

Enrico Peceguini Ruggieri

**Informação Assimétrica no mercado de saúde
suplementar brasileiro: uma abordagem
econométrica**

São Paulo, SP, Brasil

26 de Junho de 2024

Enrico Peceguini Ruggieri

Informação Assimétrica no mercado de saúde suplementar brasileiro: uma abordagem econométrica

Monografia apresentada ao departamento de Ciências Econômicas da FEA-USP para conclusão do Bacharelado em Ciências Econômicas.

Universidade de São Paulo - USP

Faculdade de Economia, Administração, Contabilidade e Atuária

Bacharelado em Ciências Econômicas

Orientadora: Profa. Dra. Denise Cavallini Cyrillo

São Paulo, SP, Brasil

26 de Junho de 2024

Enrico Peceguini Ruggieri

Informação Assimétrica no mercado de saúde suplementar brasileiro: uma abordagem econométrica/ Enrico Peceguini Ruggieri. – São Paulo, SP, Brasil, 26 de Junho de 2024-

41 p. : il. (algumas color.) ; 30 cm.

Orientadora: Profa. Dra. Denise Cavallini Cyrillo

Monografia – Universidade de São Paulo - USP

Faculdade de Economia, Administração, Contabilidade e Atuária

Bacharelado em Ciências Econômicas, 26 de Junho de 2024.

1. Risco e Incerteza. 2. Economia da Saúde

JEL: D82, G22, I11

Agradecimentos

À USP, por ser um espaço tão significativo para mim, presente em tantos momentos da minha vida pessoal e acadêmica.

À minha orientadora Denise, pelo direcionamento atencioso, pela paciência com minhas idas e vindas, e pela receptividade às minhas ideias, mesmo quando elas não seguiram os caminhos mais ortodoxos.

Aos meus colegas de faculdade, pelo apoio nos momentos difíceis e pelas amizades sinceras que não esperava encontrar ao ingressar na faculdade em uma fase mais madura da minha vida.

À minha irmã Mari, que sempre me faz pensar sobre como é possível gostar tanto de alguém tão diferente de si mesmo.

À Veri, que conhece todos os meus defeitos e responde a eles apenas com carinho, pelo enorme apoio e suporte ao longo desses quatro anos.

À minha mãe, por me ensinar a buscar sempre mais e a exigir o melhor de mim mesmo.

Ao meu pai, cuja inteligência sempre me inspirou a estudar e a me empenhar na curiosidade verdadeira pelo conhecimento.

Meu sincero muito obrigado a todos que caminharam ao meu lado para que eu pudesse chegar até aqui.

*“Viver é muito perigoso,
Porque aprender a viver é que o viver mesmo.
Travessia perigosa, mas é a da vida.
Sertão que se alteia e se abaixa.
O mais difícil não é ser bom e proceder honesto;
Difícil, mesmo, é saber definido o que quer,
e ter o poder de ir até o rabo da palavra.”
(João Guimarães Rosa)*

Resumo

A partir da motivação do fenômeno da Seleção Adversa, este trabalho explora os desafios do mercado de saúde suplementar brasileiro e sua susceptibilidade ao problema da Informação Assimétrica. Por meio da aplicação do teste de correlação positiva ([CHIAPPORI; SALANIE, 2000](#)) aos dados da Pesquisa Nacional de Saúde de 2019, o trabalho busca contribuir com a bibliografia que estuda a presença dessa falha informacional.

Como contribuição original, propõe-se também uma atenção especial às extensões do teste, realizando uma revisão metodológica dessa abordagem econométrica para verificação da Informação Assimétrica. Em especial, são exploradas as extensões de variáveis não-utilizadas de [Finkelstein e Poterba \(2014\)](#) e o modelo multinomial de [Kim et al. \(2009\)](#).

Encontram-se evidências estatisticamente significativas da presença de informação assimétrica nesse mercado. Também são identificadas variáveis, com significância estatística, que contribuem para esse problema informacional por não serem utilizadas pelas companhias de seguro na diferenciação dos prêmios. Além disso, uma Regressão Binomial Negativa revela resultados semelhantes, utilizando um modelo mais complexo que permite a diferenciação entre tipos de cobertura.

Palavras-chaves: Economia da Saúde. Informação Assimétrica. Risco e Seguro.

Códigos JEL: D82, G22, I11.

Abstract

Motivated by the phenomenon of Adverse Selection, this work explores the challenges of the Brazilian supplementary health market and its susceptibility to the Asymmetric Information problem. By applying the positive correlation test ([CHIAPPORI; SALANIE, 2000](#)) to the 2019 National Health Survey data, the study aims to contribute to the literature examining the presence of this informational failure.

As an original contribution, special attention is also proposed for extensions of the test, conducting a methodological review of this econometric approach for assessing Asymmetric Information. In particular, the unused variable extensions by [Finkelstein e Poterba \(2014\)](#) and the multinomial model by [Kim et al. \(2009\)](#) are explored.

Statistically significant evidence of asymmetric information presence in this market is found. Additionally, variables are identified, with statistical significance, that contribute to this informational problem by not being utilized by insurance companies in premium differentiation. Furthermore, a Negative Binomial Regression reveals similar results, using a more complex model that allows differentiation between coverage types.

Keywords: Health Economics, Asymmetric Information, Risk and Insurance.

JEL codes: D82, G22, I11.

Lista de ilustrações

Figura 1 – Índice de Preços ao Consumidor Amplo: Geral e subíndices relacionados ao Plano de Saúde	11
Figura 2 – Hipótese do efeito da Seleção Adversa no mercado de saúde	14
Figura 3 – Média de Sinistros por faixa etária, com faixas adicionais	30

Lista de tabelas

Tabela 1	– Resultado da Regressão Probit da Cobertura para o modelo canônico .	27
Tabela 2	– Resultado da Regressão Probit do Sinistro para o modelo canônico . .	27
Tabela 3	– Resultado da Regressão Probit da Cobertura - variável fumante	29
Tabela 4	– Resultado da Regressão Probit do Sinistro - variável fumante	29
Tabela 5	– Resultado da Regressão Probit da Cobertura - diabetes e faixas de idade adicionais	31
Tabela 6	– Resultado da Regressão Probit do Sinistro - diabetes e faixas de idade adicionais	32
Tabela 7	– Modelo Multinomial - Regressão Probit Ordenado da Cobertura	33
Tabela 8	– Modelo Multinomial - Regressão Binomial Negativa do Sinistro utilizando resíduos calculados como regressores	34

Lista de abreviaturas e siglas

ANS	Agência Nacional de Saúde
CPI	Comissão Parlamentar de Inquérito
IBGE	Instituto Brasileiro de Geografia e Estatística
IPCA	Índice de Preços ao Consumidor Amplo
MP	Ministério Público
OMS	Organização Mundial da Saúde
PNAD	Pesquisa Nacional de Amostra Domiciliar
PNS	Pesquisa Nacional de Saúde
QMLE	Quasi-Maximum Likelihood Estimator (Estimador de quase-máxima verossimilhança)
SIDRA	Sistema IBGE de Recuperação Automática

Sumário

	Resumo	5
	Abstract	6
1	INTRODUÇÃO	11
1.1	O mercado de saúde suplementar brasileiro: contexto e perspectivas	11
1.2	O problema da Informação Assimétrica em mercados de seguro . .	13
1.2.1	Seleção Adversa e Risco Moral	13
1.2.2	Perfil de Risco e Aversão ao Risco	15
2	REVISÃO BIBLIOGRÁFICA	16
3	MATERIAL E MÉTODO	19
3.1	Teste de Correlação Positiva: o modelo canônico	20
3.2	Extensões do teste de Correlação Positiva	21
3.2.1	Heterogeneidade preferências individuais e variáveis não-utilizadas	21
3.2.2	Medidas multinomiais de cobertura e sinistro	23
4	RESULTADOS	26
4.1	PNS 2019: Uma visão geral	26
4.2	Teste de Correlação Positiva	26
4.2.1	O modelo canônico	26
4.2.2	Heterogeneidade das Preferências e Variáveis não-utilizadas	28
4.2.3	Testando outras variáveis não-utilizadas	30
4.2.4	Uma proposta de modelo multinomial para a Cobertura	32
4.3	Risco Moral vs. Seleção Adversa: uma breve discussão	34
	Conclusão	36
	REFERÊNCIAS	38
	APÊNDICE A – CÓDIGOS EM R E PYTHON	41

1 Introdução

1.1 O mercado de saúde suplementar brasileiro: contexto e perspectivas

A saúde suplementar no Brasil tem, atualmente, 50.5 milhões de beneficiários, quase um quarto da população brasileira. O setor, contudo, enfrenta graves percalços, em que o elevado grau de insatisfação dos consumidores convive com dificuldades financeiras das operadoras. Enquanto os reajustes dos planos, tanto individuais quanto coletivos, consistentemente superam a média da inflação medida pelo IPCA ([OCKÉ-REIS; FIUZA; COIMBRA, 2019](#)), as operadoras relatam prejuízos consecutivos em seus resultados financeiros. O setor é regulamentado pela Agência Nacional de Saúde Suplementar (ANS), instituída pela Lei 9656/98, que define e fiscaliza os reajustes, bem como o rol de procedimentos da saúde a serem cobertos pelas operadoras.

Em relação à possibilidade de diferenciação de preços pelas seguradoras, o caso da saúde suplementar brasileira chama atenção pela legislação notavelmente restritiva nesse aspecto. A Súmula Normativa 27/2015 da ANS afirma que “é vedada a prática de seleção

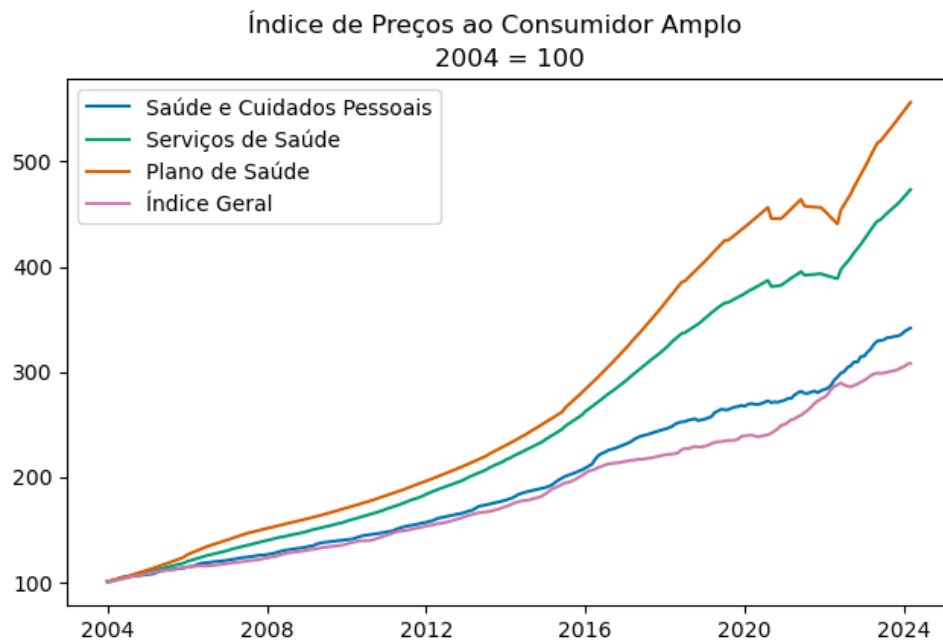


Figura 1 – Índice de Preços ao Consumidor Amplo: Geral e subíndices relacionados ao Plano de Saúde

Fonte: Elaboração própria a partir de dados do SIDRA IBGE

de riscos pelas operadoras de plano de saúde na contratação de qualquer modalidade de plano privado de assistência à saúde”(ANS, 2015). Atualmente, é permitida somente a diferenciação do prêmio em 10 faixas de idade, sendo que o Estatuto do Idoso impede qualquer distinção após os 59 anos. A judicialização também impõe dificuldades adicionais às operadoras, que se veem obrigadas a custear remédios caríssimos, mesmo quando esses não estão inclusos no rol da ANS (Conselho Nacional de Justiça, 2024).

Ao mesmo tempo, a insatisfação dos consumidores com o serviço oferecido é enorme. O número de reclamações registrado na agência reguladora cresce vertiginosamente (PIOVEZAN, 2023), com queixas sobre tratamentos negados e sobre piora nas coberturas. Além disso, o custo pago por esses serviços demanda uma parcela cada vez maior do orçamento das famílias. Na Figura 1 é possível notar como os subíndices do indicador relacionados ao plano de saúde e a serviços de saúde têm consistentemente superado o índice geral.

Uma mais recente polêmica relacionada à saúde suplementar também tem reverberado na sociedade brasileira. O cancelamento unilateral dos planos de saúde pelas seguradoras tem afetado muitos brasileiros, com concentração desproporcional em idosos, portadores de condições pré-existentes e indivíduos em tratamento de doenças graves (MUGNATTO, 2024). Assembleias Legislativas de diversas Unidades Federativas se organizam para instaurar CPIs, MPs estaduais instauram inquéritos e a ANS se mobiliza para exercer seu papel de agência reguladora. Trata-se de um problema grave que expõe indivíduos já fragilizados a situações de risco em seu momento de maior vulnerabilidade, subvertendo a ideia de proteção financeira, preconizada como princípio norteador de sistemas de saúde pela OMS (WORLD HEALTH ORGANIZATION, 2014).

É dessa difícil conjunção de fatores que emerge a motivação deste trabalho. À medida que os custos crescentes das operadoras são repassados aos consumidores, aumenta o incentivo para que indivíduos saudáveis deixem o mercado e para que indivíduos com maior risco escolham contratar o seguro, elevando cada vez mais a sinistralidade dos planos. A existência de Informação Assimétrica nesse mercado, portanto, parece uma hipótese bastante razoável.

Uma metodologia já previamente empregada pela literatura para avaliação da Informação Assimétrica no mercado de saúde suplementar é o teste de correlação positiva, proposto inicialmente por Chiappori e Salanie (2000). O ineditismo deste trabalho consiste na aplicação dessa metodologia à Pesquisa Nacional da Saúde mais recente (2019) e, principalmente, a um enfoque especial nas extensões propostas ao teste econométrico, até então pouco exploradas pela literatura.

1.2 O problema da Informação Assimétrica em mercados de seguro

Considerando o aumento na sinistralidade, o significativo aumento no custo real das apólices, bem como questões demográficas relacionadas ao envelhecimento da população brasileira, a questão que se faz presente é se existe Informação Assimétrica no mercado de saúde suplementar.

Esse fenômeno é bastante frequente em mercados de seguros e costuma ser dividido entre duas falhas de mercado: a Seleção Adversa e o Risco Moral. Enquanto o problema da seleção adversa é visto como um problema informacional *ex-ante*, o risco moral é o problema informacional *ex-post*. No caso do mercado do seguros, a Seleção Adversa motivaria o indivíduo com mais risco a aderir mais à cobertura (ou a aderir a uma apólice mais completa). O Risco Moral, por sua vez, seria o indivíduo adotar uma postura mais arriscada pelo fato de estar coberto pelo seguro. O efeito da Informação Assimétrica pode ser extremamente deletério para mercados de seguro, levando a uma seleção desvantajosa de indivíduos, aumentando cada vez mais os prêmios e a sinistralidade, e podendo levar a um colapso desse mercado.

1.2.1 Seleção Adversa e Risco Moral

O tema da seleção adversa no mercado da saúde suplementar tem sido objeto de estudo nas últimas décadas. Esse fenômeno foi inicialmente descrito no célebre artigo *The Market for Lemons* (AKERLOF, 1970), e ocorre quando indivíduos utilizam informações privadas *ex-ante* para escolherem participar ou não de um mercado. Alguns anos mais tarde, Rotschild e Stiglitz (1976) aplicaram esse conceito ao mercado de seguros, quando os indivíduos possuem informações privadas sobre seu risco que influenciam na escolha de contratar ou não um seguro.

O conceito da Seleção Adversa no mercado da saúde suplementar é de compreensão bastante intuitiva. Imaginemos um conjunto de indivíduos i em que cada um desses indivíduos possui uma dada disposição a pagar d_i por uma apólice de seguro saúde. É razoável supor que a utilidade esperada associada ao plano de saúde é tanto maior quanto pior o estado de saúde desse indivíduo. Dessa forma, a disposição a pagar também é maior de acordo com o estado de saúde do indivíduo *ex-ante* à decisão sobre aderir ou não ao plano.

Se esse estado de saúde é desconhecido pela seguradora (ou se há algum outro impedimento para diferenciação ao prêmio), é inequívoco notar como esse fenômeno pode levar ao colapso desse mercado. Havendo um aumento no preço da apólice, indivíduos com d_i menor terão um incentivo maior para deixar o plano. Isto é, ocorrerá exclusão de indivíduos com estado de saúde melhor, o que tenderá a elevar a sinistralidade do plano, levando conseqüentemente a um aumento do prêmio. Estabelece-se, assim, um “efeito

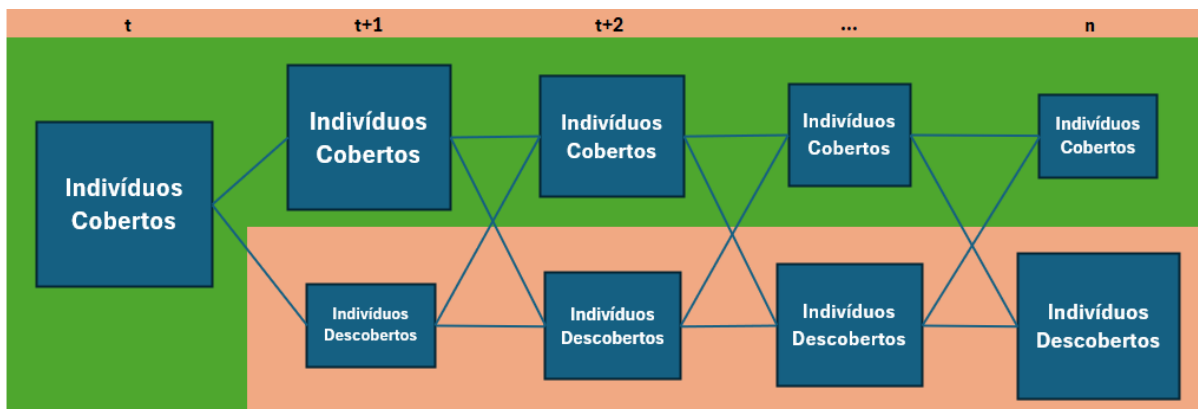


Figura 2 – Hipótese do efeito da Seleção Adversa no mercado de saúde

Fonte: Elaboração própria adaptada de [Winssen, Kleef e Ven \(2018\)](#)

bola de neve” que, no limite, faz com que haja uma adesão cada vez menor de indivíduos saudáveis e cada vez maior de indivíduos com estado de saúde cada vez pior.

Para ilustrar esse problema, um estudo digno de nota é, na verdade, o artigo que motivou esta pesquisa. Utilizando uma base de dados bastante detalhada a respeito do mercado de saúde suplementar na Holanda, [Winssen, Kleef e Ven \(2018\)](#) realizaram uma simulação iterativa, repetida por 25 períodos, analisando a entrada e saída de indivíduos e reestimando a taxa de sinistralidade e o prêmio pago em cada período. A Figura 2 representa o efeito da Seleção Adversa em um mercado de saúde suplementar como modelado pelos autores. A conclusão do artigo foi que uma diferenciação do prêmio poderia reduzir o problema da seleção adversa e evitar o colapso desse mercado.

O Risco Moral, por sua vez, como discutido anteriormente, é o problema informacional *ex-post*. Nesse fenômeno, o agente econômico adota um comportamento mais arriscado (ou aumenta a sinistralidade) após a decisão da adesão ao seguro. O artigo seminal que introduziu esse conceito no âmbito dos seguros foi [Arrow \(1963\)](#).

No caso do seguro saúde, esse conceito pode parecer, a um primeiro momento, pouco razoável. Seria ilógico supor, por exemplo, que um indivíduo ficaria mais doente por estar coberto por uma apólice. Para procedimentos eletivos, contudo, a hipótese é mais factível, podendo o agente econômico realizar procedimentos desnecessários por estar coberto por um plano, possibilidade que motiva a existência da coparticipação em diversas apólices.

Alguns autores criticam essa ideia, contudo, argumentando que o suposto “desperdício” atribuído ao risco moral é, na verdade, um reflexo da sub-utilização de serviços de saúde por indivíduos sem cobertura, incapazes de arcar com os elevados custos desse cuidado ([ROBERTSON et al., 2020](#)).

Dada a menor clareza do fenômeno do Risco Moral na saúde suplementar, a motivação deste trabalho surge a partir de uma atenção especial à Seleção Adversa nesse mercado. É importante ressaltar, contudo, que o teste de correlação positiva e suas extensões não é capaz de diferenciar esses dois fenômenos de Informação Assimétrica. A metodologia é robusta para identificar se existe informação privada dos consumidores sobre o seu risco, mas não sobre a natureza específica dessa informação.

Both adverse selection and moral hazard can generate a positive correlation between insurance coverage and claims, but these are two very different forms of asymmetric information with very different implications for public policy. With adverse selection, individuals who have private information that they are at higher risk self-select into the insurance market, generating the positive correlation between insurance coverage and observed claims. As already discussed, the government has several potential welfare-improving policy tools to possibly address such selection. With moral hazard, individuals are identical before they purchase insurance, but have incentives to behave differently after. Those with greater coverage have less incentive to take actions that reduce their expected costs, which will generate a relationship between insurance coverage and observed claims. (EINAV; FINKELSTEIN, 2011)

Apesar dessa ressalva, na seção de resultados deste trabalho será apresentada uma breve discussão, também baseada no artigo de [Finkelstein e Poterba \(2014\)](#), que propõe uma diferenciação de caráter eminentemente qualitativo entre os dois efeitos dos fenômenos de Informação Assimétrica.

Com a finalidade de ser fiel ao rigor metodológico do teste empregado, e reconhecendo a limitação dessa abordagem para diferenciar esses dois fenômenos, o termo *Informação Assimétrica* será predominantemente utilizado ao longo deste trabalho.

1.2.2 Perfil de Risco e Aversão ao Risco

Uma última definição importante para esta introdução é diferenciar a informação privada proveniente de diferentes perfis de risco da heterogeneidade das preferências individuais. *Ceteris paribus*, indivíduos com uma aversão ao risco costumam demandar mais apólices de seguro, ou apólices mais completas.

O problema que emerge, portanto, é se a correlação positiva identificada será proveniente dessa heterogeneidade na aversão ao risco dos indivíduos ou dos perfis diferentes de risco. A diferenciação é importante porque pode invalidar o resultado do teste econométrico, e será discutida mais a fundo na seção de metodologia deste trabalho.

2 Revisão Bibliográfica

Poucos anos depois da publicação do célebre artigo “The Market for Lemons”, de [Akerlof \(1970\)](#), o fenômeno da Informação Assimétrica foi estudado no contexto do mercado de seguros, com a significativa contribuição de [Rothschild e Stiglitz \(1976\)](#). Esse artigo seminal deu origem ao modelo Rothschild-Stiglitz, que descreve a ocorrência de Seleção Adversa nesse domínio e que, posteriormente, justificaria a entrega do Prêmio Nobel ao segundo autor. Desde então, a questão da informação assimétrica no mercado de seguros tem sido objeto de extensa investigação. [Arrow \(1963\)](#), por sua vez, publicou um trabalho de grande relevância sobre o Risco Moral no mercado de seguros.

Se nas primeiras décadas após a publicação desses artigos o debate se focava sobretudo em aspectos teóricos, mais recentemente a discussão tem se concentrado em aspectos empíricos. Pesquisadores têm recorrido a conjuntos de dados extensos e utilizam métodos econométricos e quantitativos para investigar a presença ou não do fenômeno de informação assimétrica no mercado de seguros.

Desde a publicação de [Chiappori e Salanie \(2000\)](#), o modelo canônico utilizado para verificação de informação assimétrica tem sido o teste de correlação positiva, que estima duas equações na forma reduzida, uma para a cobertura de seguro, e a outra para o risco de perda (sinistro). A ideia desse procedimento é testar a hipótese, decorrente do modelo de Rothschild-Stiglitz, de que contratos com uma cobertura mais completa serão escolhidos por agentes mais suscetíveis ao risco de sinistro. O teste consiste em duas equações, uma primeira que modela a cobertura do seguro em função das variáveis exógenas observáveis do indivíduo, e uma segunda que modela o risco de sinistro com essas mesmas variáveis. Se há correlação positiva nos resíduos dessas regressões tem-se uma evidência da presença de informação assimétrica. A metodologia será formalizada de forma mais aprofundada no próximo capítulo desta monografia.

Embora o artigo de [Chiappori e Salanie \(2000\)](#) tenha estudado o mercado de seguros automotivos na França, essa mesma metodologia também foi bastante aplicada a uma série de outros mercados. [Cohen e Siegelmann \(2010\)](#) fizeram uma grande revisão metodológica das publicações testando a informação assimétrica para diferentes tipos de seguro, e verificaram que o método desses dois autores era um dos mais frequentemente utilizados para essa finalidade.

Mais especificamente no mercado de saúde, alguns artigos também se destacam por sua relevância na investigação desse fenômeno. [Cutler e Zeckhauser \(2000\)](#) fazem uma extensa revisão bibliográfica e descobrem que a grande maioria dos artigos publicados até então encontrou fortes evidências da existência de informação assimétrica no mercado de

seguro saúde. Dentre os trinta artigos que os dois autores agregam, publicados ao longo de 26 anos, vinte e cinco constataam ocorrência de informação assimétrica e três apontam para um efeito ambíguo. A metodologia para a avaliação desse fenômeno de informação assimétrica no mercado de saúde é bastante variada, mas os resultados, de maneira geral, predominantemente apontam para a ocorrência do problema informacional. Uma importante exceção é o trabalho amplamente citado de [Cardon e Hendel \(2001\)](#) que, utilizando um modelo estrutural de duas etapas, não encontra evidências de informação assimétrica, sendo a maior parte da variância nos sinistros explicada por variáveis observáveis.

No Brasil, a metodologia do teste de correlação positiva foi aplicada à saúde suplementar por alguns autores ao longo dessas duas últimas décadas. [Alves \(2004\)](#) e [Resende e Zeidan \(2010\)](#) testaram a hipótese da informação assimétrica no mercado de saúde suplementar brasileiro utilizando os dados da PNAD Saúde 2003 (atual PNS). O primeiro trabalho encontrou evidências de informação assimétrica, enquanto o segundo obteve resultados diferentes. Esses dois autores, contudo, parecem ter incorrido em alguns desvios do modelo canônico ao utilizar na matriz de variáveis observáveis uma grande quantidade de variáveis exógenas não utilizadas pela companhia de seguros. Como discutido previamente, a legislação brasileira impõe uma série de restrições às possibilidades de diferenciação de prêmio que as seguradoras podem realizar. A inclusão, portanto, de uma grande quantidade de variáveis exógenas na matriz parece divergir do que sugere a bibliografia desse assunto. Em um esforço subsequente, [Fonseca \(2017\)](#) repetiu o procedimento desses autores, utilizando os dados da PNS 2013. Ao empregar uma metodologia mais robusta e mais próxima do modelo canônico, também encontrou evidências de informação assimétrica.

Outros métodos foram aplicados na investigação dos problemas informacionais no sistema de saúde suplementar brasileiro. [Sá \(2012\)](#) faz uma revisão bibliográfica de diversas publicações que utilizam os dados da PNAD Saúde 1998 e 2003, utilizando desde o método proposto por Chiappori e Salanie até outros procedimentos econométricos, como o estimador de DiD empregado por [Nishijima, Postali e Fava \(2011\)](#). Dentre os artigos reunidos por [Sá \(2012\)](#), cinco investigam a presença de risco moral, sendo que todos encontram evidências da presença dessa falha de mercado. No caso da seleção adversa, dos sete artigos que verificam a presença desse problema informacional, apenas dois encontram evidências significativas. Esse é um resultado bastante interessante, sobretudo considerando as diversas limitações institucionais para diferenciação do prêmio do seguro saúde no Brasil.

Mais recentemente, alguns esforços foram realizados no afã de corrigir algumas das limitações do teste de correlação positiva. [Cutler, Finkelstein e McGarry \(2008\)](#) sugerem a inclusão de variáveis que capturem características individuais relacionadas à aversão ao risco. A ideia dessa metodologia seria controlar o modelo por fatores que possam gerar algum tipo de endogeneidade nos modelos devido a diferenças nas preferências individuais.

Finkelstein e Poterba (2014), por sua vez, propõem a adição de mais um vetor W_i incorporando as variáveis observáveis não utilizadas pela companhia de seguros. Esse teste de “*unused observables*” buscaria endereçar o problema de variáveis que podem gerar a informação assimétrica tanto pela informação privada de diferentes perfis como por problemas de heterogeneidade em parâmetros individuais que podem afetar a demanda por seguro. Ao aplicar esse modelo no mercado de previdência privada do Reino Unido (*annuity market*), os autores identificaram que a desconsideração do local de residência (a *unused observable*) no mercado de *annuity insurance* gera o mesmo tipo de ineficiência que surge quando os contratantes detêm informações privadas sobre o risco de mortalidade. O local de residência acabaria sendo incluído como variável determinante do prêmio pelas companhias posteriormente, no Reino Unido. Uma breve aplicação dessa extensão ao caso brasileiro, utilizando a PNS de 2013, é sugerida por Fonseca (2017), mas o autor utiliza esse método somente como teste de robustez na seção final de seus resultados.

Também no âmbito das extensões do teste de correlação positiva, destaca-se a importância do artigo de Kim et al. (2009), que justificadamente problematiza a característica dicotômica dos testes de correlação positiva. Em termos práticos, os testes costumam se utilizar de variáveis binárias tanto para a cobertura como para o sinistro. Para lidar com essa limitação, os autores utilizam um probit ordenado na variável cobertura, de forma a separar apólices com maior cobertura de apólices mais limitadas. Como resultado, aplicando essa metodologia no mercado de seguro automobilístico, encontram fortes evidências da presença de informação assimétrica.

Nesse mesmo sentido, Dardanoni, Forcina e Donni (2018) propuseram um modelo multivariado para lidar com o problema da natureza multidimensional da informação privada. Essa extensão do teste de correlação positiva baseia-se numa classe flexível de regressões que analisa a associação entre a cobertura e mais de uma variável referente à sinistralidade. Os autores aplicam esse modelo a uma base de dados do *Health and Retirement Service* que estuda os planos do *Medigap* (um tipo de suplemento ao *Medicare* nos EUA), e encontram evidência de que existe correlação positiva e estatisticamente significativa entre risco e cobertura, rejeitando assim a hipótese de informação simétrica.

3 Material e Método

Para averiguação da informação assimétrica no mercado de saúde suplementar brasileiro, a principal base de dados a ser utilizada foi a Pesquisa Nacional de Saúde, conduzida pelo IBGE, de 2019. O instituto também disponibiliza os microdados dos 293.726 indivíduos entrevistados em 108.475 domicílios, sendo possível ter acesso a informações referentes a seu estado de saúde, sua adesão ao sistema de saúde suplementar, bem como dados sobre sua renda e utilização de serviços de saúde.

Com o intuito de verificar a presença de informação assimétrica, foi utilizado um teste de correlação positiva. Um enfoque especial foi dado às extensões desse teste desenvolvidas posteriormente. A primeira extensão, da heterogeneidade das preferências individuais, permite controlar o modelo por alguma variável relativa ao comportamento do indivíduo que se relacione com sua aversão ao risco. A segunda extensão, por sua vez, permite averiguar se existem variáveis que a companhia de seguro não utiliza (ou não pode utilizar) para diferenciar o prêmio que geram esse problema informacional.

Finalmente, será utilizado também um modelo multinomial que permite um ordenamento entre diferentes tipos de cobertura, bem como a utilização de uma variável quantitativa discreta como variável dependente do Sinistro.

3.1 Teste de Correlação Positiva: o modelo canônico

A metodologia utilizada para investigação da informação assimétrica será o teste de correlação positiva, já brevemente mencionado na revisão bibliográfica. O teste econômico consiste em duas equações, onde X_i é uma matriz das variáveis exógenas de cada indivíduo i que são utilizadas pela seguradora para determinação do prêmio.

A primeira equação modela a cobertura do seguro em função das variáveis contidas em X_i , enquanto que a segunda relaciona o risco de sinistro com essas mesmas variáveis. A ideia desse procedimento é testar a hipótese de que contratos com uma cobertura mais completa serão escolhidos por agentes mais suscetíveis ao risco de perda. Se há correlação nas distribuições dessas equações estimadas, isso significa que as variáveis não observáveis (contidas em ϵ e η) estão afetando tanto a escolha do indivíduo para sua cobertura como sua chance de sofrer acidente/utilizar o seguro. Se toda informação privada fosse perfeitamente capturada em X_i , essas distribuições seriam descorrelacionadas, e os resíduos das regressões seriam descorrelacionados também.

Sendo Ω e Π vetores dos parâmetros, podemos escrever:

$$Cobertura_i = X_i\Omega + \epsilon_i \quad (3.1)$$

$$Sinistro_i = X_i\Pi + \eta_i \quad (3.2)$$

No caso da PNS, os dois modelos serão uma regressão das variáveis exógenas utilizadas (faixas de idade) sobre a variável resposta se aqueles indivíduos têm ou não seguro, e uma regressão das mesmas variáveis exógenas sobre a variável resposta do quanto esses indivíduos utilizaram o plano.

Esses modelos são estimados por meio de dois probits independentes e, se rejeitada $H_0 : Cov(\hat{\epsilon}, \hat{\eta}) = 0$, rejeita-se a hipótese nula de informação simétrica. Os autores sugerem, para verificação dessa hipótese, a estatística W , onde w_i é o número de dias que o indivíduo i estaria coberto pelo seguro:

$$W = \frac{\left(\sum_{i=1}^n w_i \hat{\epsilon}_i \hat{\eta}_i \right)^2}{\sum_{i=1}^n w_i^2 \hat{\epsilon}_i^2 \hat{\eta}_i^2} \quad (3.3)$$

Nesse caso, dada a natureza *cross-sectional* dos dados, bem como a periodicidade da PNS, podemos considerar w_i como constante, extraíndo-o do somatório tanto no numerador como denominador, eliminando-o, assim, do cálculo dessa estatística. Sob a hipótese nula da não-correlação entre os resíduos, W é distribuída assintoticamente como uma distribuição

qui-quadrado com 1 grau de liberdade. Na prática, para um nível de significância de 5%, o W_{critico} será igual a 3.84. Além da estatística W , originalmente utilizada pelos autores, a existência da correlação entre os vetores dos resíduos estimados pode ser corroborada por outros coeficientes de correlação, bem como pelo grau de significância estatística desses coeficientes, como os coeficientes de Pearson, Spearman e Tau de Kendall (HEUVEL; ZHAN, 2022).

Como mencionado anteriormente, Alves (2004) e Resende e Zeidan (2010), que aplicaram esse teste ao caso brasileiro, adotaram um desvio do modelo canônico ao incluir em X_i uma série de variáveis exógenas que não podem ser legalmente utilizadas pela companhia de seguros para determinação do prêmio. Esse procedimento diverge do que sugere a bibliografia, e é pouco coerente com a formalização teórica do modelo. Neste trabalho, portanto, só serão incluídas na matriz de variáveis exógenas X_i variáveis categóricas representando as 10 faixas de idade em que a legislação brasileira autoriza diferenciação do prêmio.

3.2 Extensões do teste de Correlação Positiva

3.2.1 Heterogeneidade preferências individuais e variáveis não-utilizadas

Complementando o trabalho dos autores mencionados acima, este projeto também se propõe a direcionar um foco especial às extensões do teste de correlação positiva, em especial no que concerne à heterogeneidade das preferências individuais e à inclusão de variáveis não-utilizadas pela companhia de seguro nas regressões.

Um potencial problema do modelo canônico é que ele não faz diferença entre a Informação Privada proveniente dos diferentes perfis de risco das diferenças na aversão ao risco dos indivíduos. Formalmente, sendo os resíduos dos modelos canônicos ϵ e η , se existir tanto o problema dos diferentes perfis de risco (capturados por X_1) como diferentes aversões ao risco (X_2), podemos definí-los como:

$$\epsilon_i = X_{1,i}\omega_1 + X_{2,i}\omega_2 + \epsilon'_i \quad (3.4)$$

$$\eta_i = X_{1,i}\pi_1 + X_{2,i}\pi_2 + \eta'_i \quad (3.5)$$

Dessa forma, se existir algum tipo de heterogeneidade nas preferências não levadas em conta pelo modelo, é possível que o resultado do teste de correlação positiva leve a uma conclusão errada. Finkelstein e Poterba (2014), por exemplo, apontam a possibilidade da ocorrência do Erro Tipo II (não rejeita H_0 de informação simétrica, porém H_0 é falsa) em

uma situação em que a aversão ao risco é positivamente correlacionada com a *Cobertura*, porém negativamente correlacionada com o *Sinistro*.

A primeira extensão que busca endereçar esse problema é a proposta por [Cutler, Finkelstein e McGarry \(2008\)](#), e consiste na inclusão de variáveis que capturem características individuais relacionadas à aversão ao risco. A ideia dessa metodologia é controlar o modelo por fatores que possam gerar algum tipo de endogeneidade nos modelos devido a diferenças nas preferências individuais. Seria razoável supor, por exemplo, que um indivíduo fumante teria uma aversão menor ao risco em relação à sua saúde. Dessa forma, o ganho de utilidade desse indivíduo na contratação de um seguro seria menor do que o de um indivíduo não-fumante. A variável *fumante*, nesse caso, poderia ser modelada como um comportamento. Utilizando a notação de função indicadora dos autores, podemos definir:

$$\mathbb{1}Cobertura_i = \beta_0 + \beta_1 \cdot Comportamento_i + X_i\Omega + \epsilon_i \quad (3.6)$$

$$Sinistro_i = \alpha_0 + \alpha_1 \cdot Comportamento_i + X_i\Pi + \eta_i \quad (3.7)$$

A segunda extensão do teste de correlação positiva é a proposta em [Finkelstein e Poterba \(2014\)](#), tratando-se do teste de "*unused observables*" mencionado na revisão bibliográfica. Adiciona-se ao modelo canônico uma matriz W_i de variáveis exógenas não-utilizadas pelas seguradoras.

Na prática, podemos observar que esse modelo é bastante similar ao da primeira extensão:

$$Cobertura_i = X_i\Omega + \alpha W_i + \epsilon_i \quad (3.8)$$

$$Sinistro_i = X_i\Pi + \delta W_i + \eta_i \quad (3.9)$$

Esse modelo também endereça o problema da heterogeneidade das preferências, mas tem algumas diferenças em relação ao anterior. Primeiramente, ele tem característica mais geral, dado que a variável W pode ter uma parte tanto em X_1 (perfil de risco) como em X_2 (aversão ao risco).

Além disso, nesse caso não se utiliza a estatística W . O teste proposto pelos autores consiste em verificar os coeficientes das variáveis não-utilizadas nas regressões. Rejeitar $\{\alpha = 0, \delta = 0\}$, nesse caso, equivale a rejeitar a hipótese nula de informação simétrica. Se tanto α como δ forem diferentes de zero, isso significa que existe correlação de uma variável não-utilizada pela seguradora tanto para o risco de perda como para a cobertura do seguro.

Da mesma forma, a análise permite a investigação de uma série de candidatos a W_i , o que pode ser interessante para determinar quais variáveis não-utilizadas pelas seguradoras geram informação assimétrica. No caso brasileiro, isso é especialmente interessante, dadas as restrições legais para diferenciação de prêmio, e dado que o nível de detalhamento da base de dados utilizada proporciona uma série de variáveis candidatas. Uma breve aplicação dessa extensão ao caso brasileiro, utilizando a PNS de 2013, é sugerida por [Fonseca \(2017\)](#), mas o autor utiliza esse método somente como teste de robustez na seção final de seus resultados.

A metodologia econométrica proposta tem como objetivo contribuir com a discussão dos trabalhos citados, verificando a existência ou não da informação assimétrica no mercado da saúde suplementar brasileiro. É importante destacar, contudo, que o teste de correlação positiva tem a limitação de não fazer distinção direta entre o fenômeno da seleção adversa e do risco moral. A diferenciação entre o problema informacional *ex-ante* e *ex-post* deve ser feita, portanto, por meio de uma discussão qualitativa, ou valendo-se de outras evidências empíricas contidas nos dados ([FINKELSTEIN; POTERBA, 2014](#)).

Além disso, o método é pouco flexível para modelar situações em que a variável resposta é ordinal, um problema apontado tanto por [Kim et al. \(2009\)](#) como por [Dardanoni, Forcina e Donni \(2018\)](#). Neste trabalho, na próxima seção, focaremos na abordagem proposta pelos primeiros autores para mitigar essa deficiência.

3.2.2 Medidas multinomiais de cobertura e sinistro

Como discutido no modelo canônico, a equação da Cobertura é definida pela função:

$$Cobertura_i = X_i\Omega + \epsilon_i \quad (3.10)$$

O teste de correlação positiva, discutido até aqui, tem uma limitação digna de nota. A estimação da Cobertura e do Sinistro por probits impõe a necessidade de que as variáveis dependentes sejam definidas de forma binária. [Kim et al. \(2009\)](#), contudo, propõem uma solução que combina um probit ordenado para a equação da Cobertura com uma regressão binomial negativa para a equação do Sinistro, que permite utilizar uma variável ordinal na primeira e uma variável quantitativa discreta na segunda.

Dessa forma, para iniciar a estimação do modelo definimos primeiro $Cobertura_i^m$ como:

$$Cobertura_i^m = \begin{cases} 0 & \text{se } Cobertura_i^{m*} < \mu_1 \\ 1 & \text{se } \mu_1 < Cobertura_i^{m*} < \mu_2 \\ 2 & \text{se } \mu_2 < Cobertura_i^{m*} \end{cases}$$

onde $Cobertura_i^{m*}$ é uma variável latente e μ_1 e μ_2 são limiares para as três categorias definidas acima.

Utilizando uma variável dependente multinomial ordenada de $Cobertura_i^m \in \{0, 1, 2\}$, a estimação dos resíduos se torna mais complexa do que no modelo canônico.

Primeiramente, os autores propõem o agrupamento das três escolhas de $Cobertura_i^m \in \{0, 1, 2\}$ em dois conjuntos de escolha ordenada. Definimos, portanto, $Cobertura_i^1 = 0$ se $Cobertura_i^m = 0$, e $Cobertura_i^1 = 1$ se $Cobertura_i^m \in \{1, 2\}$. Da mesma forma, seja $Cobertura_i^2 = 0$ se $Cobertura_i^m \in \{0, 1\}$ e $Cobertura_i^2 = 1$ se $Cobertura_i^m = 2$.

A partir daí torna-se mais intuitivo definir as probabilidades previstas para cada agrupamento. Baseado no trabalho de [Kim et al. \(2009\)](#), mais uma vez, definimos os resíduos generalizados como:

$$\hat{\epsilon}_i^{m_1} = \frac{\phi(X_i\Omega - \hat{\mu}_1)}{\Phi(X_i\Omega - \hat{\mu}_1)(1 - \Phi(X_i\Omega - \hat{\mu}_1))} [Cobertura_i^1 - \Phi(X_i\Omega)] \quad (3.11)$$

$$\hat{\epsilon}_i^{m_2} = \frac{\phi(X_i\Omega - \hat{\mu}_2)}{\Phi(X_i\Omega - \hat{\mu}_2)(1 - \Phi(X_i\Omega - \hat{\mu}_2))} [Cobertura_i^2 - \Phi(X_i\Omega)] \quad (3.12)$$

Em termos práticos, $\hat{\mu}_1$ e $\hat{\mu}_2$ são limiares estimados para as categorias observadas pelo modelo de probit ordenado. Nas equações dos resíduos generalizados, $\phi(\cdot)$ e $\Phi(\cdot)$ são as funções de densidade e distribuição acumulada da distribuição normal padrão, respectivamente, o que permite que sejam facilmente calculados computacionalmente.

Nessa extensão do modelo, é proposta uma definição do Sinistro como uma função tal que:

$$Sinistro_i = g(X_i, \hat{\epsilon}_i) + \eta_i \quad (3.13)$$

onde $Sinistro_i$ é uma variável quantitativa discreta, e não mais uma variável binária.

Nesse caso, os dois resíduos obtidos (3.11 e 3.12) são incluídos na equação de Sinistro como regressores, sendo que o coeficiente de regressão estimado de $\hat{\epsilon}_i^{m_1}$ captura o efeito da assimetria de informação na escolha entre nenhuma cobertura ($Cobertura_i^m = 0$) e uma cobertura intermediária ($Cobertura_i^m = 1$), e o de $\hat{\epsilon}_i^{m_2}$ captura o efeito da assimetria de informação na escolha entre uma cobertura intermediária ($Cobertura_i^m = 1$) e uma cobertura completa ($Cobertura_i^m = 2$).

Há algumas formas possíveis de estimar essa equação. A forma mais elementar de se modelar uma variável discreta (*count variable*) seria utilizando uma Regressão de Poisson. Nesse modelo, para a equação do Sinistro, a partir de [Wooldridge \(2003\)](#), podemos definir a função de log-verossimilhança como:

$$\mathcal{L}(\Pi) = \sum_{i=1}^n \ell_i(\Pi) = \sum_{i=1}^n (S_i x_i \Pi - \exp(x_i \Pi)) \quad (3.14)$$

Os estimadores podem, então, ser obtidos maximizando a função de log-verossimilhança.

Deve-se, contudo, ressaltar que a Regressão de Poisson assume uma hipótese bastante restritiva de que todos os momentos da distribuição são iguais à média. Em particular, isso significa assumir que $E(S|X) = Var(S|X)$. Para lidar com essa limitação, uma análise de quase-máxima verossimilhança pode ser utilizada, de fácil execução por meio de linguagens de programação econométricas. Particularmente, no caso da QMLE (quasi-maximum likelihood estimation), assume-se que:

$$Var(S|X) = \varphi E(S|X) \quad (3.15)$$

onde $\varphi > 0$ é um parâmetro desconhecido ([WOOLDRIDGE, 2003](#)).

Uma solução ainda mais geral, utilizada pelo artigo de [Kim et al. \(2009\)](#), é a Regressão Binomial Negativa. A definição desse modelo assume que o primeiro e segundo momentos são dados, respectivamente, por:

$$E(S|X) = \mu \quad (3.16)$$

$$Var(S|X) = \mu + \alpha \mu^2 \quad (3.17)$$

onde α é o parâmetro da sobredispersão ([HILBE, 2011](#)). Dessa forma, enquanto a variância do modelo Poisson QMLE é definida por uma função linear de μ , a variância da Regressão Binomial Negativa é função quadrática de μ . Neste trabalho, utilizaremos a Regressão Binomial Negativa, aproximando-nos da metodologia empregada pelos autores.

4 Resultados

4.1 PNS 2019: Uma visão geral

Como mencionado na seção anterior, a PNS de 2019 contém 293.726 observações de indivíduos entrevistados em 108.475 domicílios.

Desse espaço amostral, 14.344 não responderam à pergunta "o(a) Sr(a) tem algum plano de saúde médico particular, de empresa ou órgão público?". Para efeito das regressões, essas observações serão tratadas como nulas e retiradas do estudo.

Sobre essa mesma pergunta, 58.597 indivíduos (26.5%) responderam positivamente à pergunta se tinham alguma cobertura de plano de saúde médico, uma porcentagem em linha com os dados da ANS desse mesmo período. De acordo com a ANS, uma parcela significativa dos planos são da modalidade coletiva empresarial (cerca de 70%). Essa informação está contemplada nos microdados da PNS, e será incluída em uma das extensões do teste econométrico, como motivação para a utilização de um modelo multinomial.

4.2 Teste de Correlação Positiva

4.2.1 O modelo canônico

A implementação do teste de correlação positiva nos dados da PNS 2019, bem como das extensões do teste, foi conduzida utilizando sobretudo a biblioteca *statsmodels* em Python. A matriz X foi segmentada com base nos grupos etários, respeitando as restrições legais, e dividida em faixas de idade definidas como 0-18 anos, 19-23 anos, 24-28 anos, 29-33 anos, 34-38 anos, 39-43 anos, 44-48 anos, 49-53 anos, 54-59 anos e 59 anos ou mais. Posteriormente, foram realizadas duas regressões probits, ambas demonstrando significância estatística.

É importante relembrar aqui a natureza binomial do modelo canônico. Dessa forma, faz-se necessário construir as variáveis *Cobertura* e *Sinistro*. A variável *dummy* de cobertura assumirá o valor 1 pela resposta positiva à pergunta "o(a) Sr(a) tem algum plano de saúde médico particular, de empresa ou órgão público?", e 0 em caso contrário. Já a variável *Sinistro* utilizará o mesmo limiar definido por [Fonseca \(2017\)](#), que define como sinistralidade àqueles que respondem à pergunta "Quantas vezes consultou um médico nos últimos 12 meses?" com um número igual ou superior a 3. É fácil perceber que esse limiar é arbitrário e pouco informativo sobre a real sinistralidade dos indivíduos, fato que será

Tabela 1 – Resultado da Regressão Probit da Cobertura para o modelo canônico

Dep. Variable:	cobertura_plano	No. Observations:	279382	
converged:	True	LL-Null:	-1.4349e+05	
Covariance Type:	nonrobust	LLR p-value:	0.000	
variable	coef	std err	z	P> z
Intercept	-0.9381	0.006	-170.010	0.000
C(age_group)[T.19-23]	-0.0597	0.012	-5.168	0.000
C(age_group)[T.24-28]	0.0421	0.012	3.634	0.000
C(age_group)[T.29-33]	0.1615	0.011	14.187	0.000
C(age_group)[T.34-38]	0.2212	0.011	20.167	0.000
C(age_group)[T.39-43]	0.1965	0.011	17.704	0.000
C(age_group)[T.44-48]	0.1701	0.012	14.693	0.000
C(age_group)[T.49-53]	0.1795	0.012	15.311	0.000
C(age_group)[T.54-59]	0.2143	0.011	19.134	0.000
C(age_group)[T.59+]	0.2747	0.008	32.829	0.000

Tabela 2 – Resultado da Regressão Probit do Sinistro para o modelo canônico

Dep. Variable:	sinistro	No. Observations:	279382		
converged:	True	LL-Null:	-1.8223e+05		
Covariance Type:	nonrobust	LLR p-value:	0.000		
<hr/>					
coef		std err	z	P> z	
<hr/>					
Intercept		-0.4558	0.005	-93.656	0.000
C(age_group)[T.19-23]		-0.2007	0.010	-19.398	0.000
C(age_group)[T.24-28]		-0.1082	0.011	-10.299	0.000
C(age_group)[T.29-33]		-0.0356	0.011	-3.389	0.001
C(age_group)[T.34-38]		-0.0063	0.010	-0.615	0.539
C(age_group)[T.39-43]		0.0134	0.010	1.309	0.191
C(age_group)[T.44-48]		0.0679	0.011	6.429	0.000
C(age_group)[T.49-53]		0.1648	0.011	15.508	0.000
C(age_group)[T.54-59]		0.2593	0.010	25.543	0.000
C(age_group)[T.59+]		0.4653	0.008	61.431	0.000

melhor explorado na seção do modelo multinomial deste trabalho.

Seguindo a metodologia, foi conduzida uma análise da correlação entre os resíduos das duas regressões probits realizadas, revelando coeficientes de correlação significativos.

O cálculo da estatística W proposta por [Chiappori e Salanie \(2000\)](#), efetuado computacionalmente utilizando os vetores dos resíduos, traz um valor de $W = 6092.7$, muito superior ao valor crítico de 3.84. Corroborando esse resultado, para o coeficiente de Pearson, a correlação encontrada foi de 0.1585, para o coeficiente de Spearman 0.2644, e para o Tau de Kendall 0.2160. Todos esses valores apresentaram p-valor baixo, indicando uma correlação estatisticamente significativa entre os resíduos das regressões.

Os resultados das regressões acima, bem como a significância estatística deles, permitem afirmar que existem fortes evidências da presença de Informação Assimétrica no

mercado de saúde suplementar brasileiro.

4.2.2 Heterogeneidade das Preferências e Variáveis não-utilizadas

Para o estudo dessas duas extensões do teste de correlação positiva, será proposta a utilização da variável *fumante*. Essa variável será construída a partir da seção mais detalhada da PNS, com um espaço amostral de 90.846 indivíduos. Para efeito da construção da variável *dummy*, será considerado fumante àquele indivíduo que respondeu positivamente à afirmação "Atualmente, o(a) Sr(a) fuma algum produto do tabaco?" ou "E no passado, o(a) Sr(a) fumou algum produto do tabaco diariamente?".

Para a primeira extensão, teremos o seguinte modelo:

$$\mathbb{1}Cobertura_i = \beta_0 + \beta_1 \cdot Fumante_i + X_i\Gamma + \epsilon_i \quad (4.1)$$

$$Sinistro_i = \alpha_0 + \alpha_1 \cdot Fumante_i + X_i\Pi + \eta_i \quad (4.2)$$

A hipótese aqui seria de que um indivíduo fumante seria menos avesso ao risco em relação à saúde, e atribuiria menor utilidade a um plano de saúde, por exemplo. A ideia da inclusão da variável fumante, portanto, seria a de controlar o modelo por diferentes aversões ao risco dos indivíduos.

Repetindo as regressões acima com esse modelo, chegamos a um $W = 1657.7$, o que permite concluir que o fenômeno da Informação Assimétrica permanece com significância estatística.

Relembrando a definição das duas extensões na seção 3.2.1, podemos utilizar, de forma mais geral, a mesma regressão dentro do contexto da segunda extensão. Dessa forma, consideraremos a variável fumante como variável não-utilizada, uma primeira candidata dentre uma série de variáveis que a granularidade da PNS 2019 permite que testemos. Além disso, a escolha dessa segunda extensão também implica em não definir *fumante* tão-somente como uma *proxy* da aversão ao risco do indivíduo, mas sim como uma variável que pode contribuir tanto para a heterogeneidade das preferências como para a informação privada advinda de diferentes perfis de risco.

$$Cobertura_i = X_i\Gamma + \alpha fumante_i + \epsilon_i \quad (4.3)$$

$$Sinistro_i = X_i\Pi + \delta fumante_i + \eta_i \quad (4.4)$$

Nesse caso, não será utilizada a estatística W . Para esse modelo, rejeitar $\{\alpha = 0, \delta = 0\}$ equivale a rejeitar a hipótese nula de informação simétrica.

Tabela 3 – Resultado da Regressão Probit da Cobertura - variável fumante

Dep. Variable:	cobertura_plano	No. Observations:	90846	
converged:	True	LL-Null:	-48593.	
Covariance Type:	nonrobust	LLR p-value:	1.456e-230	
variable	coef	std err	z	P> z
Intercept	-1.0401	0.032	-32.426	0.000
C(age_group)[T.19-23]	0.0486	0.038	1.275	0.202
C(age_group)[T.24-28]	0.1564	0.037	4.267	0.000
C(age_group)[T.29-33]	0.3094	0.036	8.682	0.000
C(age_group)[T.34-38]	0.3692	0.035	10.508	0.000
C(age_group)[T.39-43]	0.3763	0.035	10.682	0.000
C(age_group)[T.44-48]	0.3270	0.036	9.175	0.000
C(age_group)[T.49-53]	0.3122	0.036	8.726	0.000
C(age_group)[T.54-59]	0.3630	0.035	10.334	0.000
C(age_group)[T.59+]	0.4235	0.033	12.733	0.000
smoking_status	-0.3620	0.015	-23.566	0.000

Tabela 4 – Resultado da Regressão Probit do Sinistro - variável fumante

Dep. Variable:	sinistro	No. Observations:	90846	
converged:	True	LL-Null:	-61082.	
Covariance Type:	nonrobust	LLR p-value:	0.000	
variable	coef	std err	z	P> z
Intercept	-0.7042	0.029	-24.613	0.000
C(age_group)[T.19-23]	0.2163	0.034	6.428	0.000
C(age_group)[T.24-28]	0.2637	0.033	8.073	0.000
C(age_group)[T.29-33]	0.3070	0.032	9.589	0.000
C(age_group)[T.34-38]	0.3160	0.032	9.994	0.000
C(age_group)[T.39-43]	0.3320	0.032	10.471	0.000
C(age_group)[T.44-48]	0.4041	0.032	12.645	0.000
C(age_group)[T.49-53]	0.4895	0.032	15.301	0.000
C(age_group)[T.54-59]	0.5853	0.031	18.620	0.000
C(age_group)[T.59+]	0.7559	0.030	25.413	0.000
smoking_status	-0.2072	0.013	-15.887	0.000

Os resultados obtidos permitem que rejeitemos $\{\alpha = 0, \delta = 0\}$ para qualquer nível de significância usual, de forma que podemos identificar a variável *fumante* como fonte de Informação Assimétrica.

4.2.3 Testando outras variáveis não-utilizadas

Como discutido anteriormente, uma possível aplicação da segunda extensão do teste de correlação positiva é a verificação de demais variáveis não-utilizadas pelas companhias de seguro como potenciais fontes de Informação Assimétrica.

Para a escolha das variáveis candidatas, serão escolhidas características individuais que não são utilizadas para determinação do prêmio por razões institucionais. A legislação brasileira proíbe a seleção de risco e, para endereçar essa limitação, testaremos o diagnóstico de diabetes - uma síndrome metabólica associada a uma série de complicações crônicas - como variável candidata. Além disso, é também de grande relevância no contexto institucional do mercado de saúde suplementar brasileiro a impossibilidade de reajuste após os 60 anos de idade, o que será abordado por meio da inclusão de faixas de idade adicionais que contemplem esse aumento do risco de sinistro. Na figura 3, é possível observar que ocorre um acréscimo significativo na média de sinistros nas faixas etárias a partir dos 60 anos.

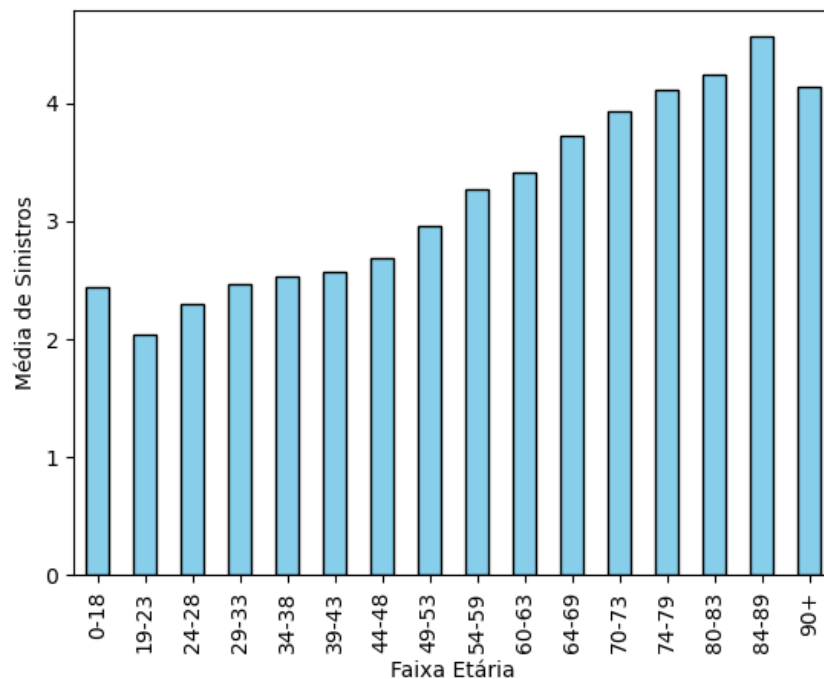


Figura 3 – Média de Sinistros por faixa etária, com faixas adicionais

Fonte: Elaboração Própria a partir dos dados da PNS 2019

Utilizando os dados da PNS 2019, podemos agora testar na matriz X_i as variáveis diabéticas, além de variáveis de idade que diferenciem a partir dos 60 anos. O espaço amostral para essa primeira variável é de 84.073 respondentes à pergunta "Algum médico já lhe deu o diagnóstico de diabetes?", à qual 7.374 entrevistados (8.77%) responderam positivamente. Para as demais faixas de idade, definimos as variáveis categóricas a partir

da mesma variável utilizada no modelo canônico, agora com um número maior de faixas de idade.

Tabela 5 – Resultado da Regressão Probit da Cobertura - diabetes e faixas de idade adicionais

Dep. Variable:	cobertura_plano	No. Observations:	84073	
converged:	True	LL-Null:	-46357.	
Covariance Type:	nonrobust	LLR p-value:	1.194e-73	
variable	coef	std err	z	P> z
Intercept	-0.9314	0.035	-26.263	0.000
C(age_group)[T.19-23]	-0.0183	0.042	-0.438	0.661
C(age_group)[T.24-28]	0.0683	0.040	1.709	0.087
C(age_group)[T.29-33]	0.2221	0.039	5.709	0.000
C(age_group)[T.34-38]	0.2790	0.038	7.266	0.000
C(age_group)[T.39-43]	0.2819	0.038	7.330	0.000
C(age_group)[T.44-48]	0.2187	0.039	5.629	0.000
C(age_group)[T.49-53]	0.1944	0.039	4.991	0.000
C(age_group)[T.54-59]	0.2356	0.038	6.146	0.000
C(age_group)[T.60-63]	0.2387	0.040	5.949	0.000
C(age_group)[T.64-69]	0.2656	0.039	6.792	0.000
C(age_group)[T.70-73]	0.2950	0.042	7.005	0.000
C(age_group)[T.74-79]	0.3389	0.042	8.137	0.000
C(age_group)[T.80-83]	0.4007	0.048	8.264	0.000
C(age_group)[T.84-89]	0.4315	0.051	8.418	0.000
C(age_group)[T.90+]	0.4897	0.066	7.389	0.000
diabetes_status	-0.0294	0.017	-1.697	0.090

Tabela 6 – Resultado da Regressão Probit do Sinistro - diabetes e faixas de idade adicionais

Dep. Variable:	sinistro	No. Observations:	84073	
converged:	True	LL-Null:	-57130.	
Covariance Type:	nonrobust	LLR p-value:	0.000	
variable	coef	std err	z	P> z
Intercept	-0.6105	0.032	-18.904	0.000
C(age_group)[T.19-23]	0.1744	0.038	4.642	0.000
C(age_group)[T.24-28]	0.1955	0.036	5.385	0.000
C(age_group)[T.29-33]	0.2271	0.036	6.381	0.000
C(age_group)[T.34-38]	0.2344	0.035	6.662	0.000
C(age_group)[T.39-43]	0.2308	0.035	6.548	0.000
C(age_group)[T.44-48]	0.3005	0.035	8.473	0.000
C(age_group)[T.49-53]	0.3607	0.035	10.163	0.000
C(age_group)[T.54-59]	0.4268	0.035	12.207	0.000
C(age_group)[T.60-63]	0.4635	0.037	12.655	0.000
C(age_group)[T.64-69]	0.5127	0.036	14.353	0.000
C(age_group)[T.70-73]	0.5769	0.039	14.931	0.000
C(age_group)[T.74-79]	0.6291	0.038	16.417	0.000
C(age_group)[T.80-83]	0.7247	0.045	15.980	0.000
C(age_group)[T.84-89]	0.7284	0.048	15.114	0.000
C(age_group)[T.90+]	0.7507	0.064	11.817	0.000
diabetes_status	0.5471	0.016	33.613	0.000

Podemos notar que a variável *diabético* não passa no teste de variáveis não-utilizadas. Ela é estatisticamente significativa para a regressão do Sinistro (ser diabético correlacionado com maior utilização do plano de saúde), mas não é estatisticamente significativa para a regressão da Cobertura. Dessa forma, não é possível rejeitar a hipótese nula da informação simétrica para essa variável.

As divisões em faixas etárias para além do que é permitido atualmente pela legislação são estatisticamente significantes para ambas as regressões e, portanto, de acordo com a extensão de [Finkelstein e Poterba \(2014\)](#), são uma fonte de informação assimétrica.

4.2.4 Uma proposta de modelo multinomial para a Cobertura

Como discutido previamente, [Kim et al. \(2009\)](#) propõem um modelo multinomial para a cobertura do seguro, de forma a contemplar a possibilidade de escolha do segurado dentre um menu de coberturas opcionais. No caso do artigo dos autores, o mercado analisado é o de seguros automobilísticos.

Neste trabalho, propomos a aplicação do modelo desses autores ao caso da PNS 2019. No caso do seguro saúde no Brasil, é sabido que uma parte significativa das apólices é paga pelo empregador. Uma outra parte, contudo, é inteiramente paga pelo indivíduo. Essa peculiaridade da saúde suplementar é potencialmente um problema para a verificação

da Informação Assimétrica, dado que haverá diferença na utilidade do indivíduo sendo ele o responsável pelo pagamento da apólice ou não, para uma mesma cobertura. Se a diferenciação entre os tipos de apólice não é feita, o resultado do teste de correlação positiva pode levar a conclusões equivocadas, visto que a decisão de aderir ou não ao plano não necessariamente partiu do próprio indivíduo.

Para mitigar esse problema, a cobertura da saúde suplementar é modelada pela proposta multinomial de [Kim et al. \(2009\)](#). Sendo $Cobertura_i^m$ a variável que representa diferentes tipos de cobertura, podemos definir $Cobertura_i^m = 0$, para indivíduos sem cobertura, $Cobertura_i^m = 1$ para indivíduos com cobertura não diretamente paga por eles, e $Cobertura_i^m = 2$ para indivíduos cuja cobertura é paga por eles mesmos.

Para construir essa variável, utilizaremos a pergunta da PNS 2019 "Quem paga a mensalidade deste plano de saúde", e atribuiremos o valor 2 para a variável somente para aqueles que responderam "Somente o titular ou outro morador do domicílio".

Construída essa variável, é possível estimar um modelo de probit ordenado utilizando as mesmas variáveis independentes do modelo canônico. A implementação do modelo é feita por meio da linguagem R.

Tabela 7 – Modelo Multinomial - Regressão Probit Ordenado da Cobertura

variable	coef	std error	t
C(age_group)[T.19-23]	0.5828	0.02101	27.74
C(age_group)[T.24-28]	0.9986	0.01882	53.06
C(age_group)[T.29-33]	1.1975	0.01806	66.30
C(age_group)[T.34-38]	1.2542	0.01757	71.40
C(age_group)[T.39-43]	1.2347	0.01762	70.07
C(age_group)[T.44-48]	1.2010	0.01801	66.69
C(age_group)[T.49-53]	1.2073	0.01804	66.94
C(age_group)[T.54-58]	1.2580	0.01741	72.27
C(age_group)[T.59+]	1.3495	0.01502	89.84
intercept	coef	std err	t
0 1	2.1203	0.0133	159.2695
1 2	2.5655	0.0137	187.3893

Os t-valores nos permitem aceitar que existe significância estatística nos resultados obtidos. Seguindo o modelo proposto pelos autores, a partir dos interceptos estimados pelo Probit Ordenado, podemos calcular os resíduos generalizados a partir das seguintes equações:

$$\hat{\epsilon}_i^{m_1} = \frac{\phi(X_i\Omega - \hat{\mu}_1)}{\Phi(X_i\Omega - \hat{\mu}_1)(1 - \Phi(X_i\Omega - \hat{\mu}_1))} [Cobertura_i^1 - \Phi(X_i\Omega)] \quad (4.5)$$

$$\hat{\epsilon}_i^{m_2} = \frac{\phi(X_i\Omega - \hat{\mu}_2)}{\Phi(X_i\Omega - \hat{\mu}_2)(1 - \Phi(X_i\Omega - \hat{\mu}_2))} [Cobertura_i^2 - \Phi(X_i\Omega)] \quad (4.6)$$

sendo $\hat{\mu}_1 = 2.1203$ e $\hat{\mu}_2 = 2.5655$.

Como discutido na seção de Metodologia, para permitir a utilização de uma variável quantitativa discreta na equação de Sinistro, utilizaremos uma Regressão Binomial Negativa, um modelo mais geral do que a Regressão de Poisson.

Tabela 8 – Modelo Multinomial - Regressão Binomial Negativa do Sinistro utilizando resíduos calculados como regressores

variable	coef	std err	z	Pr(> z)
(Intercept)	1.147558	0.004013	285.976	< 2e-16 ***
C(age_group)[T.19-23]	-0.048826	0.008397	-5.815	6.07e-09 ***
C(age_group)[T.24-28]	0.036746	0.008464	4.342	1.41e-05 ***
C(age_group)[T.29-33]	0.064822	0.008414	7.704	1.32e-14 ***
C(age_group)[T.34-38]	0.068681	0.008127	8.451	< 2e-16 ***
C(age_group)[T.39-43]	0.064389	0.008131	7.919	2.39e-15 ***
C(age_group)[T.44-48]	0.090464	0.008327	10.864	< 2e-16 ***
C(age_group)[T.49-53]	0.153246	0.008277	18.515	< 2e-16 ***
C(age_group)[T.54-59]	0.216001	0.007809	27.661	< 2e-16 ***
C(age_group)[T.59]+	0.302227	0.005808	52.034	< 2e-16 ***
residuo especial1	0.072102	0.001561	46.177	< 2e-16 ***
residuo especial2	-0.005298	0.002449	-2.163	0.0305 *

Podemos verificar, mais uma vez, que há significância estatística, para um nível de significância $\alpha = 5\%$, para os dois resíduos utilizados como regressores. O modelo utilizado não só foi adequado, permitindo uma modelagem mais completa da Cobertura e do Sinistro, mas também reiterou o resultado obtido nas regressões anteriores de que existe o fenômeno da Informação Assimétrica no mercado de saúde suplementar brasileiro.

4.3 Risco Moral vs. Seleção Adversa: uma breve discussão

Como discutido anteriormente, o teste de correlação positiva é robusto na identificação da informação privada sobre o tipo de risco dos indivíduos e, portanto, robusto para verificação da Informação Assimétrica nos mercados de seguro. Todavia, é incapaz de verificar se essa assimetria está associada ao fenômeno da Seleção Adversa ou do Risco Moral.

The unused observables test, like the positive correlation test, is a test for asymmetric information. Without additional information, rejecting the null hypothesis of symmetric information **does not offer evidence on the question of whether asymmetric information results from moral hazard or from selection**. In some cases, such additional information may be available. For example, when a researcher has evidence

suggesting that an unused observable variable is correlated with the risk of loss even among individuals who have identical insurance coverage, then finding that individuals with certain values of the unused observable select more insurance coverage supports the presence of selection based on ex-ante private information rather than moral hazard based on ex-post private information. (FINKELSTEIN; POTERBA, 2014)

Os autores, no texto acima, defendem que, havendo informação adicional que permita concluir que uma variável não-utilizada é correlacionada com o risco de sinistro, mesmo em indivíduos com exatamente a mesma cobertura de seguro, haveria evidência da informação privada *ex-ante*, isto é, do fenômeno da seleção adversa. No caso das variáveis não-utilizadas com significância estatística discutidas nos resultados anteriores, essa correlação é bastante trivial. É inconteste a correlação positiva entre utilização de serviços de saúde e idade, bem como entre a utilização de serviços de saúde e o tabagismo (KALSETH; HALVORSEN, 2020; SIMONS et al., 2023). De acordo com os autores do teste das variáveis não-utilizadas, isso seria uma evidência a favor do caráter *ex-ante* da informação privada identificada pelo teste econométrico.

É importante ressaltar, contudo, que dado o enfoque deste trabalho na abordagem econométrica e, principalmente, dada a natureza eminentemente qualitativa da diferenciação proposta acima, reconhecemos essa como uma limitação desta monografia.

Conclusão

As evidências encontradas neste estudo corroboram, com significância estatística, a hipótese da presença de Informação Assimétrica no mercado de saúde suplementar no Brasil. As variáveis *fumante* e as faixas de idade adicionais são identificadas como fontes dessa falha de mercado, de acordo com os resultados do teste de correlação positiva e suas extensões. Essas variáveis, não utilizadas para a diferenciação dos prêmios, demonstram serem relevantes para esse problema informacional.

Surpreendentemente, a variável *diabético* não teve o mesmo resultado. Embora tenha sido encontrada significância estatística para a equação de sinistro, o mesmo não ocorreu para a cobertura. Esse resultado sugere que a informação privada sobre o estado de risco dos diabéticos não é utilizada na decisão de aderir a uma apólice. Não existem evidências, portanto, para rejeitar a hipótese nula de informação simétrica para essa variável.

Um modelo multinomial, utilizando uma Regressão Binomial Negativa, também indicou a presença significativa de Informação Assimétrica, evidenciada pela significância estatística dos regressores dos dois resíduos. Este resultado é promissor, pois abre caminho para uma modelagem mais completa que permita a diferenciação entre tipos de cobertura, além de possibilitar uma análise do sinistro para além de uma abordagem binária.

A diferenciação entre Seleção Adversa e Risco Moral foi brevemente discutida neste trabalho, mas também reconhecida como uma limitação relevante do teste de correlação positiva. Desenvolver de forma empírica o argumento qualitativo de [Finkelstein e Poterba \(2014\)](#), exposto na subseção 4.3, seria certamente um objetivo louvável para uma futura pesquisa.

A PNS compreende uma quantidade imensa de variáveis. Pelo escopo desse trabalho, e para evitar um prolongamento excessivo dessa investigação, apenas um número seletivo dessas variáveis foi testada como fonte de Informação Assimétrica. Um trabalho que explore variáveis adicionais, até mesmo utilizando outras bases de dados, seria certamente bem-vindo.

Finalmente, a modelagem envolvendo mais de um tipo de cobertura também foi aqui proposta como aplicação original dessa extensão do teste econométrico ao caso da PNS. Essa metodologia pode certamente ser explorada mais a fundo por outros trabalhos, assim como pode ser aplicada a outros mercados de seguro que não o da saúde suplementar.

Os resultados aqui obtidos, bem como a constatação da grave conjuntura que afeta a saúde suplementar brasileira, suscitam a lembrança da simulação iterativa de [Winssen](#),

Kleef e Ven (2018) e do inevitável colapso de um mercado com Seleção Adversa. Uma maior possibilidade de diferenciação do prêmio poderia mitigar essa falha de mercado da saúde suplementar? Quanto dessa crise sendo vivenciada pela saúde suplementar pode, de fato, ser atribuída ao problema da Informação Assimétrica? Até que ponto as companhias de seguro devem poder selecionar o risco dos segurados?

As respostas para essas perguntas ultrapassam em muito o escopo deste trabalho e, a bem da verdade, parecem depender não só de conceitos econômicos, mas também de conceitos morais e das concepções pessoais de cada cidadão. É, todavia, a partir dessas indagações que esperamos que este estudo possa oferecer alguma contribuição. As transformações demográficas, laborais e tecnológicas vivenciadas pelo país devem tornar esse tema cada vez mais imperativo, bem como devem obrigar a sociedade brasileira a tomar decisões cada vez mais difíceis.

Referências

- AKERLOF, G. A. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, v. 84(3), p. 488–500, 1970. Citado 2 vezes nas páginas 13 e 16.
- ALVES, S. L. Estimando seleção adversa em planos. *Revista Economia*, 2004. Citado 2 vezes nas páginas 17 e 21.
- ANS. Súmula normativa 27/2015 da agência nacional de saúde suplementar. v. 11 de junho de 2015, 2015. Citado na página 12.
- ARROW, K. J. Uncertainty and the welfare economics of medical care. *The American Economic Review*, American Economic Association, v. 53, n. 5, p. 941–973, 1963. ISSN 00028282. Disponível em: <<http://www.jstor.org/stable/1812044>>. Citado 2 vezes nas páginas 14 e 16.
- CARDON, J. H.; HENDEL, I. Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, v. 32(3), p. 408–427, 2001. Citado na página 17.
- CHIAPPORI, P.-A.; SALANIE, B. Testing for asymmetric information in insurance markets. *Journal of Political Economy*, v. 108(1), p. 56–78, 2000. Disponível em: <<https://doi.org/10.1086/262111>>. Citado 5 vezes nas páginas 5, 6, 12, 16 e 27.
- COHEN, A.; SIEGELMANN, P. Testing for adverse selection in insurance markets. *Journal of Risk and Insurance*, v. 77(1), p. 39–84, 2010. Citado na página 16.
- Conselho Nacional de Justiça. *Saúde suplementar pontua impacto de processos judiciais para equilíbrio do setor*. 2024. Acessado em: 5 de junho de 2024. Disponível em: <cnj.jus.br/saude-suplementar-pontua-impacto-de-processos-judiciais-para-equilibrio-do-setor/>. Citado na página 12.
- CUTLER, D. M.; FINKELSTEIN, A.; MCGARRY, K. Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. *The American Economic Review*, v. 98(2), p. 158–161, 2008. Citado 2 vezes nas páginas 17 e 22.
- CUTLER, D. M.; ZECKHAUSER, R. J. The anatomy of health insurance. *Handbook of Health Economics*, v. 1, p. 563–643, 2000. Citado na página 16.
- DARDANONI, V.; FORCINA, A.; DONNI, P. L. Testing for asymmetric information in insurance markets: A multivariate ordered regression approach. *The Journal of Risk and Insurance*, v. 85(1), p. 107–125, 2018. Citado 2 vezes nas páginas 18 e 23.
- EINAV, L.; FINKELSTEIN, A. Selection in insurance markets: theory and empirics in pictures. *Journal of Economic Perspectives*, v. 25, n. 1, p. 115–138, Winter 2011. Citado na página 15.
- FINKELSTEIN, A.; POTERBA, J. Testing for asymmetric information using “unused observables” in insurance markets: Evidence from the u.k. annuity market. *Journal of*

- Risk and Insurance*, v. 81(4), p. 709–734, 2014. Citado 10 vezes nas páginas 5, 6, 15, 18, 21, 22, 23, 32, 35 e 36.
- FONSECA, R. B. d. A. Informational frictions in the brazilian health insurance market. *Dissertação de Mestrado submetida à EPGE*, 2017. Citado 4 vezes nas páginas 17, 18, 23 e 26.
- HEUVEL, E. van den; ZHAN, Z. Myths about linear and monotonic associations: Pearson's r, spearman's , and kendall's . *The American Statistician*, Taylor & Francis, v. 76, n. 1, p. 44–52, 2022. Disponível em: <<https://doi.org/10.1080/00031305.2021.2004922>>. Citado na página 21.
- HILBE, J. M. *Negative Binomial Regression*. 2nd. ed. Cambridge: Cambridge University Press, 2011. ISBN 978-0521198158. Citado na página 25.
- KALSETH, J.; HALVORSEN, T. Health and care service utilisation and cost over the life-span: a descriptive analysis of population data. *BMC Health Services Research*, v. 20, n. 1, p. 435, 2020. Disponível em: <<https://doi.org/10.1186/s12913-020-05295-2>>. Citado na página 35.
- KIM, H. et al. Evidence of asymmetric information in the automobile insurance market: Dichotomous versus multinomial measurement of insurance coverage. *The Journal of Risk and Insurance*, v. 76(2), p. 343–366, 2009. Citado 8 vezes nas páginas 5, 6, 18, 23, 24, 25, 32 e 33.
- MUGNATTO, S. *Aumentam reclamações de consumidores sobre cancelamentos unilaterais de planos de saúde*. 2024. Acessado em: 5 de junho de 2024. Disponível em: <camara.leg.br/noticias/1062863-aumentam-reclamacoes-de-consumidores/-sobre-cancelamentos-unilaterais-de-planos-de-saude/>. Citado na página 12.
- NISHIJIMA, M.; POSTALI, F. A. S.; FAVA, V. L. Consumo de serviços médicos e marco regulatório no mercado de seguro de saúde brasileiro. *Revista Pesquisa e Planejamento Econômico*, v. 41(3), p. 509–531, 2011. Citado na página 17.
- OCKÉ-REIS, C. O.; FIUZA, E. P. S.; COIMBRA, P. H. H. Inflação dos planos de saúde 2000-2018. *Nota Técnica do Instituto de Pesquisa Econômica Aplicada (IPEA)*, v. 54, 2019. Citado na página 11.
- PIOVEZAN, S. Reclamações contra planos de saúde disparam e chegam a quase 900 por dia. *Folha de São Paulo*, v. 09 de Setembro de 2023, 2023. Citado na página 12.
- RESENDE, M.; ZEIDAN, R. Adverse selection in the health insurance market: some empirical evidence. *The European journal of health economics*, v. 11(4), p. 413–418, 2010. Citado 2 vezes nas páginas 17 e 21.
- ROBERTSON, C. T. et al. Distinguishing moral hazard from access for high-cost healthcare under insurance. *PLoS ONE*, v. 15, n. 4, p. e0231768, 2020. Citado na página 14.
- ROTSCHILD, M.; STIGLITZ, J. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, v. 90(4), p. 629–649, 1976. Citado 2 vezes nas páginas 13 e 16.

SIMONS, K. et al. Age and gender patterns in health service utilisation: Age-period-cohort modelling of linked health service usage records. *BMC Health Services Research*, v. 23, p. 480, 2023. Disponível em: <<https://doi.org/10.1186/s12913-023-09456-x>>. Citado na página 35.

Sá, M. C. de. Risco moral e seleção adversa de beneficiários no mercado de saúde suplementar. *XXXII Encontro nacional de engenharia de produção*, 2012. Citado na página 17.

WINSEN, K. P. M. van; KLEEF, R. C. van; VEN, W. P. M. M. van de. Can premium differentiation counteract adverse selection in the dutch supplementary health insurance? a simulation study. *The European journal of health economics : HEPAC : health economics in prevention and care*, v. 19(5), p. 757–768, 2018. Citado 2 vezes nas páginas 14 e 37.

WOOLDRIDGE, J. M. *Introductory Econometrics: A Modern Approach*. 1st. ed. Cincinnati, OH: South-Western College Publishing, 2003. Citado 2 vezes nas páginas 24 e 25.

WORLD HEALTH ORGANIZATION. *53rd Directing Council, 66th Session of the Regional Committee of WHO for the Americas*. Washington, D.C., USA, 2014. 53rd Directing Council, 66th Session of the Regional Committee of WHO for the Americas. Citado na página 12.

APÊNDICE A – Códigos em R e Python

Códigos dos modelos implementados e do tratamento dos dados disponíveis em: github.com/enricoruggieri/informationasymmetryPNS

Os modelos canônico e as extensões da heterogeneidade das preferências e das variáveis não-utilizadas foram implementados em Python, utilizando sobretudo a biblioteca *statsmodels*.

O modelo multinomial foi implementado em R, devido às bibliotecas mais adequadas para a Regressão Probit Ordenado e para a Regressão Binomial Negativa necessárias para esse modelo.