a riqueza do indivíduo i e $u_i(w)$ sua correspondente função utilidade do tipo vNM. O tempo de contrato é representado por t_i . Assume-se que o seguro é acionado de acordo com uma distribuição de Poisson com uma taxa anual λ_i . Em outras palavras, λ_i é o risco inerente a cada indivíduo e, por hipótese, de conhecimento próprio. Assume-se também que λ_i independe da escolha da franquia, ou seja, não há risco moral. Por fim, a última hipótese estabelece que, em caso de acidente, a indenização paga deve ser maior que d_i^h . No resto desta seção, o subscrito i será omitido por conveniência.

Este modelo estabelece que tanto o prêmio como o risco são proporcionais ao tempo de contrato. Cohen e Einav (2007) explicam que esta abordagem possui três vantagens. A primeira é que ajuda a lidar com contratos cancelados, ou que possuem um período de tempo menor. A segunda, refere-se ao fato de permitir que a escolha da franquia seja independente das incertezas de longo prazo, possibilitando o foco nas preferências sobre o risco no curto prazo. A terceira vantagem resulta de uma conveniência analítica e computacional.

A utilidade esperada que o indivíduo obtém da escolha do contrato (p, d) é dado por:

$$v(p, d) \equiv (1 - \lambda t)u(w - pt) + (\lambda t)u(w - pt - d) \quad (5.1.1)$$

Com isso, pode-se caracterizar o conjunto de parâmetros que torna o indivíduo indiferente entre os contratos com franquia normal e reduzida. Isso permite definir um limite inferior (superior) para o nível de aversão ao risco dos indivíduos que escolheram franquia

reduzida (normal) para uma dado λ . Aplicando o limite em relação a t e utilizando a regra de *L'Hopital*, observa-se:

$$\begin{aligned}\lambda &= \lim_{t \rightarrow 0} \frac{\frac{1}{t}(u(w - p^h t)) - u(w - p^l t)}{((u(w - p^h t) - u(w - p^h t - d)) - (u(w - p^h t) - u(w - p^l t - d^h)))} \\ &= \frac{(p^l - p^h)u'(w)}{u(w - d^l) - u(w - d^h)}\end{aligned}\tag{5.1.2}$$

rearranjando

$$(p^l - p^h)u'(w) = \lambda(u(w - d^l) - u(w - d^h))\tag{5.1.3}$$

A expressão 5.1.3 possui uma simples interpretação: o lado direito representa o ganho esperado de utilidade por unidade de tempo de escolher a franquia baixa, enquanto o lado esquerdo equivale ao custo dessa escolha por unidade. Para que o indivíduo seja indiferente entre os dois contratos os ganhos esperados devem ser iguais aos custos. O coeficiente de aversão ao risco absoluto do indivíduo indiferente pode ser calculado a partir da hipótese de que a terceira derivada da utilidade vNM não é muito grande. Com isso, aplicando uma expansão de Taylor nos dois termos do lado direito da equação 5.1.3 obtém-se, de maneira geral, $u(w - d) \approx u(w) - du'(w) + (d^2/2)u''(w)$, o que implica em:

$$\frac{p^l - p^h}{\lambda}u'(w) \approx (d^h - d^l)u'(w) - \frac{1}{2}(d^h - d^l)(d^h + d^l)u''(w)\tag{5.1.4}$$

Renomeando as variáveis, $\Delta d \equiv d^h - d^l > 0$, $\Delta p \equiv p^l - p^h > 0$ e $\bar{d} \equiv \frac{1}{2}(d^h + d^l)$, temos:

$$\frac{\Delta p}{\lambda \Delta d}u'(w) \approx u'(w) - \bar{d}u''(w)\tag{5.1.5}$$

ou

$$r \equiv \frac{-u''(w)}{u'(w)} \approx \frac{\frac{\Delta p}{\lambda \Delta d} - 1}{\bar{d}}\tag{5.1.6}$$

Em que r é o coeficiente de aversão ao risco absoluto dada a riqueza w . Sendo assim, a equação 5.1.6 define o conjunto de indiferença que relaciona, a partir dos dados referentes a prêmio e franquia, as variáveis risco e coeficiente de aversão ao risco absoluto $(r^*(\lambda), \lambda)$.

Ambas são específicas dos indivíduos, já que dependem da escolha do menu de contratos, que varia entre os segurados. Desta forma, seja um indivíduo i representado por um par (r_i, λ_i) , a quem é oferecido um menu de contratos $\{(p_i^h, d_i^h), (p_i^l, d_i^l)\}$, então este escolherá o contrato de baixa franquia se, e somente se, seu coeficiente de aversão ao risco absoluto satisfaz $r_i > r_i^*(\lambda)$.

5.2 Modelo Econométrico

O objetivo é estimar a distribuição conjunta entre risco e o coeficiente de aversão ao risco absoluto, (λ_i, r_i) , na população de segurados, condicional as variáveis observadas. Para isso, assume-se que (λ_i, r_i) seguem uma distribuição lognormal bivariada, de modo que:

$$\ln \lambda_i = x_i' \beta + \epsilon_i \quad (5.2.1)$$

$$\ln r_i = x_i' \gamma + \nu_i \quad (5.2.2)$$

com

$$\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & \rho \sigma_\lambda \sigma_r \\ \rho \sigma_\lambda \sigma_r & \sigma_r^2 \end{bmatrix} \right) \quad (5.2.3)$$

Sendo r e λ variáveis latentes não observadas. Por isso, a fim de estimar a distribuição conjunta de tais variáveis, é necessário definir a relação destas com as observadas. Primeiramente, admite-se que o número de acionamentos do seguro realizado pelo indivíduo i resulta de uma distribuição de Poisson, tal que:

$$acionamentos_i \sim Poisson(\lambda_i, t_i) \quad (5.2.4)$$

Em que t_i é tempo de contrato do seguro. Já o coeficiente de aversão ao risco absoluto relaciona-se com a escolha da franquia - reduzida ou normal - por meio do modelo teórico, de modo que o indivíduo escolherá o contrato com maior cobertura, e portanto menor

franquia, se seu coeficiente de aversão ao risco absoluto for maior que o limite estipulado pela abordagem teórica. Isto é:

$$\begin{aligned} Pr(cobertura_i = 1) &= Pr\left(r_i > \frac{\frac{\Delta p_i}{\lambda_i \Delta d_i} - 1}{\bar{d}_i}\right) \\ &= Pr\left(\exp(x_i' \gamma + \nu_i) > \frac{\frac{\Delta p_i}{\exp(x_i' \beta + \epsilon_i) \Delta d_i} - 1}{\bar{d}_i}\right) \end{aligned} \quad (5.2.5)$$

Da equação 5.2.5 percebe-se que o indivíduo escolhe o contrato com maior cobertura - franquia reduzida - somente se $r_i > r_i^*(\lambda_i)$, tal que $r_i^*(.)$ é definido pela equação 5.1.6. Como há heterogeneidade não observada em λ_i , isto é $\varepsilon_i \neq 0$, este modelo permite que fatores não observados pela seguradora expliquem o risco individual. Em outras palavras, esta abordagem admite seleção adversa. Caso contrário, a equação 5.2.5 se reduziria à um Probit, já que o risco seria perfeitamente estimado por meio das variáveis observadas, $\hat{\lambda}(x_i)$.

A função de verossimilhança do modelo descrito nesta seção é representada por:

$$L(\text{acionamento}_i, \text{cobertura}_i | \theta) = Pr(\text{acionamento}_i, \text{cobertura}_i | \lambda_i, r_i) Pr(\lambda_i, r_i | \theta)$$

Em que θ é vetor de parâmetros a ser estimado. Contudo, a estimação via máxima verossimilhança não é trivial. Devido à existência de heterogeneidade não observada no risco e, também, na aversão ao risco, a estimação torna-se um processo computacionalmente penoso, já que é necessário realizar a integração em relação às duas dimensões. Em contrapartida, a amostragem de Gibbs, que utiliza método de Monte Carlo via Cadeias de Markov (MCMC), é muito atrativa para este caso. Cohen e Einav (2007) argumentam que esta metodologia é ideal para este caso, pois permite o aumento dos dados das variáveis latentes (TANNER; WONG, 1987). Sendo assim, pode-se simular (λ_i, r_i) e, em seguida, tratar tais simulações como parte dos dados. Além disso, a hipótese de log-normalidade implica que $F(\ln(\lambda_i) | r_i)$ e $F(\ln(r_i) | \lambda_i)$ seguem uma distribuição normal, o que colabora para diminuição do esforço computacional.

A metodologia da amostragem de Gibbs está descrita no apêndice A. Sua intuição básica é regredir as equações 5.2.1 e 5.2.2 condicional à λ_i e r_i para cada indivíduo. Para obter aleatoriamente observações sobre (λ_i, r_i) , realiza-se diversas iterações. Condicional à λ_i , a distribuição a *posteriori* de $\ln(r_i)$ segue uma distribuição normal truncada, em que o ponto de truncagem depende do *menu* oferecido ao segurado e sua direção (se a distribuição está acima ou abaixo do ponto de truncagem) advém da escolha da franquia. Coletar uma amostra da distribuição a *posteriori* de $\ln(\lambda_i)$ condicional à r_i é mais complicado, pois existem dois pontos de truncagem. O primeiro originário da seleção adversa (similar ao r_i) e o segundo devido à hipótese da equação 5.2.4 sobre a distribuição do número de acionamentos realizados que traz uma informação adicional sobre o distribuição a *posteriori* de λ_i . Para obter uma amostra sobre a distribuição desconhecida de λ_i utilizar-se-á o "*sliced sampler*" (DAMLEN; WAKEFIELD; WALKER, 1999).

Os resultados apresentados neste trabalho são fruto de 100.000 iterações da amostragem de Gibbs. Como a estimativa da distribuição das variáveis latentes dependem de um chute inicial e converge após determinado número de iterações, as 10.000 primeiras amostragens - referentes ao total - foram descartadas.

6 Inferência Bayesiana

6.1 Introdução à Inferência Bayesiana

Em linhas gerais, Inferência Bayesiana é uma abordagem que consiste em atualizar crenças iniciais por meio das informações disponíveis. Seu uso é muito atrativo, pois requer apenas a aplicação de alguns princípios estatísticos simples. Seja um modelo econométrico caracterizado pela distribuição conjunta de todas as variáveis de interesse analisadas. Destas, as que podem ser conhecidas via amostragem são denotadas por um vetor y , T -dimensional, enquanto as desconhecidas compõem o vetor θ com dimensão K , tal que $\theta \in \Theta \subseteq \mathbb{R}^K$. Seja θ e y variáveis aleatórias contínuas, então, em termos de densidade, tem-se:

$$p(y, \theta) = p(\theta)p(y|\theta) = p(y)p(\theta|y) \quad (6.1.1)$$

Em que $p(\theta)$ é a *densidade (ou distribuição) à priori* e $p(\theta|y)$ a *densidade (ou distribuição) à posteriori*. Analisando $p(y|\theta)$ como uma função de θ para determinado y , qualquer função proporcional a ela é denominada função de verossimilhança. Pode-se denotar a função de verossimilhança como $L(\theta)$, de modo que $L(\theta) = p(y|\theta)$ ⁷. Além disso, $p(y)$ é definido como a distribuição marginal dos dados.

As definições acima relacionam-se com teorema de Bayes de maneira que a *densidade à posteriori* pode ser escrita como:

$$p(\theta|y) = \frac{L(\theta)p(\theta)}{p(y)} \quad (6.1.2)$$

A distribuição marginal dos dados, $p(y)$ não envolve θ , ou seja, atua apenas como um fator de proporcionalidade ⁸, e, por isso, torna-se dispensável para realização de inferência

⁷ Neste caso deve-se incluir a constante de integração para $y|\theta$ na descrição de função de verossimilhança

⁸ O fator de proporcionalidade garante que a integral da *densidade à posteriori* assuma valor unitário.

sobre θ . Com isso o teorema de Bayes⁹ é frequentemente reescrito como:

$$p(\theta|y) \propto L(\theta)p(\theta) \quad (6.1.3)$$

Desta última relação, percebe-se que a *distribuição à posteriori* é proporcional ao produto da função verossimilhança e da *distribuição à priori*. Esse contexto, evidencia que a *distribuição à posteriori* é a densidade de maior interesse na Inferência Bayesiana, pois representa as crenças atualizadas - após a utilização dos dados - sobre o vetor de interesse θ . Por fim, pode-se reduzir a método Bayesiano à um algoritmo de cinco passos:

1. Formular o modelo econômico como um conjunto de distribuições de probabilidade condicionais a diferentes valores do vetor de interesse $\theta \in \Theta$.
2. Organizar suas crenças sobre θ em uma distribuição de probabilidade *à priori* sobre o mesmo vetor de parâmetros.
3. Coletar os dados e introduzi-los na família de distribuições da etapa 1.
4. Usar o teorema de Bayes para calcular as novas crenças sobre θ .
5. Criticar o modelo.

Na maioria das vezes, o interesse na *posteriori* recai sobre os diferentes componentes do vetor de parâmetros θ . Mas, para obter as distribuições marginais associadas à θ , é necessário integrar em relação aos demais parâmetros. O problema é que não existe uma distribuição posterior - e suas eventuais distribuições marginais - para todas as classes de pares de verossimilhança e priores. Nesses casos, pode-se realizar amostragens sobre a distribuição *a posteriori* a fim de inferir a respeito dos elementos de θ . A Lei dos Grandes Números garante que as estimativas irão convergir em probabilidade para o verdadeiro valor a medida que o número de amostragens tende para o infinito. Uma das formas de realizar amostragens sobre uma distribuição de probabilidade é via *Markov Chain Monte Carlo* (MCMC).

⁹ Para mais detalhes ver Koop, Poirier e Tobias (2007)

6.2 MCMC - Monte Carlo via Cadeia de Markov

Os métodos de Inferência Bayesiana consistem em estimar uma distribuição posterior a partir de uma função de verossimilhança e das distribuições a priori sobre os parâmetros de interesse, por meio do teorema de Bayes. Como nem sempre existe uma forma analítica para a distribuição *à posteriori*, é comum o uso de aproximação numérica para avaliá-la. Nesse sentido, o método MCMC se torna útil, pois consiste em realizar amostragens sobre uma distribuição estacionária (a posteriori de interesse) de uma cadeia de Markov simulada.

6.2.1 Cadeia de Markov e Matriz de Transição

Uma cadeia de Markov é uma sequência de variáveis aleatórias $(X_0, X_1, X_2, \dots)$, tal que a probabilidade de distribuição de qualquer uma, condicional às realizações anteriores, depende apenas da realização imediatamente anterior. Isto é, se χ é o espaço amostral de $\{X_t\}$ e t é o conjunto de índices, então:

$$P(X_{t+1}|x_0, x_1, \dots, x_t) = P(X_{t+1}|x_t) \quad (6.2.1)$$

Para todo t . O valor de X_t denomina-se *estado da cadeia* em t e a probabilidade associada à expressão 6.2.1 é a *probabilidade de transição*. A Cadeia é dita homogênea - daqui em diante será assumida esta hipótese - se a probabilidade de transição não depender do período, t .

Na maioria das aplicações do MCMC em inferência Bayesiana, o estado de X é caracterizado por um vetor de variáveis aleatórias (vetor de parâmetros), cujo cada componente representa o elemento correspondente de X . É evidente que, caso haja homogeneidade, uma Cadeia de Markov pode ser completamente descrita por um estado inicial e por uma regra - denominada Kernel de transição - sobre como os estados da cadeia se movem ao longo do tempo. Para caso discreto, essa regra é caracterizada pelo conjunto de probabilidades de transição que compõe a *Matriz de Transição*, ou *Matriz Estocástica*, K :

$$K(x, y) = P(X_{t+1} = y | X_t = x), \quad x, y \in \chi \quad (6.2.2)$$

Cada linha da matriz K , ou melhor cada x , fornece uma função de probabilidade (para y), na qual define-se $p_{ij} = P(X_{t+1} = j | x_t = i)$ como o elemento (i, j) da mesma matriz.

$$p_{ij} \geq 0$$

$$\sum_i p_{ij} = 1$$

Com isso, percebe-se que todos os elementos da matriz de transição são não-negativos e as linhas - por serem distribuições de probabilidade - somam um.

6.2.2 Distribuição de Probabilidade

Seja p_{t+1} a distribuição de probabilidade de X_{t+1} , então é possível defini-la em termos da matriz de transição e da correspondente distribuição de X_t . No caso discreto para M estados:

$$P(X_{t+1} = j) = \sum_{i=1}^M P(X_t = i) P(X_{t+1} = j | X_t = i), \quad j = 1, 2, \dots, M$$

Isto é,

$$p_{t+1}(j) = \sum_{i=1}^M p_t(i) K(i, j)$$

Em que $p_{t+1}(j)$ [$p_t(i)$] denota a probabilidade marginal de que a cadeia, no período $t+1$ [t], esteja no estado j [i]. Como dito anteriormente, $K(i, j)$ representa a probabilidade de $X_{t+1} = j$ dado que $X_t = i$, (o elemento p_{ij} da matriz de transição). Em termos matriciais para uma cadeia discreta e finita:

$$p_{t+1} = p_t \cdot K \quad (6.2.3)$$

Tal que K é a matriz de transição de dimensão $M \times M$, p_t e p_{t+1} são vetores linha que descrevem a probabilidade de X_t e X_{t+1} , respectivamente. Pós-multiplicando os dois lados

da expressão 6.2.3 pela matriz de transição m vezes e realizando um exercício recursivo observa-se:

$$p_{t+m} = p_t \cdot K^m$$

Nesse caso, p_{t+m} representa a probabilidade da cadeia estar no estado j no período $t + m$, enquanto K^m denomina-se *matriz de transição em m passos*¹⁰. Sendo assim, pode-se descrever o elemento $p_{ij}(m) = p(X_{t+m}|X_m = i)$ como a probabilidade de sair do estado i para o estado j em m passos. Portanto, admitindo que p_t é a distribuição inicial ($t = 0$) nota-se que simular uma Cadeia de Markov é necessário apenas conhecer o kernel K e a distribuição inicial, p_0 .

6.2.3 Distribuição Estacionária e Convergência

Diante da expressão 6.2.3 é natural se perguntar sobre a existência de uma distribuição de probabilidade de modo que $p_t = p_{t+1} = p$, ou seja, uma vez que a cadeia atinge a distribuição p ela se torna invariante. Nesse caso, p denomina-se *distribuição estacionária*. Matematicamente temos:

$$p = p \cdot K$$

ou em termos contínuos:

$$p(y) = \int_{\chi} K(x, y)p(x)dx \quad (6.2.4)$$

A distribuição estacionária é de grande interesse para o algoritmo MCMC, já que o método realiza amostragens de uma função de probabilidade posterior estacionária. Contudo, é necessário assegurar a convergência da cadeia para a distribuição estacionária, independente da função de probabilidade inicial. Seja a cadeia discreta, $p_t = p_0 \cdot K^t$, a medida que p_t se aproxima de um vetor constante, K^t se aproxima de uma matrix com

¹⁰ Este resultado é obtido a partir do teorema das equações de *Chapman-Komolgorov*. Ver Wasserman (2004)

linhas idênticas, correspondentes às distribuições estacionárias. Pode-se definir, então, convergência como o comportamento limite da matriz de transição em t passos, tal que:

$$\lim_{t \rightarrow \infty} p_{ij}(t) = p_j$$

Para todo i e j em uma cadeia finita com distribuição estacionária p e independente do função de probabilidade inicial p_0 ¹¹.

6.2.4 Definições

A seguir são apresentadas algumas definições¹² importantes para o método MCMC:

1. O estado i acessa j ($i \rightarrow j$) se $p_{ij}(n) > 0$ para algum n . Se $i \rightarrow j$ e $j \rightarrow i$, então i e j são *comunicáveis* ($i \leftrightarrow j$). Um conjunto de estados comunicáveis denomina-se *classe de comunicação*. Uma cadeia é *irredutível* se existe apenas uma classe de comunicação.
2. O *tempo de recorrência* para o estado i é:

$$T_i = \inf\{n \geq 1; X_n = i\}$$

Ou seja, representa o tempo transcorrido para que uma cadeia no estado i retorne ao mesmo estado. Logo, $T_i \geq 1$.

3. Um estado é *recorrente* (ou *persistente*) se

$$P(X_n = i | X_0 = i) = 1$$

para qualquer $n \geq 1$. Ou ainda,

$$P_i(T_i < \infty)$$

Caso contrário, o estado é dito *transiente*.

¹¹ Para detalhes ver Lancaster (2004).

¹² Ver Wasserman (2004)

4. Um estado i é aperiódico se $p_{ii}(n) > 0$ para todo n suficientemente grande. Uma cadeia é aperiódica quando todos os seus estados são aperiódicos.
5. Tempo médio de recorrência de um estado recorrente i é dado por

$$m_i = E(T_{ii}) = \sum_n n f_{ii}(n)$$

Em que $f_{ij}(n) = P(X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j | X_0 = i)$. Um estado recorrente é dito nulo se $m_i = \infty$, caso contrário o estado é não nulo.

6. Um estado é *ergódico* se ele for recorrente, não nulo e aperiódico. uma cadeia é ergódica se todos os estados forem ergódicos.

Em suma, o método MCMC consiste em construir uma Cadeia de Markov, na qual p , a distribuição de interesse de uma variável aleatória X , seja estacionária e única. Sendo assim, as realizações da distribuição estacionária (posteriori) resultante de uma Cadeia de Markov - com n passos e uma distribuição inicial¹³ (priori) - devem convergir em probabilidade para a distribuição de interesse p da variável aleatória X . Este resultado é evidenciado pelo teorema: ¹⁴

- Seja uma Cadeia de Markov ergódica e irredutível que converge para uma distribuição estacionária e única, p . Se g é uma função limitada, então, com probabilidade um,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^n g(X_n) \rightarrow E_p(g) \equiv \sum_j g(j) p_j$$

6.3 Algoritmo MCMC

Como dito anteriormente o MCMC trata-se de um método numérico amplamente utilizado devido à dificuldade em encontrar expressões analíticas nos cálculos Bayesianos.

¹³ Definida de forma aleatória ou determinística

¹⁴ Conhecido como Lei dos Grandes Números da Cadeia de Markov. Para detalhes ver Wasserman (2004)

Sua ideia fundamental consiste em simular uma cadeia cuja a distribuição limite é única e estacionária. Desse forma, o problema se resume a encontrar uma matriz de transição correspondente à essa distribuição de interesse.

6.3.1 Amostragem de *Gibbs*

Segundo Lancaster (2004), a Amostragem de *Gibbs* é um método geral que permite encontrar a Matriz de Transição cuja distribuição de interesse é multivariada e estacionária. Nesse caso, pretende-se realizar amostragens sobre, por exemplo, a distribuição conjunta $p(y_1, y_2)$, em que y_j pode representar um vetor. Assim como na expressão 6.2.2, é possível definir o kernel de transição para o caso multivariado:

$$K(x, y) = K(x_1, x_2, y_1, y_2)$$

Tal que, quando a cadeia encontra-se no estado (x_1, x_2) , K fornece a probabilidade de passar para o estado (y_1, y_2) . Para isto, é necessário escolher um kernel K , de forma que a distribuição estacionária da cadeia seja a distribuição posterior $p(y_1, y_2)$. Uma forma simples e intuitiva de simular esta cadeia consiste em realizar amostragens sobre as distribuição condicionais $p_{Y_1|Y_2}(y_1|y_2)$. Nesse caso, Y_1 e Y_2 representam as variáveis aleatórias, enquanto y_1 e y_2 são as realizações, de forma que $x = (x_1, x_2)$ captam a realizações precedentes. Na prática, determina-se o número de iterações que serão realizadas, T , e, em seguida, aplica-se o algoritmo repetidamente. A Amostragem de *Gibbs* consiste em:

1. Obter uma amostra y_1 a partir da distribuição condicional de Y_1 dado $Y_2 = x_2$, $p_{Y_1|Y_2}(y_1|x_2)$.
2. Utilizando a realização de y_1 , obter uma amostra y_2 a partir da distribuição condicional de Y_2 dado $Y_1 = y_1$, $p_{Y_2|Y_1}(y_2|y_1)$.
3. Repetir os estágios 1 e 2 por T vezes, atualizando as realizações.

Apesar do exemplo acima ilustrar o caso bivariado, o algoritmo pode ser generalizado para mais dimensões. Na verdade, esta é uma característica importante do método, já que

a função de probabilidade condicional $p_{Y_i|Y_j}(y_i|y_j, i \neq j)$ é a única densidade utilizada para simular a cadeia, então, permite-se que, mesmo em um problema com várias dimensões, todas as simulações sejam univariadas. O kernel de transição da Amostragem de *Gibbs* é definido como:

$$\mathbf{P}(x, y) = p_{Y_1|Y_2}(y_1|x_2)p_{Y_2|Y_1}(y_2|y_1) \quad (6.3.1)$$

A intuição do método é evidente: para se mover ao longo da cadeia deve-se realizar amostragens sobre as distribuições condicionais das variáveis de interesse. Para verificar se 6.3.1 tem $p(y_1, y_2)$ como distribuição estacionária, basta substituir na expressão 6.2.4:

$$\begin{aligned} \int \mathbf{P}(x, y)p(x)dx &= \int p_{Y_1|Y_2}(y_1|x_2)p_{Y_2|Y_1}(y_2|y_1)p_{Y_1|Y_2}(x_1, x_2)dx_1dx_2 \\ &= \int p_{Y_1|Y_2}(y_1|x_2)p_{Y_2|Y_1}(y_2|y_1)p_{Y_2}(X_2)dx_2 \\ &= p_{Y_2|Y_1}(y_2|y_1) \int p_{Y_1|Y_2}(y_1|x_2)dx_2 \\ &= p_{Y_2|Y_1}(y_2|y_1)p_{Y_1}(y_1) = p_{Y_1, Y_2}(y_1) = p(y) \end{aligned}$$

Logo, realizar amostragens sobre distribuições condicionais sequencialmente proporciona um kernel de uma cadeia com distribuição estacionária dada pela correspondente distribuição conjunta. Além disso, caso a cadeia seja irredutível e ergódica haverá convergência para esta distribuição estacionária e única. A Amostragem de Gibbs é conhecida por ser um caso especial do algoritmo Metropolis-Hastings, ainda que com algumas características distintas.¹⁵

6.3.2 Algoritmo *Metropolis-Hastings*

O Algoritmo *Metropolis-Hastings* é um método MCMC específico que também consiste em realizar amostragens sobre uma distribuição de posterior de interesse. Seja $q(y|x)$ uma distribuição arbitrária e conhecida, de modo que se consiga amostrar sobre ela, então, a densidade condicional $q(y|x)$ denomina-se distribuição proposta. A fim de criar uma sequência de observações $X_1, X_2, \dots$, determina-se um número de iterações, T , e aplica-se o seguinte algoritmo repetidamente:

¹⁵ Para detalhes ver Robert e Casella (2013).

1. Definir um valor inicial x_0 arbitrariamente. Suponha que já foi gerada uma sequência $X_0, X_1, \dots, X_i$. Para obter X_{i+1} , deve-se:
2. gerar um candidato a partir da distribuição proposta $Y \sim q(y|X_i)$.
3. Avaliar $r \equiv r(X_i, Y)$, tal que:

$$r(x, y) = \min \left\{ \frac{f(y) q(x|y)}{f(x) q(y|x)} \right\}$$

4. Gerar uma variável aleatória, u , uniformemente distribuída entre 0 e 1, $U \sim (0, 1)$. Se $U < r$ defina $X_{i+1} = Y$, caso contrário defina $X_{i+1} = X_i$.
5. Repetir os estágios por T vezes, atualizando as realizações.

Se a densidade proposta for simétrica, $q(y|x) = q(x|y)$, então, r é simplificada para:

$$r = \min \left\{ \frac{f(y)}{f(x_i)}, 1 \right\}$$

Nesse caso, o algoritmo é denominado algoritmo de *Metropolis*. Assim como a amostragem de *Gibbs* o algoritmo de *Metropolis* também converge para uma distribuição estacionária.¹⁶

6.3.3 Sliced Sampler

A Amostragem de *Gibbs* permite obter amostras de distribuições multivariadas, desde que se conheça a forma analítica das funções de probabilidade condicionais. Quando a distribuição (condicional) posterior de interesse assume uma forma pouco familiar, o método denominado *Sliced Sample* torna-se útil. Neal (2003) fornece a intuição de como implementar o algoritmo de maneira clara. Seja $f(x)$ a função densidade sob a qual se deseja retirar amostras, tal que $x \in X \subseteq \mathfrak{R}$. A ideia consiste em introduzir um variável real auxiliar, y , e definir uma distribuição conjunta uniforme na região $U = \{(x, y) : 0 < y < f(x)\}$ abaixo da curva definida por $f(x)$. A distribuição conjunta de (x, y) é definida por:

¹⁶ Para provas ver Robert e Casella (2013).

$$p(x, y) = \begin{cases} 1/Z, & \text{se } 0 < y < f(x) \\ 0, & \text{c.c.} \end{cases}$$

Em que $Z = \int f(x)dx$. Assim, a distribuição marginal de x é expressa por:

$$p(x) = \int_0^{f(x)} (1/Z)dy = f(x)/Z$$

Portanto, a amostragem de x pode ser feita por meio da amostragem conjunta de (x, y) e, em seguida, ignorando y . A fim de gerar amostras independentes dos pontos (x, y) , que estão uniformemente distribuídos em U , pode-se, por exemplo, realizar a amostragem de *Gibbs*. Para isto, deve-se obter, alternativamente, uma amostra da distribuição de y dado x , que é uniforme no intervalo $(0, f(x))$, e uma amostra da distribuição de x dado y , que é uniforme na região $S = \{x : y < f(x)\}$. Ainda sim, gerar amostras independentes uniformemente distribuídas em S pode ser complicado e, por isso, que Neal (2003) sugere o seguinte processo:

1. Amostrar um valor de y uniformemente distribuído no intervalo $(0, f(x_0))$, em que x_0 representa o valor inicial. Desta forma, define-se a 'fatia' (*slice*) horizontal: $S = \{x : y < f(x)\}$.
2. Encontrar um intervalo, $I = (L, R)$, em torno de x_0 que contenha a maior parte da 'fatia'.
3. Por fim, deve-se amostrar um novo ponto x_1 , da parte da 'fatia' pertencente ao intervalo I .

Os passos 2 e 3 podem ser implementados de diversas maneiras, mas vale ressaltar um método para cada estágio. Em 2, o procedimento *stepping out* consiste em posicionar um intervalo com banda w randomicamente em torno de x_0 e, então, ampliá-los com incrementos de tamanho w até que ambos os extremos do intervalo saiam da "fatia". Em 3, o método *shrinkage* realiza a amostragem de x_1 a partir de I . O valor é aceito se $y < f(x_1)$, caso contrário o intervalo I é reduzido para $L = x_1$ (se $x_1 < x_0$) ou $R = x_1$ (se $x_1 > x_0$).

7 Resultados

7.1 Estimação - Forma Reduzida

Uma idéia inicial a respeito do nível de aversão ao risco absoluto pode ser obtida por meio dos valores médios não condicionais de Δp , Δd , λ e $\bar{d}$ ¹⁷. Substituindo esses valores na equação 5.1.6 encontra-se um coeficiente de $2,3 \cdot 10^{-3}$, que pode ser interpretado como ponto de indiferença médio, de modo que cerca de 14% dos segurados possuem um coeficiente de aversão ao risco mais elevado.¹⁸

A tabela 3 representa uma análise na forma reduzida entre as variáveis observadas e as latentes. A coluna 1 reporta valores de uma regressão de poisson do número de acionamentos sobre as características observadas. Esta equação é análoga a equação do risco do modelo principal. Os resultados sugerem que os homens possuem menor probabilidade de se envolver em um acidente, enquanto os jovens - com idade entre 18 e 25 anos - são os mais propensos à acionar o seguro.

A coluna 2 apresenta estimativas de probit em que a variável dependente assume valor 1 caso o segurado escolha a franquia reduzida. A variável $\ln \left(\left[\Delta p_i / \left(\hat{\lambda}(x_i) \Delta d_i \right) - 1 \right] / \bar{d}_i \right)$ foi usada como regressor normalizada para -1 .¹⁹ A estimativa de $\hat{\lambda}(x_i)$ foi obtida mediante o computo do valor predito da regressão de poisson da coluna 1. Diante disso, percebe-se que este modelo não permite heterogeneidade não observada no risco, isto é, assemelha-se a equação da aversão ao risco do modelo principal, caso a hipótese restritiva de ausência de heterogeneidade não observada seja verdadeira. Dentre os 112630 segurados analisados 83 obtiveram $\hat{\lambda}(x_i) > (\Delta p_i / \Delta d_i)$, o que implicaria na escolha da franquia reduzida para qualquer nível de aversão ao risco. Contudo, apenas 7 dos 83 indivíduos escolheram maior cobertura, evidenciando a presença de seleção adversa e, consequentemente, heterogeneidade não observada no risco.

¹⁷ O valor não condicional de λ foi computado por máxima verossimilhança, utilizando os dados de acionamentos e duração dos contratos

¹⁸ Percentual dos segurados que escolheram franquia reduzida.

¹⁹ O próprio nível da variável é uma normalização, logo, aplicou-se apenas o sinal negativo.

Os resultados da tabela 2 vão ao encontro das estimação do modelo principal, destacando-se as mulheres como mais avessas ao risco e os jovens entre 18 e 25 anos como os menos avessos. A coluna 2 também reporta valores dos efeitos marginais dos regressores sobre probabilidade de escolher franquia reduzida, assim como a coluna 3, que, entretanto, não utiliza $\ln r_i$ como regressor de seu probit correspondente, não possuindo assim uma interpretação estrutural.²⁰

Tabela 3: Ausência de Heterogeneidade no Risco

		Regressão Poisson			Regressão Probit ^a			
		Var. Dep: Acionamentos			Var. Dep: 1 se franquia reduzida			
Variáveis		Coef.	D.P.	IRR	Coef.	D.P.	dP/dX	dP/dX
	Constante	-2,548***	0,736	0,078	-5,808***	0,233		
Demográficas:	Gênero	-0,097***	0,032	0,907	-0,091***	0,010	-0,018***	-0,043***
	Idade 18-25	-	-	-	-	-	-	-
	Idade 26-35	-0,369***	0,048	0,691	0,316***	0,019	0,068***	0,051***
	Idade 36-45	-0,552***	0,052	0,575	0,412***	0,020	0,093***	0,056***
	Idade 46-55	-0,544***	0,054	0,580	0,416***	0,021	0,095***	0,060***
	Idade 55-65	-0,798***	0,067	0,450	0,670***	0,022	0,174***	0,095***
	Idade >65	-0,810***	0,093	0,445	0,643***	0,027	0,172***	0,100***
Características do Carro:	Log (IS Casco)	0,056	0,070	1,057	0,080***	0,021	0,016***	0,029***
	Modelo Ano <2008	-0,136**	0,058	0,873	-0,061***	0,018	-0,012***	-0,019***
	Modelo Ano 2008-2010	0,046	0,038	1,047	-0,074***	0,012	-0,015***	-0,014***
	Modelo Ano >2010	-	-	-	-	-	-	-
	Potência Motor	0,039	0,059	1,040	-0,124***	0,018	-0,025***	-0,003
Pseudo R		0,011			0,081			0,032
Obs.		112630			112547			112630
Log-Verossimilhança		-17117,7			-41771,4			-44056,6

** Significante à 5%.

*** Significante à 1%.

^a Foram realizados duas regressões diferentes. O primeiro probit (coluna 2) utilizou um regressor adicional, $\ln([\Delta p_i / (\hat{\lambda}(x_i) \Delta d_i) - 1] / \bar{d}_i)$ (com $\hat{\lambda}$ predito na coluna 1) que possibilita uma interpretação estrutural e a comparação entre estes coeficientes e o do modelo principal. Contudo, vale lembrar que esta regressão não permite heterogeneidade não observada no risco. A coluna 3 reporta os efeitos marginais do probit sem o regressor adicional, possibilitado a comparação entre os resultados das duas colunas.

7.2 Estimação - Inferência Bayesiana

Para a estimação do modelo realizou-se a normalização das variáveis de controle X , por isso a interpretação deve ser realizada em termos de desvios-padrão. Esta abordagem

²⁰ A coluna 3 utiliza todos os segurados da base, mas admite que a escolha da cobertura não depende de preços, risco e aversão ao risco.

torna a escala dos regressores irrelevante, o que permite alta comparabilidade entre os coeficientes.

Os resultados do modelo estão representados pela tabela 4. Observando apenas as variáveis da equação do risco percebe-se que a probabilidade de acidente é menor para o sexo masculino (cerca de 8% menor). Em relação a faixa etária, vale destacar não só que os segurados com idade entre 18 e 25 anos têm um risco maior de acidente, mas também que a probabilidade de acidente diminui com a idade, isto é, faixas etárias mais avançadas são menos arriscadas. Este resultado pode estar associado ao nível de experiência dos motoristas. As demais variáveis de controle - importância segurada, que atua como uma *proxy* para o valor do carro, ano do modelo do veículo e potência do motor - não apresentam efeito significativo na probabilidade de acidente. As *dummies* relacionadas aos três primeiros dígitos do endereço postal explicam o risco, pois a região por onde o veículo circula interfere na probabilidade de acidente. Contudo, estas variáveis não foram utilizadas para explicar a equação de aversão ao risco. Esses resultados não foram reportados na tabela.²¹

Quando o foco é a equação da aversão ao risco nota-se alguns resultados interessantes. O coeficiente do gênero sugere o coeficiente de aversão ao risco absoluto dos homens é 24 % menor que os das mulheres. Este resultado vai ao encontro do obtido por Cohen e Einav (2007) que encontraram um coeficiente 20% maior para as mulheres. As *dummies* de faixa etária revelam um resultado importante: a faixa etária mais avançada é a mais avessa ao risco, em contrapartida, os jovens entre 18 e 25 anos são os menos avessos. A variável importância segurada sugere que quanto maior o valor do veículo mais avesso ao risco o indivíduo.

A cada iteração da amostragem de *Gibbs*, computou-se a média e o desvio padrão das amostragens aleatórias de λ_i e r_i , bem como a correlação entre estas variáveis. A tabela 4 reporta as médias e os desvios-padrão das quantidades computadas a cada iteração da amostragem de *Gibbs*. Dessa forma, estas estimativas não são condicionais às características observáveis dos segurados, ou seja, não é possível obtê-las diretamente dos parâmetros. Observa-se que a média do coeficiente de aversão ao risco absoluto é de 0,0025 e a mediana

²¹ Os resultados do CEP encontram-se no apêndice C.

Tabela 4: Resultados - Estimação Bayesiana

Variáveis		$\ln \lambda$	$\ln r$	Matriz Var-Covar	
Demográficas:	Constante	-3,233 (0,0211) ***	-8,3726 (0,0649) ***	σ_λ	0,8113 (0,0163)
	Gênero	-0,0783 (0,0279) ***	-0,2431 (0,0584) ***	σ_r	2,1793 (0,0542)
	Idade 18-25	-	-	ρ	-0,3900 (0,0305)
	Idade 26-35	-0,2070 (0,0420) ***	0,3788 (0,0987) ***	Estatísticas não condicionais	
	Idade 36-45	-0,3021 (0,0488) ***	0,4205 (0,1114) ***	Média λ	0,0563 (0,0008)
	Idade 46-55	-0,3534 (0,0484) ***	0,5500 (0,1078) ***	Mediana λ	0,0394 (0,0008)
	Idade 55-65	-0,5159 (0,0611) ***	0,9689 (0,1231) ***	D.P. λ	0,0569 (0,0015)
	Idade >65	-0,6254 (0,0830) ***	1,0932 (0,1561) ***	Média r	0,0025 (0,0001)
Características do Carro:	Log (IS Casco)	-0,0387 (0,0673)	0,5301 (0,1400) ***	Mediana r	0,0002 ($1,49 \cdot 10^{-5}$)
	Modelo Ano <2008	-0,1491 (0,0478) ***	-0,0343 (0,0979)	D.P. r	0,0146 (0,0014)
	Modelo Ano 2008-2010	-0,0395 (0,0327)	0,0231 (0,0671)	Corr (λ, r)	-0,0836 (0,0046)
	Modelo Ano >2010	-	-		
	Potência Motor	-0,0236 (0,0509)	-0,0930 (0,1019)		

*** Significante à 1%.
Nota: 112630 observações.
Fonte: elaboração própria.
Nota: Desvios-padrão obtidos por meio de amostragens da distribuição posterior nos parênteses.

muito próxima de zero 0,0002.

Além disso, observa-se uma elevada heterogeneidade não observada de aversão ao risco (σ_r) e do risco (σ_λ), mas a primeira é muito maior que a segunda. Este resultado corrobora com o obtido por Cohen e Einav (2007). Os autores argumentam que este sub-produto da estimação indica que a heterogeneidade não observada da aversão ao risco é uma fonte de seleção mais importante que a do risco.

Por fim, a tabela 4 também aponta uma correlação negativa de $-0,39$ entre a aversão ao risco não observada e o risco não observado, ρ , e uma correlação não condicional de $-0,083$ entre λ e r . A princípio esse resultado é intuitivo, pois é natural supor que indivíduos mais avessos ao risco terão atitudes mais cautelosas e, portanto, correm menos riscos. Finkelstein e McGarry (2006) observaram essa mesma evidência analisando o mercado de seguros de saúde americano. Entretanto, os próprios Cohen e Einav (2007) enumeram motivos para que o sinal da correlação seja ambíguo.

Primeiro, no mercado de seguros de automóveis, o risco dos agentes não depende apenas de atitudes precaucionárias, ou de fatores inerentes ao segurado, tal como habilidade inata em dirigir, mas também da interação dos hábitos de condução dos motorista. Segundo, existem fatores não observados que podem estar positivamente correlacionado às duas dimensões (risco e aversão). Por exemplo, se indivíduos mais avessos ao risco dirigem com mais frequência, então, possuem maior risco. Por fim, a correlação não condicional entre as variáveis é sensível a hipótese realizada sobre a distribuição dos acionamentos e

está sujeita aos coeficientes das covariadas. Isto é, os coeficientes de uma mesma covariada podem afetar o risco e a aversão ao risco na mesma direção, logo, mesmo que ρ seja negativo a correlação (λ, r) pode não ser.

De qualquer forma, os resultados apresentados neste trabalho apontam para uma correlação negativa tanto para fatores não observados, ρ , quanto para as duas variáveis de interesse (λ, r) . Em relação a dispersão do coeficiente de aversão ao risco absoluto, a tabela 5 reporta a interpretação dos diferentes percentis da estimação não condicional da aversão ao risco. A interpretação foi obtida para uma função utilidade do tipo quadrática, $\{u(w) = w - bw^2\}$, o que comporta a hipótese de que a expansão de Taylor de segunda ordem - utilizada no modelo teórico - é exata para cada indivíduo.

Tabela 5: Estimativas de Aversão ao risco

Especificação	Aversão ao risco absoluta ($R\$$)	Interpretação ($R\$$) ^a
Estimação pelas médias	$2,3 \cdot 10^{-3}$	80,54
Inferência Bayesiana:		
Média	$2,5 \cdot 10^{-3}$	79,58
25° percentil	$1,1 \cdot 10^{-4}$	98,83
Mediana	$4,2 \cdot 10^{-4}$	95,95
75° percentil	$1,5 \cdot 10^{-3}$	86,83
90° percentil	$4,7 \cdot 10^{-3}$	65,94
95° percentil	$9,3 \cdot 10^{-3}$	44,08
Cohen e Einav:	Aversão ao risco absoluta ($US\$$)	Interpretação ($US\$$)
Média	$6,7 \cdot 10^{-3}$	56,05
25° percentil	$2,3 \cdot 10^{-6}$	99,98
Mediana	$2,6 \cdot 10^{-5}$	99,74
75° percentil	$2,9 \cdot 10^{-4}$	97,14
90° percentil	$2,7 \cdot 10^{-3}$	78,34
95° percentil	$9,9 \cdot 10^{-3}$	49,73

^a Para interpretar a aversão ao risco absoluta (ARA) calculou-se $\left\{x : u(w) = \frac{1}{2}u(w + 100) + \frac{1}{2}u(w - x)\right\}$. Logo, para os diferentes níveis de ARA, reportou-se valores de x , que torna o indivíduo indiferente entre participar de uma loteria em que se tem 50% de chance de ganhar $R\$100$ e 50% de perder x . Para realizar a interpretação assumiu-se que o indivíduo possui uma utilidade quadrática do tipo $\{u(w) = w - bw^2\}$. Além disso, é importante notar que a interpretação não depende do nível de riqueza do indivíduo, pois a análise parte do coeficiente absoluto de aversão ao risco.

Sendo assim, um segurado com o coeficiente de aversão ao risco absoluto médio será indiferente entre participar ou não de uma loteria em que ele tem 50% de chance de ganhar $R\$100,00$ e 50% de chance de perder $R\$79,58$. Cohen e Einav (2007) obtiveram que o indivíduo médio é indiferente em participar (ou não) de uma loteria 50-50 em que ele ganha $US\$100,00$ ou perde $US\$56,05$. Apesar da comparação não ser imediata devido à diferença na unidade de medida ($US\$$ e $R\$$), pode-se observar que em ambos os resultados o indivíduo mediano é quase neutro ao risco, ainda que o segurado médio possua considerável nível de aversão em loterias dessa magnitude.

Figura 1: Estimativa da Distribuição Conjunta - Risco e Aversão

Por fim, estimou-se, a partir dos valores das médias e desvios-padrão, a densidade conjunta das duas variáveis latentes. Para tal, a distribuição conjunta das variáveis foi simulada de acordo com os resultados e, em seguida, realizou-se a estimação da densidade kernel representada na figura 1. Este resultado é interessante, pois sintetiza a dispersão das variáveis latentes estimadas com este trabalho, destacando-se pela concentração da distribuição conjunta em valores próximos à zero.

8 Conclusão

O presente trabalho aplicou a metodologia proposta por Cohen e Einav (2007) a fim de estimar a distribuição de aversão ao risco absoluto a partir de dados sobre seguros de automóveis da região metropolitana de São Paulo. As informações básicas utilizadas para realização dessa abordagem são: escolha da cobertura, conjunto de contratos oferecidos pela seguradora (par prêmio e franquia) e número de acionamentos realizados pelo segurado. Além disso, utilizou-se características individuais dos segurados como covariadas na equação do risco e da aversão ao risco.

Os resultados obtidos sugerem que a média da aversão ao risco é baixa (0,0025) e a mediana muito próxima de zero (0,0002). Além disso, observou-se elevada heterogeneidade não observada tanto no risco quanto na aversão ao risco. Tais estimativas são relativamente similares às obtidas por Cohen e Einav (2007), entretanto a correlação (negativa) entre a equação do risco e da aversão ao risco difere dos resultados encontrados pelos autores. O sinal desse coeficiente, como discutido anteriormente, é ambíguo.

Em relação às covariadas percebe-se que os homens são menos avessos ao risco que as mulheres e que coeficiente de aversão ao risco é maior para segurados maiores de 66 anos. Além disso, mulheres e jovens entre 18 e 25 possuem maior probabilidade de sinistro.

A interpretação dos resultados fica mais clara quando define-se uma forma funcional para a utilidade dos indivíduos. Assumiu-se, então, uma utilidade do tipo quadrática ($u(w) = w - bw^2$), pois sua terceira derivada assume valor zero. Esta hipótese foi realizada no modelo teórico e possibilitou uma relação na forma reduzida entre o risco e aversão ao risco. Os resultados mostram que um segurado com o coeficiente de aversão ao risco absoluto médio (mediano) será indiferente entre participar ou não de uma loteria em que ele tem 50% de chance de ganhar R\$100,00 e 50% de chance de perder R\$79,58 (R\$95,95).

Apesar dos valores obtidos, a extrapolação dos resultados não é tão imediata. Rabin e Thaler (2001) argumentaram que as escolhas na vida são realizadas em diferentes contextos e, por isso, podem depender de diferentes parâmetros da função utilidade. Nesse sentido, estudos recentes - já citados neste trabalho - visam avaliar as estabilidade das

preferências sobre o risco em diferentes mercados (BARSEGHYAN; PRINCE; TEITELBAUM, 2011; EINAV et al., 2010). Ainda sim, segundo a teoria clássica, as decisões sob incerteza dependem de um único parâmetro que permeia todas as escolhas durante a vida. Diante disso, é importante notar o mérito deste modelo, dado que o mesmo estima a aversão ao risco a partir do relevante mercado de seguros de automóveis. Sua aplicabilidade em diferentes países e regiões é o passo seguinte mais natural e principal fundamento desta pesquisa.

Referências

- BARSEGHIYAN, L. et al. The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, v. 103, n. 6, p. 2499–2529, 2013.
- BARSEGHIYAN, L.; PRINCE, J.; TEITELBAUM, J. C. Are risk preferences stable across contexts? evidence from insurance data. *American Economic Review*, v. 101, n. 2, p. 591–631, 2011.
- CHETTY, R. A new method of estimating risk aversion. *The American Economic Review*, JSTOR, v. 96, n. 5, p. 1821–1834, 2006.
- CHIAPPORI, P.-A. et al. Asymmetric information in insurance: General testable implications. *The RAND Journal of Economics*, Wiley Online Library, v. 37, n. 4, p. 783–798, 2006.
- CHIAPPORI, P.-A.; SALANIÉ, B. Testing for asymmetric information in insurance markets. *Journal of political Economy*, JSTOR, v. 108, n. 1, p. 56–78, 2000.
- COHEN, A.; EINAV, L. Estimating risk preferences from deductible choice. *THE AMERICAN ECONOMIC REVIEW*, 2007.
- CUTLER, D. M.; FINKELSTEIN, A.; MCGARRY, K. *Preference heterogeneity and insurance markets: Explaining a puzzle of insurance*. [S.l.], 2008.
- DAMLEN, P.; WAKEFIELD, J.; WALKER, S. Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 61, n. 2, p. 331–344, 1999.
- DEVROYE, L. Sample-based non-uniform random variate generation. In: ACM. *Proceedings of the 18th conference on Winter simulation*. [S.l.], 1986. p. 260–265.
- EINAV, L. et al. *How General Are Risk Preference? Choices under Uncertainty in Different Domains*. [S.l.], 2010.
- FINKELSTEIN, A.; MCGARRY, K. Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, Princeton, NJ: American Economic Association, 1911-, v. 96, n. 4, p. 938–958, 2006.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, n. 6, p. 721–741, 1984.
- JULLIEN, B.; SALANIÉ, B. Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy*, JSTOR, v. 108, n. 3, p. 503–530, 2000.
- KOOP, G.; POIRIER, D. J.; TOBIAS, J. L. *Bayesian econometric methods*. [S.l.]: Cambridge University Press, 2007.
- LANCASTER, T. *An introduction to modern Bayesian econometrics*. [S.l.]: Blackwell Oxford, 2004.

- LANDSBERGER, M.; MEILIJSON, I. A general model of insurance under adverse selection. *Economic Theory*, Springer, v. 14, n. 2, p. 331–352, 1999.
- NEAL, R. M. Slice sampling. *Annals of statistics*, JSTOR, p. 705–741, 2003.
- PARAVISINI, D.; RAPPOPORT, V.; RAVINA, E. *Risk aversion and wealth: Evidence from person-to-person lending portfolios*. [S.l.], 2010.
- RABIN, M.; THALER, R. H. Anomalies: risk aversion. *Journal of Economic perspectives*, JSTOR, p. 219–232, 2001.
- ROBERT, C.; CASELLA, G. *Monte Carlo statistical methods*. [S.l.]: Springer Science & Business Media, 2013.
- ROTHSCHILD, M.; STIGLITZ, J. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, 1976.
- SAHA, A. Risk preference estimation in the nonlinear mean standard deviation approach. *Economic Inquiry*, Wiley Online Library, v. 35, n. 4, p. 770–782, 1997.
- SMART, M. Competitive insurance markets with two unobservables. *International Economic Review*, Wiley Online Library, v. 41, n. 1, p. 153–170, 2000.
- SMITH, V. L.; WALKER, J. M. Rewards, experience and decision costs in first price auctions. *Economic Inquiry*, Wiley Online Library, v. 31, n. 2, p. 237–244, 1993.
- TANNER, M. A.; WONG, W. H. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, Taylor & Francis, v. 82, n. 398, p. 528–540, 1987.
- WASSERMAN, L. *All of statistics: a concise course in statistical inference*. [S.l.]: Springer, 2004.

A Descrição da Amostragem de *Gibbs* (GS)

Seguindo o exposto no artigo de Cohen e Einav (2007), este apêndice irá demonstrar a amostragem de *Gibbs*, o qual foi utilizado na estimação do modelo proposto neste presente trabalho. Esse método foi proposto por Geman e Geman (1984) e tornou-se popular entre os estatísticos após 1990, o qual teve impacto significativo no desenvolvimento e aplicações práticas da Estatística Bayesiana.

Uma das principais vantagens do GS é poder permitir o aumento de dados de variáveis latentes. No presente contexto, esta abordagem permite o aumento dos dados sobre aversão ao risco e ao risco de cada indivíduo, ou seja, $\{\lambda_i, r_i\}_{i=1}^n$ são tratados como parâmetros adicionais.

Portanto, o modelo pode ser escrito da seguinte maneira:

$$\ln \lambda_i = x_i' \beta + \varepsilon_i$$

$$\ln r_i = x_i' \gamma + v_i$$

$$choice_i = \begin{cases} 1 & \text{se } r_i > r_i^*(\lambda_i) \\ 0 & \text{se } r_i < r_i^*(\lambda_i) \end{cases}$$

$$Acionamentos_i \sim Poisson(\lambda_i, t_i)$$

$$\begin{pmatrix} \varepsilon_i \\ \nu_i \end{pmatrix} \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{bmatrix} \right)$$

$$\delta \equiv \begin{bmatrix} \beta \\ \gamma \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_r \\ \rho\sigma_\lambda\sigma_r & \sigma_r^2 \end{bmatrix}, X \equiv \begin{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \end{bmatrix}, y \equiv \begin{bmatrix} \lambda \\ r \end{bmatrix}, u_i \equiv \begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix}$$

O conjunto de parâmetros que pretende-se estimar uma distribuição a *posteriori* é dado por $\theta = \{\delta, \Sigma, \{u_i\}_{i=1}^n\}$. A distribuição a *priori* especifica que $\{\delta, \Sigma\}$ são independentes de

$\{u_i\}_{i=1}^n$; e $\{\delta, \Sigma\}$ tem uma distribuição a *priori* convencional difusa. Adota-se a distribuição a *priori* hierárquica para $\{u_i\}_{i=1}^n$:

$$\{u_i\}_{i=1}^n \mid \Sigma \sim N(0, \Sigma)$$

$$\Sigma^{-1} \sim Wishart_2(a, Q)$$

Então, condicional em todos os outros parâmetros, tem-se:

$$\Sigma^{-1} \mid \delta, \{u_i\}_{i=1}^n \sim Wishart_2\left(a + n - k, \left(Q^{-1} + \sum_i u_i u_i'\right)^{-1}\right)$$

$$\delta \mid \Sigma, \{u_i\}_{i=1}^n \sim N((X'X)^{-1}(X'y), \Sigma^{-1} \otimes (X'X)^{-1})$$

Para Σ^{-1} também é usada uma distribuição a *priori* convencional difusa de tal forma que $a = 0$ e $Q^{-1} = 0$.

A GS é menos trivial nos casos que envolvem a amostragem da distribuição condicional dos parâmetros aumentados, $\{u_i\}_{i=1}^n$. Todos os indivíduos são independentes entre si, de modo que, condicional aos outros parâmetros, não há dependência de dados aumentados dos demais indivíduos. Desta forma, precisa-se apenas descrever a probabilidade condicional de u_i .

Note que, condicional em δ , tem-se: $\varepsilon_i = \ln \lambda_i - x_i' \gamma$ e $v_i = \ln r_i - x_i' \beta$. Logo, pode-se apenas focar na distribuição a *posterior* de λ_i e r_i . Essas distribuições a *posterior* são:

$$Pr(r_i \mid \gamma, \beta, \Sigma, \lambda_i, data) \propto$$

$$\begin{cases} \phi\left[\ln r_i, x_i' \gamma + \rho \frac{\sigma_r}{\sigma_\lambda} (\ln \lambda_i - x_i' \beta), \sqrt{\sigma_r^2 (1 - \rho^2)}\right] & \text{se } escolha_i = I(r_i < r_i^*(\lambda_i)) \\ 0 & \text{se } escolha_i \neq I(r_i < r_i^*(\lambda_i)) \end{cases}$$

e

$$Pr(\lambda_i \mid \gamma, \beta, \Sigma, \lambda_i, data) \propto$$

$$\begin{cases} p(\lambda_i, aciona., t_i) \phi\left[\ln \lambda_i, x_i' \beta + \rho \frac{\sigma_\lambda}{\sigma_r} (\ln r_i - x_i' \gamma), \sqrt{\sigma_\lambda^2 (1 - \rho^2)}\right] & \text{se } escolha_i = I(r_i < r_i^*(\lambda_i)) \\ 0 & \text{se } escolha_i \neq I(r_i < r_i^*(\lambda_i)) \end{cases}$$

Em que $p(\lambda, acionamento, t) = \lambda^{acionamento} \exp(-xt)$ é proporcional a probabilidade da função de densidade da distribuição de Poisson, $\phi(x, \mu, \sigma) = \exp[-(1/2)((x - \mu)/\sigma)^2]$ é proporcional a função de densidade de probabilidade da normal e $I(.)$ é uma função indicadora.

A distribuição a *posterior* para $\ln r_i$ é uma normal truncada, para a qual usamos uma simples amostragem da "função de densidade acumulada invertida" (DEVROYE, 1986). A *posterior* para a $\ln \lambda_i$ é menos trivial, com isso será usada uma "sliced sampler" (DAMLÉN; WAKEFIELD; WALKER, 1999). A ideia básica é reescrever:

$$Pr(\lambda_i) = b_0(\lambda_i)b_1(\lambda_i)b_2(\lambda_i)$$

Em que:

$$\begin{aligned} b_0(\lambda_i) &\text{ é uma distribuição normal truncada;} \\ b_1(\ln \lambda_i) &= \lambda_i^{acionamento_i} = (\exp(\ln \lambda_i))^{acionamento_i}; \\ b_2(\ln \lambda_i) &= \exp(-\lambda_i t_i) = \exp(-t_i \exp(\ln \lambda_i)). \end{aligned}$$

O aumento dos dados será realizado por meio de duas variáveis adicionais, u_i^1 e u_i^2 , que são distribuídas uniformemente em $[0, b_1(\lambda_i)]$ e $[0, b_2(\lambda_i)]$, respectivamente. E a probabilidade será:

$$\begin{aligned} Pr(\lambda_i, u_i^1, u_i^2) &= b_0(\lambda_i)b_1(\lambda_i)b_2(\lambda_i)[I(0 \leq u_i^1 \leq b_1(\lambda_i)/b_1(\lambda_i))][I(0 \leq u_i^2 \leq b_2(\lambda_i)/b_2(\lambda_i))] \\ &= b_0(\lambda_i)I(0 \leq u_i^1 \leq b_1(\lambda_i))I(0 \leq u_i^2 \leq b_2(\lambda_i)) \end{aligned}$$

Ao passo que $b_1(\cdot)$ e $b_2(\cdot)$ são funções monotônicas condicionais em u_i^1 e u_i^2 , isso significa que $b_1^{-1}(u_i^1) = ((\ln u_i^1)/acionamentos_i)$ é o limite inferior de $\ln \lambda_i$ (para $acionamentos_i > 0$) e $b_2^{-1}(u_i^2) = \ln(-\ln u_i^2) - \ln t_i$ é o limite superior de $\ln \lambda_i$. Então, a amostragem de λ_i de uma normal truncada ocorre após a modificação dos limites de acordo com u_i^1 e u_i^2 .

B Estatísticas Descritivas das *dummies* de CEP

Tabela 6: Estatísticas Descritivas das *dummies* de CEP

Região	Média	D.P.	Min.	Max.
Centro	0,002	0,048	0	1
Bom Retiro	0,004	0,063	0	1
Vila Buarque	0,013	0,113	0	1
Consolação	0,011	0,103	0	1
Jardins	0,014	0,117	0	1
Liberdade	0,010	0,100	0	1
Santana	0,009	0,095	0	1
Vila Maria	0,006	0,078	0	1
Jaçanã/ Tucuruvi	0,014	0,118	0	1
Tremembé	0,016	0,127	0	1
Mandaqui	0,016	0,125	0	1
Casa Verde	0,008	0,090	0	1
Cachoeirinha	0,007	0,086	0	1
Limão	0,006	0,080	0	1
Brasilândia	0,011	0,103	0	1
Freguesia do Ó	0,017	0,131	0	1
Brás	0,010	0,098	0	1
Moóca	0,019	0,137	0	1
São Lucas	0,012	0,110	0	1
Anália Franco	0,019	0,135	0	1
Carrão e Aricanduva	0,011	0,107	0	1
Vila Matilde	0,019	0,137	0	1
Penha	0,015	0,123	0	1
Cangaíba	0,007	0,084	0	1
Ermelino Matarazzo	0,008	0,091	0	1
São Mateus	0,014	0,116	0	1
Moema/ Vila Mariana (OESTE)	0,019	0,135	0	1
Saúde/ Vila Mariana (LESTE)	0,024	0,154	0	1
Ipiranga	0,016	0,126	0	1
Jabaquara	0,022	0,145	0	1
Cidade Ademar	0,015	0,121	0	1
Itaim Bibi	0,022	0,146	0	1
Campo Belo	0,017	0,128	0	1

continua na próxima página

Tabela 6 – continuação da página anterior

Região	Média	D.P.	Min.	Max.
Santo Amaro	0,012	0,109	0	1
Cidade Dutra/ Grajaú	0,025	0,156	0	1
Guarapiranga	0,007	0,085	0	1
Lapa e Perdizes	0,020	0,140	0	1
Pirituba/ Jaraguá	0,014	0,116	0	1
Perus	0,005	0,074	0	1
Jaguaré/ Leopoldina	0,010	0,100	0	1
Pinheiros	0,016	0,124	0	1
Butantã	0,012	0,109	0	1
Morumbi	0,014	0,116	0	1
Campo Limpo	0,018	0,134	0	1
Capão Redondo	0,017	0,128	0	1
Osasco (REGIÃO 1)	0,008	0,091	0	1
Osasco (REGIÃO 2)	0,013	0,114	0	1
Osasco (REGIÃO 3)	0,009	0,097	0	1
Carapicuíba	0,012	0,110	0	1
Barueri	0,012	0,109	0	1
Santana do Parnaíba	0,007	0,085	0	1
Itapevi/ Jandira	0,007	0,086	0	1
Cotia/ Vargem Grande Pta,/ Taboão da Serra	0,026	0,159	0	1
Embu e Itapeirica da Serra	0,011	0,106	0	1
Embu-Guaçu	0,003	0,052	0	1
Guarulhos (REGIÃO 1)	0,019	0,137	0	1
Guarulhos (REGIÃO 2)	0,036	0,185	0	1
Guarulhos (REGIÃO 3)	0,013	0,112	0	1
Mairiporã	0,003	0,050	0	1
Caieiras/ Cajamar/ Jordanesia/ Polvilho	0,004	0,064	0	1
Franco da Rocha	0,003	0,055	0	1
Francisco Morato	0,002	0,046	0	1
São Miguel Paulista	0,014	0,116	0	1
Itaim Paulista	0,008	0,087	0	1
Itaquera	0,016	0,126	0	1
São Mateus	0,009	0,095	0	1
Guaianazes	0,013	0,112	0	1
Santo André (REGIÃO 1)	0,014	0,117	0	1
Santo André (REGIÃO 2)	0,021	0,143	0	1

continua na próxima página

Tabela 6 – continuação da página anterior

Região	Média	D.P.	Min.	Max.
Santo André (REGIÃO 3)	0,013	0,111	0	1
Mauá	0,014	0,118	0	1
Ribeirão Pires/ Rio Grande da Serra	0,005	0,071	0	1
São Caetano do Sul	0,015	0,120	0	1
São Bernardo do Campo (REGIÃO 1)	0,009	0,094	0	1
São Bernardo do Campo (REGIÃO 2)	0,019	0,138	0	1
São Bernardo do Campo (REGIÃO 3)	0,024	0,152	0	1
Diadema	0,014	0,115	0	1

C Resultado das *dummies* de CEP - Estimação Bayesiana

Tabela 7: Resultado das *dummies* de CEP - Estimação Bayesiana

Região	$\ln \lambda$	Região	$\ln \lambda$
Centro		Jaguaré/ Leopoldina	-0,1016 (0,1545)
Bom Retiro	-0,1581 (0,1816)	Pinheiros	0,1142 (0,1448)
Vila Buarque	0,0348 (0,1498)	Butantã	-0,1440 (0,1524)
Consolação	0,1864 (0,1468)	Morumbi	-0,0169 (0,1483)
Jardins	-0,0217 (0,1487)	Campo Limpo	-0,1930 (0,14683)
Liberdade	-0,0321 (0,1525)	Capão Redondo	-0,3204 (0,1496)
Santana	-0,1017 (0,1563)	Osasco (REGIÃO 1)	-0,0898 (0,1571)
Vila Maria	-0,2211 (0,1692)	Osasco (REGIÃO 2)	-0,2533 (0,1527)
Jaçanã/ Tucuruvi	-0,1838 (0,1505)	Osasco (REGIÃO 3)	-0,2269 (0,1590)
Tremembé	-0,1437 (0,1480)	Carapicuíba	-0,2793 (0,1561)
Mandaqui	-0,0787 (0,1459)	Barueri	-0,2038 (0,1548)
Casa Verde	-0,2271 (0,1616)	Santana do Parnaíba	-0,2517 (0,1668)
Cachoeirinha	-0,1556 (0,1595)	Itapevi/ Jandira	-0,3528 (0,1681)
Limão	-0,3268 (0,1763)	Cotia/ Vargem Grande Pta./ Taboão da Serra	-0,2755 (0,1443)
Brasilândia	-0,3525 (0,1595)	Embu e Itapeirica da Serra	-0,4098 (0,1608)
Freguesia do Ó	-0,2239 (0,1488)	Embu-Guaçu	-0,0487 (0,1872)
Brás	-0,1561 (0,1610)	Guarulhos (REGIÃO 1)	-0,2108 (0,1479)
Moóca	-0,1726 (0,1473)	Guarulhos (REGIÃO 2)	-0,3288 (0,1421)
São Lucas	-0,4429 (0,1639)	Guarulhos (REGIÃO 3)	-0,3592 (0,1529)
Anália Franco	-0,1676 (0,1471)	Mairiporã	-0,1091 (0,1954)
Carrão e Aricanduva	-0,3473 (0,1608)	Caieiras /Cajamar /Jordanesia /Polvilho	-0,1801 (0,1858)
Vila Matilde	-0,3106 (0,1498)	Franco da Rocha	0,0262 (0,1813)
Penha	-0,1098 (0,1485)	Francisco Morato	-0,2095 (0,2156)
Cangaíba	-0,3415 (0,1761)	São Miguel Paulista	-0,4102 (0,1610)
Ermelino Matarazzo	-0,2795 (0,1664)	Itaim Paulista	-0,3025 (0,1699)
São Mateus	-0,4722 (0,1656)	Itaquera	-0,3932 (0,1570)
Moema/ Vila Mariana (OESTE)	0,0595 (0,1446)	São Mateus	-0,3600 (0,1726)
Saúde/ Vila Mariana (LESTE)	-0,1072 (0,1453)	Guaianazes	-0,4695 (0,1639)
Ipiranga	-0,2796 (0,1523)	Santo André (REGIÃO 1)	-0,3524 (0,1524)
Jabaquara	-0,1758 (0,1456)	Santo André (REGIÃO 2)	-0,4593 (0,1530)
Cidade Ademar	-0,2839 (0,1518)	Santo André (REGIÃO 3)	-0,3882 (0,1580)
Itaim Bibi	0,0476 (0,1428)	Mauá	-0,4375 (0,1568)
Campo Belo	-0,1024 (0,1462)	Ribeirão Pires/ Rio Grande da Serra	-0,3900 (0,1803)
Santo Amaro	-0,0375 (0,1491)	São Caetano do Sul	-0,1872 (0,1497)
Cidade Dutra/ Grajaú	-0,2709 (0,1462)	São Bernardo do Campo (REGIÃO 1)	-0,2962 (0,1671)
Guarapiranga	-0,3448 (0,1680)	São Bernardo do Campo (REGIÃO 2)	-0,2663 (0,1475)
Lapa e Perdizes	-0,0219 (0,1455)	São Bernardo do Campo (REGIÃO 3)	-0,3142 (0,1468)
Pirituba/ Jaraguá	-0,2539 (0,1547)	Diadema	-0,4155 (0,1589)
Perus	-0,2436 (0,1725)		