



Mariana Arozo Benício de Melo

**Modelos autorregressivos generalizados
orientados por score aplicados a seguros:
previsão para número de sinistros, severidade e
sinistro agregado**

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do grau de
Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica
da PUC-Rio.

Orientador : Prof. Cristiano Augusto Coelho Fernandes
Co-orientador: Prof. Eduardo Fraga Lima de Melo

Rio de Janeiro
Dezembro de 2017



Mariana Arozo Benício de Melo

**Modelos autorregressivos generalizados
orientados por score aplicados a seguros:
previsão para número de sinistros, severidade e
sinistro agregado**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Cristiano Augusto Coelho Fernandes

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Eduardo Fraga Lima de Melo

Co-orientador

UERJ

Prof. Antonio Carlos Figueiredo Pinto

Instituto de Administração e Gerência – PUC-Rio

Prof. Adrian Heringer Pizzinga

UFF

Prof. Luciano Vereda Oliveira

UFF

Prof. César da Rocha Neves

UERJ

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 11 de Dezembro de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Mariana Arozo Benício de Melo

Graduou-se em 2001 em Ciências Atuariais com mérito *Magna Cum Laude* pela Universidade Federal do Rio de Janeiro. Em 2005, obteve o grau de Mestre em Administração com ênfase em Finanças pelo IAG/ PUC-Rio. Mariana é analista técnica da Superintendência de Seguros Privados (Susep) desde 2002. Sua área de interesse é ciências atuariais, modelagem estatística, probabilidade, aplicações de séries temporais a seguros.

Ficha Catalográfica

Melo, Mariana Arozo Benício de

Modelos autorregressivos generalizados orientados por score aplicados a seguros: previsão para número de sinistros, severidade e sinistro agregado / Mariana Arozo Benício de Melo; orientador: Cristiano Augusto Coelho Fernandes; co-orientador: Eduardo Fraga Lima de Melo. – Rio de Janeiro: PUC-Rio, Departamento de Engenharia Elétrica, 2017.

v., 123 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Modelos GAS;. 3. Sinistro Agregado;. 4. Teoria do Risco Coletivo;. 5. Soma Aleatória;. 6. Transformada Rápida de Fourier;. 7. Avaliação de Mercado;. 8. Cópulas;. 9. Seguros Financeiros.. I. Fernandes, Cristiano Augusto Coelho. II. Melo, Eduardo Fraga Lima de. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Aos meus amados filhos,
Guiherme e Manuela.

Agradecimentos

Em primeiro lugar, gostaria de agradecer ao Eduardo, pelo apoio e incentivo desde o primeiro dia de doutorado, pelas trocas, sugestões, compreensão e paciência.

Agradeço aos meus filhos Guilherme e Manuela, minha inspiração de vida, pelo amor, apoio e torcida genuína e calorosa a cada etapa superada e pela compreensão com as ausências e falhas nesse período de dedicação ao doutorado. Espero ter passado exemplo de dedicação e esforço e plantado neles a semente da busca pela formação acadêmica.

Aos meu orientador Prof. Cristiano Fernandes, por ter me recebido no doutorado, pelas excelentes aulas, pelos ensinamentos, orientação e parceria.

Aos meus pais, com toda minha gratidão, pelo apoio incondicional e incentivo à minha formação acadêmica desde sempre. À família e amigos, pela ajuda e motivação. Destaco aqui o suporte das minhas queridas amigas da Rui Barbosa, Carlinha (*in memoriam*), Tuti, Raquel, Maria, Chris, Lu e Andréa, que sempre me motivaram e me incentivaram, dando forças nos muitos momentos difíceis.

À Susep, pela oportunidade de cursar o doutorado. Aos colegas do Gabin e da Codin na Susep, pelo apoio, incentivo e compreensão no período de elaboração da tese. Agradecimento especial ao Alexandre Penner, pela confiança, reconhecimento e suporte.

Aos membros da banca de exame de proposta de tese, pelas contribuições e sugestões de grande valia para a elaboração da tese.

Aos funcionários do DEE/PUC-Rio, pela cordialidade, atenção e paciência.

À ESNS - Escola Nacional de Seguros, pelo suporte e incentivo.

Resumo

Melo, Mariana Arozo Benício de; Fernandes, Cristiano Augusto Coelho (Orientador); Melo, Eduardo Fraga Lima de (Coorientador). **Modelos autorregressivos generalizados orientados por score aplicados a seguros: previsão para número de sinistros, severidade e sinistro agregado**. Rio de Janeiro, 2017. 123p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo desta tese é apresentar novas alternativas para modelagem de variáveis aleatórias no setor de seguros, utilizando o arcabouço dos modelos orientados por *score* com parâmetros variantes no tempo. No primeiro artigo, propomos um modelo dinâmico para a distribuição do sinistro agregado, que corresponde à soma aleatória dos valores de sinistros (severidade) em determinado período de tempo. A obtenção da distribuição do sinistro agregado é um problema clássico na teoria do risco e fundamental para precificação de seguros, cálculo de provisões e de probabilidade de ruína. No entanto, a obtenção da expressão analítica para essa distribuição de probabilidade é uma tarefa difícil. Neste trabalho, especificamos distribuições não-Gaussianas, tanto para o número de sinistros como para severidade, sob o arcabouço GAS (*Generalized Autoregressive Score*), e, por meio do uso da Transformada Rápida de Fourier obtemos, numericamente, a distribuição do sinistro agregado. O segundo artigo trata da incorporação do efeito de variáveis macroeconômicas na modelagem de variáveis relevantes no setor de seguros, em linha com requisito internacional de avaliação de provisões de forma consistente com mercado, a qual leva em consideração as informações disponíveis nos mercados financeiros e de capital relevantes, utilizando metodologias e parâmetros consistentes com esses mercados. Modelamos uma série bivariada de número de sinistros (duas linhas de negócios) de seguros financeiros com modelos autorregressivos e utilizamos cópulas para modelar a estrutura de dependência das séries temporais condicionado aos modelos ajustados nas marginais. Com esta abordagem, é possível simular números de sinistros futuros de mais de uma carteira, podendo esse resultado ser utilizado em uma avaliação consistente de provisões e da saúde financeira da seguradora.

Palavras-chave

Modelos GAS; Sinistro Agregado; Teoria do Risco Coletivo; Soma Aleatória; Transformada Rápida de Fourier; Avaliação de Mercado; Cópulas; Seguros Financeiros.

Abstract

Melo, Mariana Arozo Benício de; Fernandes, Cristiano Augusto Coelho (Advisor); Melo, Eduardo Fraga Lima de (Co-Advisor). **Generalized autoregressive score driven models applied to insurance: forecasting claim frequency, claim severity and aggregate claims.** Rio de Janeiro, 2017. 123p. Tese de doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The objective of this thesis is to present new alternatives for modeling random variables in the insurance industry, using the framework of the score driven models with time-varying parameters. In the first paper, we propose a dynamic model for the aggregate claims distribution, which corresponds to a random sum of claims severity in a certain period of time. Obtaining the aggregate claims distribution is a classic problem in the Risk Theory and fundamental for premium estimation, measurement of obligations and ruin probability valuation. However, obtaining the analytic expression for this probability distribution is a hard task. In this work, we specify non-Gaussian distributions for both the number of claims and for the claims severity, under the GAS framework, and, through the use of the fast Fourier transform, we obtain, numerically, the aggregate claims distribution. The second paper deals with the incorporation of macroeconomic variables on the modeling of relevant variables in the insurance sector, in line with the international requirements for market consistent valuation of insurance liabilities, which means that one should take into account the available information in relevant financial and capital markets, using methodologies and parameters consistent with these markets. We model a bivariate time series (two lines of business) of financial insurance with autoregressive models and use copulas models to consider the dependency structure of the time series conditioned to the fitted models for the marginals. Within this approach, it is possible to simulate the numbers of claims from more than one portfolio, and this result can be used in a consistent valuation of liabilities and of the financial health of an insurer.

Keywords

GAS Models; Aggregate Claims; Collective Risk Theory; Random Sum; Fast Fourier Transform; Market Valuation; Copulas; Credit Insurance.

Sumário

1	Introdução	17
2	Distribuição do sinistro agregado via modelos orientados por score com parâmetros variantes no tempo	21
2.1	Introdução	21
2.2	Abordagem atuarial para sinistro agregado	24
2.3	Modelos orientados por score	26
2.4	Modelos condicionais	28
2.4.1	Modelos para número de sinistros	28
2.4.2	Modelos para severidade	30
2.4.3	Transformada Rápida de Fourier	34
2.5	Aplicação	36
2.5.1	Dados	36
2.5.2	Resultados	37
2.5.2.1	Diagnósticos	39
2.5.3	Previsão	42
2.5.4	Distribuição do sinistro agregado	47
2.6	Considerações Finais	47
3	Avaliação consistente de mercado para seguros financeiros: uma abordagem multivariada usando modelos orientados por score, fatores macroeconômicos e cópulas	51
3.1	Introdução	51
3.2	Metodologia	54
3.2.1	Modelagem para as distribuições marginais	55
3.2.1.1	Modelos Lineares Generalizados Autorregressivos	56
3.2.1.2	Modelos orientados por score	59
3.2.2	Modelagem para estrutura de dependência condicional via cópulas	63
3.2.2.1	Modelos de cópulas	63
3.2.2.2	Avaliação de dependência e do ajuste do modelo de cópulas	67
3.3	Aplicação	69
3.3.1	Dados	69
3.3.2	Resultados	72
3.3.2.1	Modelagem da série de seguro de fiança locatícia	73
3.3.2.2	Modelagem da série de seguro de crédito interno	79
3.3.2.3	Modelagem da estrutura de dependência via cópulas	85
3.3.3	Previsão	91
3.4	Considerações Finais	99
4	Conclusão	101
	Referências bibliográficas	105
A	Apêndice - Capítulo 2	114

A.1	Resíduos quantílicos	114
A.2	Algoritmo para previsão no modelo GAS	117
B	Apêndice - Capítulo 3	119
B.1	Resíduos quantílicos	119
B.2	Cópuas bivariadas	120
B.3	Modelo t-GAS	122
B.4	Algoritmo para previsão no modelo GAS com cópuas	122

Lista de figuras

Figura 2.1	Séries temporais de números de sinistros, severidade média e sinistro agregado no período de 01/2006 a 02/2014.	36
Figura 2.2	Box-plot mensal da série de número de sinistros com dados de 01/2006 a 02/2014.	37
Figura 2.3	Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Poisson(1,1) no período de 01/2006 a 02/2013.	39
Figura 2.4	Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Binomial Negativa(1,1) no período de 01/2006 a 02/2013.	40
Figura 2.5	Média dos valores de severidade observados, em linha sólida, versus valores estimados do parâmetro variante no tempo μ_t , em linha pontilhada, pelo modelo GAS Gama(1,1) no período de 01/2006 a 02/2013.	40
Figura 2.6	Média do logaritmo dos valores de severidade observados, em linha sólida, versus valores estimados do parâmetro variante no tempo μ_t , em linha pontilhada, pelo modelo GAS Lognormal(1,1) no período de 01/2006 a 02/2013.	41
Figura 2.7	Previsão <i>fora da amostra</i> do modelo GAS Poisson(1,1) para número de sinistros no período de 03/2013 a 02/2014.	43
Figura 2.8	Previsão <i>fora da amostra</i> do modelo GAS Binomial Negativa(1,1) para número de sinistros no período de 03/2013 a 02/2014.	44
Figura 2.9	Previsão <i>fora da amostra</i> do modelo GAS Gama(1,1) para severidade no período de 03/2013 a 02/2014.	44
Figura 2.10	Previsão <i>fora da amostra</i> do modelo GAS Lognormal(1,1) para severidade no período de 03/2013 a 02/2014.	45
Figura 2.11	Quantis e médias das distribuições do sinistro agregado dos 12 meses de fora da amostra (03/2013 a 02/2014) obtidas a partir dos modelos GAS Poisson(1,1) e GAS Gama(1,1) e do uso da FFT.	48
Figura 2.12	Distribuições do sinistro agregado para os 12 meses de fora da amostra (03/2013 a 02/2014), obtidas com o uso da FFT e do modelo GAS Poisson(1,1) para número de sinistros e modelo GAS Gama(1,1) para severidade.	49
Figura 3.1	Esquema para avaliação de sinistros de seguro de riscos financeiros bivariado usando modelos orientados por <i>score</i> , GLM autorregressivo, fatores macroeconômicos e cópulas.	55
Figura 3.2	Séries temporais de número de sinistros e taxa de ocorrência do seguro de fiança locatícia no período de 01/2006 a 12/2015.	69
Figura 3.3	Séries temporais de número de sinistros e taxa de ocorrência do seguro de crédito interno no período de 01/2006 a 12/2015.	70

- Figura 3.4 Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Poisson(1,1) para o seguro de fiança locatícia no período de 01/2006 a 12/2014, com variáveis exógenas IPCA e PIB. 78
- Figura 3.5 Taxa de ocorrência de sinistro observada, em linha sólida, versus taxa de ocorrência estimada, em linha pontilhada, pelo modelo GAS Poisson(1,1) para o seguro de fiança locatícia no período de 01/2006 a 12/2014, com variáveis exógenas IPCA e PIB. 79
- Figura 3.6 Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Binomial Negativa(1,1) para o seguro de crédito interno no período de 01/2006 a 12/2014, com variáveis exógenas PIB e câmbio. 85
- Figura 3.7 Taxa de ocorrência de sinistros observada, em linha sólida, versus taxa de ocorrência estimada, em linha pontilhada, pelo modelo GAS Binomial Negativa(1,1) para o seguro de crédito interno no período de 01/2006 a 12/2014, com variáveis exógenas PIB e câmbio. 86
- Figura 3.8 Gráfico de dispersão das séries de pseudo-observações de seguro de fiança locatícia e crédito interno. 87
- Figura 3.9 Kendall-plot sob hipótese nula de independência para as séries de pseudo-observações de seguro de fiança locatícia e de crédito interno. 87
- Figura 3.10 Densidade de uma cópula Survival Gumbel com parâmetro $\theta = 1.234$. 89
- Figura 3.11 Evolução da estimativa do parâmetro ρ_t para o modelo de cópula t -GAS(1,1). 90
- Figura 3.12 Previsão *fora da amostra* para número de sinistro de seguro de fiança locatícia, no período de 01/2015 a 12/2015, com modelo GAS Poisson(1,1) com variáveis exógenas IPCA e PIB. 92
- Figura 3.13 Previsão *fora da amostra* para taxa de ocorrência de sinistro de seguro de fiança locatícia, no período de 01/2015 a 12/2015, com modelo GAS Poisson(1,1) com variáveis exógenas IPCA e PIB. 92
- Figura 3.14 Previsão *fora da amostra* para número de sinistro de seguro de crédito interno, no período de 01/2015 a 12/2015, com modelo GAS Binomial Negativa(1,1), com variáveis exógenas PIB e câmbio. 93
- Figura 3.15 Previsão *fora da amostra* para taxa de ocorrência de sinistro de seguro de crédito interno, no período de 01/2015 a 12/2015, com modelo GAS Binomial Negativa(1,1), com variáveis exógenas PIB e câmbio. 93

- Figura 3.16 Valores “observados” (calculados com base nos valores observados de número de sinistros e nos valores arbitrados para severidade) de despesa total (S_t^*), em preto, e valores de VaR 95%, dentro da amostra (jan/2006 a dez/2014), um passo à frente, considerando modelagem com dependência, em azul. 97
- Figura 3.17 Previsão fora da amostra para para despesa total (S_t^*) e intervalo de confiança 95% considerando modelagem independente e modelagem com dependência para jan/2015 a dez/2015. 98
- Figura 3.18 Quantis e médias das distribuições simuladas para despesa total (S_t^*), nos 12 meses à frente (01/2015 a 12/2015), considerando modelagem independente e modelagem com dependência. 98
- Figura A.1 Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Poisson(1,1) para número de sinistros no período 01/2006 a 02/2013. 114
- Figura A.2 Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Binomial Negativa(1,1) para número de sinistros no período 01/2006 a 02/2013. 115
- Figura A.3 Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Gama(1,1) para severidade no período 01/2006 a 02/2013. 116
- Figura A.4 Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Lognormal(1,1) para severidade no período 01/2006 a 02/2013. 117
- Figura B.1 Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Poisson(1,1) para o seguro fiança locatícia no período de 01/2006 a 12/2014, com variáveis exógenas IPCA e PIB. 119
- Figura B.2 Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Binomial Negativa(1,1) para o seguro de credito interno no período de 01/2006 a 12/2014, com variáveis exógenas PIB e câmbio. 120

Lista de tabelas

Tabela 2.1	Estimativas de máxima verossimilhança e erros padrões das estimativas referentes aos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1) para número de sinistros e aos modelos GAS Gama(1,1) e GAS Lognormal(1,1) para severidade.	38
Tabela 2.2	Resultados <i>dentro da amostra</i> , período de 01/2006 a 02/2013, dos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1) para número de sinistros e modelos GAS Gama(1,1) e GAS Lognormal(1,1) para severidade.	38
Tabela 2.3	P-valores dos testes aplicados aos resíduos quantílicos dos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1), para número de sinistros, e modelos GAS Gama(1,1) e GAS Lognormal(1,1) para severidade - período 01/2006 a 02/2013. Teste Ljung-Box até <i>lag</i> 30.	42
Tabela 2.4	Previsão <i>fora da amostra</i> e MAPE dos modelos para número de sinistros (GAS Poisson(1,1), GAS Binomial Negativa(1,1) e modelos estáticos) e modelos para severidade (GAS Gama(1,1), GAS Lognormal(1,1) e modelos estáticos) - período de 03/2013 a 02/2014.	46
Tabela 2.5	Valores observados de sinistro agregado (S_t^*), no período de 03/2013 a 02/2014, e valores de VaR 95%, para o mesmo período (fora da amostra), obtidos a partir dos modelos GAS Poisson e Gama e modelos estáticos Poisson e Gama com aproximação Normal para sinistro agregado.	48
Tabela 3.1	Estatísticas descritivas das variáveis exógenas no período de 01/2006 a 12/2015.	72
Tabela 3.2	Correlação entre as variáveis exógenas.	72
Tabela 3.3	AIC dos modelos com distribuição Poisson ajustados à série de sinistros de fiança locatícia.	73
Tabela 3.4	AIC dos modelos com distribuição binomial negativa ajustados à série de sinistros de fiança locatícia.	74
Tabela 3.5	Medidas de aderência dentro da amostra dos modelos GAS Poisson(1,1), com série exógena “média de 3 meses” (média dos últimos três meses), para seguro de fiança locatícia no período de 01/2006 a 12/2014. A partir do modelo completo com as quatro variáveis exógenas (Selic, IPCA, PIB, Câmbio), excluímos, uma a uma, a variável cujo coeficiente se mostrou menos significativo.	75
Tabela 3.6	Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de fiança locatícia via modelos GAS Poisson(1,1) sem variáveis exógenas, com quatro variáveis exógenas e com apenas as variáveis IPCA e PIB. São apresentados, também, as medidas de aderência AIC e BIC, o valor do logaritmo da verossimilhança e a medida de erro MAPE.	76

Tabela 3.7	P-valores dos testes aplicados aos resíduos quantílicos dos modelos GAS Poisson(1,1) com série exógena “média de 3 meses” (média dos últimos três meses) para o seguro de fiança locatícia no período 01/2006 a 12/2014.	77
Tabela 3.8	Medidas de aderência e p-valores dos testes aplicados aos resíduos quantílicos do modelo final selecionado GAS Poisson(1,1) com variáveis exógenas IPCA e PIB para o seguro de fiança locatícia no período 01/2006 a 12/2014.	77
Tabela 3.9	Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de fiança locatícia via modelo GAS Poisson(1,1) com variáveis exógenas IPCA e PIB.	78
Tabela 3.10	AIC dos modelos com distribuição Poisson ajustados à série de sinistros de crédito interno.	80
Tabela 3.11	AIC dos modelos com distribuição binomial negativa ajustados à série de sinistros de crédito interno	80
Tabela 3.12	Medidas de aderência dentro da amostra dos modelos GAS Binomial Negativa(1,1), com série exógena “média de 6 meses” (média dos últimos seis meses), para seguro de crédito interno no período de 01/2006 a 12/2014. A partir do modelo completo com as quatro variáveis exógenas (Selic, IPCA, PIB, Câmbio), excluímos, uma a uma, a variável cujo coeficiente se mostrou menos significativo.	81
Tabela 3.13	Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de crédito interno via modelos GAS Binomial Negativa(1,1) sem variáveis exógenas, com quatro variáveis exógenas e com apenas as variáveis PIB e câmbio. São apresentados, também, as medidas de aderência AIC e BIC, o valor do logaritmo da verossimilhança e a medida de erro MAPE.	82
Tabela 3.14	Testes aplicados nos resíduos quantílicos dos modelos GAS Binomial Negativa(1,1), série “média de 6 meses” (média dos últimos seis meses), para seguro de crédito interno no período de 01/2006 a 12/2014.	82
Tabela 3.15	Medidas de aderência e p-valores dos testes aplicados aos resíduos quantílicos do modelo final selecionado GAS Binomial Negativa(1,1) com variáveis exógenas PIB e câmbio para o seguro de crédito interno no período 01/2006 a 12/2014.	83
Tabela 3.16	Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de crédito interno via modelo GAS Binomial Negativa(1,1), no período de 01/2006 a 12/2014, com variáveis exógenas PIB e Câmbio, série “média de 6 meses” (média dos últimos seis meses).	84
Tabela 3.17	Medidas e testes para dependência.	86
Tabela 3.18	Estimativas dos parâmetros e os respectivos erros padrões, bem como os valores do logaritmo da verossimilhança e das medida AIC e BIC, para os modelos de cópulas.	88
Tabela 3.19	Teste Cramer-von Mises para a cópula Survival Gumbel.	89

Tabela 3.20 Estimativas de máxima verossimilhança e erros padrões para o modelo de cópula t -GAS(1,1).	90
Tabela 3.21 Previsão fora da amostra e MAPE, no período de 01/2015 a 12/2015, dos modelos GAS Poisson(1,1) com variáveis exógenas IPCA e PIB, para seguro de fiança locatícia, e GAS Binomial Negativa(1,1), com variáveis exógenas PIB e câmbio, para seguro de crédito interno.	95
Tabela 3.22 Valores “observados” de despesa total (S_t^*), em Reais, e valores de VaR 95%, fora da amostra, considerando modelagem independente e modelagem com dependência para jan/2015 a dez/2015.	99

Lista de Abreviaturas

AIC – *Akaike Information Criteria*
Bacen – Banco Central do Brasil
BIC – *Bayesian Information Criteria*
BIS – *Bank for International Settlements*
DCS – *Dynamic Conditional Score*
DFT – *Discrete Fourier Transform*
EIOPA – *European Insurance and Occupational Pensions Authority*
FFT – *Fast Fourier Transform*
GAS – *Generalized Autoregressive Score*
IAIS – *International Association of Insurance Supervisors*
IASB – *International Accounting Standards Board*
IFM – *Inference Functions for Margins*
IFRS – *International Financial Reporting Standards*
MAPE – *Mean Absolute Percentage Error*
PIT – *Probability Integral Transformation*
Susep – Superintendência de Seguros Privados

1

Introdução

A indústria bancária passou por uma enorme transformação técnica com a discussão e a implantação dos Acordos de Basiléia, comandada pelo BIS (*Bank for International Settlements*). Tendo iniciado em 1988 com o chamado Basiléia I (BIS [1]), estes acordos foram sendo revistos ao longo do tempo (Basiléia II – BIS [2] e Basiléia III – BIS [3]), incorporando evoluções técnicas, com o objetivo de tornar a regulação do sistema bancário mais preparada para enfrentar crises ou quebras, minimizando os efeitos nocivos destes eventos à economia em geral. A evolução destes acordos demandou do mercado financeiro um relevante esforço e investimento em pesquisa e desenvolvimento de modelos matemáticos para avaliação dos riscos a que estão expostas estas instituições.

Tal fenômeno tem sido observado agora no mercado segurador, com alguns anos de atraso em relação ao mercado bancário. Um processo de revisão dos padrões de gestão de riscos das companhias seguradoras apenas se iniciou nos primeiros anos deste século. Até então, o mercado de seguros não adotava práticas de gestão com base no risco. Modelos adotados para resguardar a saúde financeira de seguradoras eram completamente insensíveis ao risco tomado. A abordagem *benchmark* que tem guiado a revisão do padrão de avaliação de riscos com vista à manutenção da capacidade financeira deste mercado é chamado de Solvência II (ver Sandström [4]). Tal nome foi colocado de fato para que fosse feito paralelo com os acordos de capital de Basiléia. Ao contrário do BIS, que emite princípios a serem seguidos por autoridades supervisoras de bancos em todo o mundo, o Solvência II é um projeto europeu, comandado por um organismo (EIOPA – *European Insurance and Occupational Pensions Authority*) ligado à União Européia. Entretanto, pela importância do mercado europeu, o Solvência II é fonte de inspiração para emissão de princípios pela Associação Internacional de Supervisores de Seguros (IAIS, na sigla em inglês), esta sim com mandato em relação aos supervisores de todo o mundo.

No que se refere ao mercado segurador, em razão da introdução de novos princípios de regulação de solvência, a demanda por ferramentas para gestão, avaliação e mensuração de riscos tem aumentado sensivelmente. Neste sentido, modelos matemáticos, estatísticos e atuariais têm sido aplicados para modelar a dinâmica dos riscos aos quais sociedades seguradoras estão expostas. Como

estes requisitos de solvência têm cobrado maior aderência destes modelos à experiência observada das companhias, tem-se observado o uso de técnicas estatísticas cada vez mais apuradas para a captura de fatores determinantes no perfil de risco das operações securitárias. Neste ponto, devemos chamar a atenção para duas características fundamentais observadas no mercado segurador, e também no mercado financeiro.

O primeiro se refere à dinâmica temporal dos fatores de risco que seguradoras enfrentam. Ao modelar estes riscos, assumir premissa de que o comportamento de variáveis é a mesma ao longo do tempo é forte e de difícil aceitação ou comprovação. Desta forma, a introdução de preditores que se alteram com o passar do tempo é natural e, mais, um requisito para capturar a evolução de variáveis condicionadas no tempo. O objetivo desta tese é apresentar novas alternativas para modelagens de variáveis aleatórias no setor de seguros, utilizando como base o arcabouço dos modelos orientados por *score* com parâmetros variantes no tempo.

O segundo ponto se refere à estrutura de dependência. A geralmente utilizada premissa de independência torna os cálculos bem mais simples. Entretanto, pode ser bastante perigosa sob uma realidade de integração entre mercados, eventos e práticas de negócios (Breyman et al. [5], Jouanin et al. [6] e Embrechts et al. [7]). Especificamente para companhias de seguros, que vendem apólices de curto, médio e longo prazos, tal premissa pode ser desastrosa (Frees & Valdez [8], Venter [9] e Frees & Wang [10]). Portanto, é cristalino que o estudo de dependência entre riscos seja componente importante na efetiva gestão de riscos. Por isso, nesta tese faremos uso de ferramenta para modelar este inter-relacionamento entre variáveis aleatórias. Atualmente, a principal linha de pesquisa nesta área se desenvolve sobre funções de cópulas (ver Nelsen [11]). Ademais, serão avaliadas dinâmicas no tempo também com modelos GAS.

Os modelos GAS, também chamados de modelos DCS (*Dynamic Conditional Score*), são uma nova classe de modelos orientados por observações (segundo a classificação de Cox et al. [12]) para séries temporais proposto por Creal et al. [13, 14] e Harvey & Chakravarty [15]. Essa nova classe fornece um arcabouço geral para modelos de séries temporais com parâmetros variantes no tempo. As principais características são facilidade de estimação e inferência (a verossimilhança possui forma fechada) e generalidade: uma vez computado o *score* da densidade condicional de y_t com respeito ao parâmetro variante no tempo, já se pode aplicar o modelo. Diferentemente da modelagem tradicional para a distribuição do número de sinistros e severidade, a utilização de modelos com parâmetros variantes no tempo permite que a previsão seja mais precisa,

visto que o modelo vai acompanhando a evolução dos parâmetros ao longo do tempo.

No mercado de seguros, o cálculo do prêmio e da probabilidade de ruína depende fundamentalmente da distribuição do sinistro agregado. Neste mercado de muita competição, é extremamente importante obter estimativas acuradas das perdas que podem advir dos contratos de seguro. Do ponto de vista matemático, o sinistro agregado é uma soma aleatória de variáveis aleatórias. Nesse tipo de soma, a aleatoriedade é de natureza dupla: os valores dos sinistros são variáveis aleatórias e o número de sinistros também é aleatório. Entretanto, obter a forma analítica da distribuição de probabilidade dessa soma é uma difícil tarefa e restringe o leque de família de distribuições das variáveis aleatórias individuais que podem ser utilizadas.

No primeiro artigo (capítulo 2), obtemos a distribuição do sinistro agregado, que corresponde à soma aleatória dos valores de sinistros (severidade) em determinado período de tempo. Trata-se de um problema clássico na teoria do risco e fundamental para precificação de seguros, cálculo de provisões e de probabilidade de ruína. No entanto, a obtenção da expressão analítica para essa distribuição de probabilidade é uma tarefa impossível em alguns casos. Neste trabalho, uma nova abordagem é proposta para a modelagem da distribuição preditiva do sinistro agregado. Combinamos os modelos orientados por *score*, para especificar distribuições não-Gaussianas, tanto para o número de sinistros como para severidade, com o uso da Transformada Rápida de Fourier para, então, obtermos numericamente a distribuição do sinistro agregado. Para aplicação da modelagem proposta, foram utilizados dados da carteira de seguro de automóveis de uma seguradora brasileira. Modelamos o número de sinistro por meio das distribuições Poisson e binomial negativa e a severidade por meio das distribuições gama e lognormal. Em todos os casos, trabalhamos com parâmetros variantes no tempo.

O segundo artigo (capítulo 3) propõe uma avaliação consistente de mercado para os seguros financeiros. Em linha com práticas internacionais, avaliação consistente significa levar em consideração todas as informações disponíveis nos mercados financeiros e de capital relevantes, utilizando metodologias e parâmetros consistentes com esses mercados. Em nossa pesquisa, utilizamos essas fontes de informação na modelagem dos sinistros de seguros financeiros. Um dos principais *inputs* para o mercado avaliar de forma consistente os passivos de seguro de crédito é chegar a uma previsão confiável do número de *default* (sinistros) decorrentes das apólices em vigor. Utilizamos dados reais de duas linhas de negócios de seguradoras brasileiras, com importante participação no mercado de seguros financeiros. As linhas de negócios utilizadas foram o seguro

de fiança locatícia e o seguro de crédito interno. Ambos apresentam como uma característica marcante o fato de o sinistro ser o resultado de um evento de *default*. Para o seguro de fiança locatícia, sinistros são decorrentes do *default* do inquilino, enquanto que sinistros de seguro de crédito interno resultam de não cumprimento de contratos financeiros.

Nossa abordagem tem a vantagem de usar dados macroeconômicos, a fim de atender a consistência da previsão com o mercado financeiro e, uma vez que estamos lidando com série temporal bivariada (duas linhas de negócios), também fazemos uso de cópulas para considerar a estrutura de dependência entre as linhas de negócio ao simular números de sinistros futuros de mais de uma carteira. Numa primeira etapa, as séries temporais são inicialmente modeladas separadamente, cada qual usando preditores autorregressivos e regressores exógenos guiados por fatores macroeconômicos. Ajustamos diferentes tipos de modelos para as distribuições condicionais marginais: modelos lineares generalizados autorregressivos e modelos orientados por *score*. Na segunda etapa, cópulas foram utilizadas para modelar a estrutura de dependência das séries temporais condicionado aos modelos ajustados na primeira etapa. Por fim, foram realizadas previsões para números de sinistros e taxa de ocorrência para as duas séries trabalhadas, bem como um exercício de previsão de despesa total considerando a modelagem de dependência.

2

Distribuição do sinistro agregado via modelos orientados por score com parâmetros variantes no tempo

Nesse capítulo, uma nova abordagem é proposta para modelar a distribuição preditiva do sinistro agregado, que representa a soma de todos os valores de sinistros ocorridos em determinado período de tempo. Utilizamos modelos autorregressivos generalizados orientados por *score* para especificar distribuições não-Gaussianas tanto para número de sinistros como para severidade dos sinistros. Com o uso da Transformada Rápida de Fourier, obtemos numericamente a distribuição do sinistro agregado de forma eficiente. A metodologia proposta foi aplicada a dados reais de uma seguradora brasileira de seguro de automóveis. O número de sinistros foi modelado com distribuições Poisson e binomial negativa, enquanto que a severidade de sinistro foi modelada com distribuições gama e lognormal. Em todos os modelos, parâmetros apropriados foram tratados como variantes no tempo segundo o mecanismo de orientação por *score*.

2.1

Introdução

No mercado de seguros, a mensuração de obrigações e a avaliação de riscos atuariais dependem fundamentalmente da distribuição do sinistro agregado. A entrada em vigor de princípios de supervisão de solvência assim como de mensuração de contratos de seguros têm demandado abordagens mais sofisticadas para se avaliar as despesas de uma seguradora com sinistros. Dentro deste conceito, as obrigações, assim como os ativos, devem estar mensurados de forma consistente e coerente, sempre procurando aplicar técnicas de avaliação econômica, chamada também de “avaliação realista”. A adequada mensuração destes itens é peça preponderante na aferição da saúde financeira de uma seguradora. Conforme IAIS [16] e IASB [17], a mensuração de obrigações deve refletir uma avaliação prospectiva de fluxos de caixa futuros da obrigação, incluindo o risco destes fluxos e o valor do dinheiro no tempo. Por isso, além da obtenção do valor esperado destas despesas, é fundamental aferir o risco inerente a estes fluxos de caixa, cuja distribuição de probabilidade passa a ser o grande objetivo.

Basicamente, as provisões técnicas devem corresponder a uma estimativa corrente associada a uma margem de risco. Define-se estimativa corrente como a média dos fluxos de caixa futuros, ponderados pela sua probabilidade. Já a margem de risco deve ser mensurada tal que garanta que o valor das provisões técnicas seja equivalente ao montante que as seguradoras deveriam normalmente exigir para assumir e cumprir as obrigações de seguro. Dentro deste contexto, quando os princípios emitidos por IAIS [16] e IASB [17] dispõem sobre margem de risco, o foco passa a ser a obtenção da distribuição completa do sinistro agregado. Da mesma forma, quando o interesse recai sobre a avaliação de riscos atuariais, por exemplo para calcular capital baseado no risco de subscrição, a distribuição completa do sinistro agregado fornece diretamente um valor em risco (VaR) ou outra medida de risco de interesse (como por exemplo, $TVaR$ - *Tail VaR*).

Do ponto de vista matemático, o sinistro agregado é uma soma aleatória de variáveis aleatórias. Nesse tipo de soma, a aleatoriedade é de natureza dupla: os valores dos sinistros e o número de sinistros são ambas variáveis aleatórias. Entretanto, obter a forma analítica da distribuição de probabilidade dessa soma é uma difícil tarefa e restringe o leque de famílias de distribuições das variáveis aleatórias individuais que podem ser utilizadas.

Nesse capítulo, propomos uma nova abordagem para obtenção da distribuição preditiva do sinistro agregado. Combinamos o uso de modelos GAS (*Generalized Autoregressive Score*) para número de sinistros e severidade ao uso da técnica da Transformada Rápida de Fourier para obtenção da distribuição completa do sinistro agregado. Os modelos GAS, também chamados de modelos DCS (*Dynamic Conditional Score*), são uma nova classe de modelos orientados por observações (segundo a classificação de Cox et al. [12]) para séries temporais proposta por Creal et al. [13, 14] e Harvey & Chakravarty [15]. Segundo Cox et al. [12], os modelos para séries temporais podem ser classificados em modelos orientados por parâmetros ou modelos orientados por observações. Nos modelos orientados por parâmetros, a equação de evolução do parâmetros tem sua própria fonte exógena de erro. Por outro lado, nos modelos orientados por observações, a equação de evolução do parâmetro não possui fonte de erro, de modo que a partir dos dados disponíveis até o tempo t é possível obter o parâmetro estocástico no instante $t + 1$.

Essa nova classe fornece um arcabouço geral para modelos com parâmetros variantes no tempo. As vantagens desses modelos são: facilidade de estimação, uma vez que a verossimilhança possui forma fechada, e flexibilidade, uma vez que pode ser usado com diferentes distribuições preditivas, sejam contínuas ou discretas. Diferentemente da modelagem tradicional para

a distribuição do número de sinistros e severidade, a utilização de modelos com parâmetros variantes no tempo permite que a previsão seja mais precisa, visto que o modelo permite a evolução dos parâmetros ao longo do tempo. Tendo em vista que a distribuição de sinistro agregado, exceto em raríssimos casos, não apresenta forma fechada, a utilização da Transformada Rápida de Fourier se mostra uma ferramenta eficiente computacionalmente e precisa para obtenção desta distribuição. A Transformada Rápida de Fourier é um método que permite a obtenção de uma distribuição a partir de sua função característica, conforme descrito em Klugman et al. [18], Robertson [19], Waller et al. [20] e Wang [21]. Conforme mencionado por Kass et al. [22], a Transformada Rápida de Fourier se apresenta como um método muito mais rápido para obtenção da distribuição do sinistro agregado do que métodos recursivos.

A metodologia proposta foi aplicada a dados da carteira de seguro de automóveis de uma seguradora brasileira. As características dos riscos segurados são homogêneas. Modelamos o número de sinistro por meio das distribuições Poisson e binomial negativa e a severidade por meio das distribuições gama e lognormal. Em todos os casos, trabalhamos com parâmetros variantes no tempo. Além da aplicação proposta neste capítulo, o método apresentado pode ser aplicado para outros ramos de seguros ou outros problemas que envolvam soma de variáveis aleatórias e cálculo de distribuições compostas.

Aqui cabe destacar que nossa abordagem pode ser facilmente estendida para avaliação de risco operacional e de crédito. No campo do risco operacional, essa soma aleatória também se apresenta como um problema importante, uma vez que a estimação do capital de risco operacional sob a abordagem da distribuição da perda (*Loss Distribution Approach* - Frachot et al. [23]) requer o cálculo da distribuição da perda operacional agregada. Muitos bancos adotam essa abordagem para quantificar o capital de risco operacional sob o arcabouço regulatório de supervisão bancária. Já para o risco de crédito, a abordagem conhecida como CreditRisk+ (Suisse [24]) adota uma abordagem atuarial para as perdas decorrentes de uma carteira de crédito, na qual o sinistro se caracteriza pelo não pagamento de uma dívida.

O capítulo está organizado da seguinte maneira: na seção 2.2, apresentamos a Teoria do Risco Coletivo e abordagens utilizadas na literatura. Na seção 2.3, introduzimos os modelos orientados por *score* com parâmetros variantes no tempo. Os modelos propostos para a caracterização probabilística da soma aleatória e o uso da Transformada Rápida de Fourier são discutidas na seção 2.4. Uma aplicação com dados reais é apresentada na seção 2.5, na qual os resultados e seleção de modelos são discutidos. Por fim, a seção 2.6 apresentam as considerações finais. No Apêndice A, estão os gráficos relativos aos resíduos

quantílicos dos modelos trabalhados e o algoritmo para previsão com modelos GAS.

2.2

Abordagem atuarial para sinistro agregado

O sinistro agregado representa o total de indenizações pagas para todos os sinistros ocorridos em um determinado período de tempo em uma carteira de seguros. A distribuição de sinistro agregado tem sido objeto de muitos trabalhos na literatura atuarial. Alguns autores, como Hewitt Jr [25], basearam a distribuição de sinistro agregado em dados de sinistros agregados observados. Um problema com essa abordagem é conseguir volumes de dados suficientes, sendo muitas vezes necessário combinar experiência de diferentes seguradoras que podem ter comportamentos distintos. Uma alternativa, então, é calcular a distribuição de sinistro agregado a partir das distribuições de número de sinistros e de severidade.

No modelo de risco coletivo, consideramos um portfólio que produz N sinistros em um determinado período de tempo. Conforme representação de Bowers et al. [26] (p. 367), o sinistro agregado, S , é a soma de um número aleatório de pagamentos individuais também aleatórios:

$$S = X_1 + X_2 + \dots + X_N, \quad (2-1)$$

onde N é a variável aleatória que representa o número de sinistros no período e X_i é a variável aleatória que representa a severidade do sinistro i . Obviamente, no caso em que $N = 0$, então $S = 0$.

Neste tipo de abordagem, faz-se necessário conhecer a distribuição de probabilidade do valor de cada sinistro (severidade), independentemente da apólice à qual o sinistro pertence, bem como a distribuição do número total de sinistros produzidos em uma carteira. As hipóteses adotadas são que X_1, X_2, \dots são i.i.d e independentes de N . A distribuição do sinistro agregado é obtida a partir das distribuições do número de sinistros e da severidade, que são modeladas separadamente. Em raríssimos casos, a distribuição do sinistro agregado pode ser obtida de forma analítica, como por exemplo no caso da severidade ser modelada com a distribuição exponencial e o número de sinistros com distribuição geométrica (ver Klugman et al. [18], p. 153-154). Portanto, na maioria das vezes, se faz necessário recorrer a aproximações, métodos numéricos ou recursivos para este fim.

Neste capítulo, trabalhamos os dados de severidade e número de sinistros como séries temporais. Dessa forma, temos o modelo de teoria do risco coletivo (equação 2-1) reescrito como:

$$S_t^* = X_{1t} + X_{2t} + \dots + X_{N_t t}, \quad t = 1, \dots, T, \quad (2-2)$$

no qual S_t^* é o sinistro agregado no tempo t , N_t é o número de sinistros no tempo t e X_{it} , $i=1, \dots, N_t$, os valores de sinistros ocorridos no tempo t .

Os momentos de S_t^* podem ser obtidos em termos dos momentos das variáveis N_t e X_{it} , sendo a média dada por

$$E[S_t^*] = E[E[S_t^*|N_t]] = E[N_t]E[X_{it}], \quad (2-3)$$

e a variância por

$$\text{Var}[S_t^*] = E[\text{Var}[S_t^*|N_t]] + \text{Var}[E[S_t^*|N_t]] = E[N_t]\text{Var}[X_{it}] + E[X_{it}]^2\text{Var}[N_t]. \quad (2-4)$$

A função de distribuição acumulada de S_t^* é dada por

$$\begin{aligned} F_{S_t^*}(s) &= P(S_t^* \leq s) = \sum_{n=0}^{\infty} P(X_{1t} + X_{2t} + \dots + X_{N_t t} \leq s \mid N_t = n)P(N_t = n) \\ &= \sum_{n=0}^{\infty} P(X_{1t} + X_{2t} + \dots + X_{nt} \leq s)P(N_t = n), \end{aligned}$$

e também pode ser expressa por meio de convolução:

$$F_{S_t^*}(s) = \sum_{n=0}^{\infty} P(X_{1t} + X_{2t} + \dots + X_{nt} \leq s)P(N_t = n) = \sum_{n=0}^{\infty} F_{X_{it}}^{*n}(s)P(N_t = n),$$

na qual $F_{X_{it}}^{*n}$ é a n -ésima convolução da função de distribuição de X_{it} .

Como, na maioria dos casos, não é possível obter a distribuição do sinistro agregado de forma analítica, a literatura atuarial apresenta diversos métodos para atingir esse objetivo. De maneira geral, os métodos podem ser divididos em métodos de aproximação, métodos recursivos e métodos numéricos (simulação e métodos de inversão).

Kass et al. [22] e Klugman et al. [18] apresentam alguns métodos utilizados para obtenção da distribuição do sinistro agregado, como aproximações usando o Teorema do Limite Central, fórmula recursiva de Panjer (ver Panjer [27]) e método de inversão usando a Transformada Rápida de Fourier. As aproximações mais utilizadas são aproximação Normal e aproximação Gama. Chaubey et al. [28] apresentam uma aproximação utilizando a distribuição Normal inversa. Lau [29] também aborda a aproximação para distribuição do sinistro agregado utilizando os quatro primeiros momentos das distribuições de severidade e de número de sinistros. No que se refere a métodos recursivos, o método mais difundido foi proposto por Panjer [27]. Sundt [30] faz um levantamento dos métodos recursivos presentes na literatura atuarial para o

cômputo exato e aproximado de distribuições de sinistro agregado univariadas e multivariadas.

Heckman & Meyers [31] discutem as distribuições de sinistro agregado sob a perspectiva da Teoria do Risco Coletivo e apresentam um algoritmo para obtenção da distribuição acumulada do sinistro agregado a partir das distribuições do número de sinistros e da severidade que envolve a inversão da função característica. Shevchenko [32] também faz uma análise dos diferentes métodos de obtenção da distribuição da perda agregada, focando no ponto de vista de risco operacional.

Neste trabalho, o objetivo é concentrado na obtenção de distribuições preditivas dos números de sinistros, severidade e sinistro agregado. Harvey & Fernandes [33] modelaram as variáveis aleatórias número de sinistros e severidade com distribuições preditivas via modelos estruturais para séries temporais, além de apresentarem uma extensão para modelagem de grupos. Dickson et al. [34] e Meyers [35] também trabalharam na obtenção da distribuição preditiva do sinistro agregado, mas sob o ponto de vista bayesiano. Nossa abordagem é natural, uma vez que as previsões são baseadas em distribuições condicionais do futuro dado o passado e incorporam a incerteza dos parâmetros. Uma vez modelados o número de sinistros e a severidade, condicionados ao conjunto de informação até $t - 1$, \mathcal{F}_{t-1} , a distribuição do sinistro agregado foi obtida por meio da Transformada Rápida de Fourier. Cálculos de prêmio, de probabilidade de ruína, de capital baseado no risco, entre outros, podem não ser corretamente dimensionados quando a modelagem não leva em consideração a incerteza dos parâmetros.

2.3

Modelos orientados por score

Tendo em vista nosso objetivo de obter a distribuição de sinistro agregado em um contexto de séries temporais, é natural considerar a possibilidade de tanto a distribuição do número de sinistros como a distribuição de severidade terem parâmetros variantes no tempo. Com esse propósito, utilizamos uma classe de modelos recentemente desenvolvida para séries temporais, conhecida como modelos GAS (*Generalized Autoregressive Score*), propostos por Creal et al. [13, 14] e Harvey & Chakravarty [15]. Os modelos GAS fornecem um arcabouço geral para modelar distribuições com parâmetros variantes no tempo, cuja equação de atualização é guiada por valores passados do parâmetro e do *score* ponderado. O *score*, no contexto dos modelos GAS, é definido como a primeira derivada da função do log da densidade/função de probabilidade com respeito ao vetor de parâmetro variante no tempo.

Também é possível introduzir variáveis exógenas na equação de atualização dos parâmetros. A partir dessa modelagem, os parâmetros estocásticos são perfeitamente previsíveis a partir da informação passada, tal como em modelos GARCH.

Seja $\{y_t\}_{t \in \mathbb{N}}$ uma sequência estocástica, a densidade/função de probabilidade condicional de y_t no modelo GAS é dada por:

$$y_t \sim p(y_t | f_t, \mathcal{F}_{t-1}; \theta), \quad (2-5)$$

na qual $\{f_t, \mathcal{F}_{t-1}\}$ é o conjunto de informações disponíveis no tempo t , sendo $\mathcal{F}_{t-1} = \{Y^{t-1}, F^{t-1}\}$, $Y^{t-1} = \{y_1, \dots, y_{t-1}\}$, $F^{t-1} = \{f_1, \dots, f_{t-1}\}$, f_t é o vetor de parâmetro variante no tempo e θ é o vetor de parâmetros fixos, $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$. A equação de atualização para f_t é dada pela seguinte fórmula:

$$f_{t+1} = w + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1}. \quad (2-6)$$

na qual w é um vetor de constantes, A_i e B_j são matrizes de coeficientes, cujas dimensões dependem de f_t , e s_t é o vetor *score* (∇_t) ponderado pela matriz S_t ,

$$s_t = S_t \nabla_t, \quad (2-7)$$

sendo

$$\nabla_t = \frac{\partial \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t}. \quad (2-8)$$

A matriz de ponderação S_t , em geral, é escolhida entre $I_{t|t-1}^{-1}$, $I_{t|t-1}^{-1/2}$ ou I , sendo $I_{t|t-1} = E_{t-1} [\nabla_t \nabla_t']$ a matriz de informação e I a matriz identidade. O fato de o modelo estar baseado no *score* possibilita que seja explorada toda a estrutura da densidade/função de probabilidade preditiva e não apenas alguns momentos.

Uma característica importante da equação de atualização de f_t é que, sob correta especificação, o *score* ponderado s_t é uma diferença martingal, ou seja, $E_{t-1} [s_t] = 0$. Adicionalmente, no caso de utilização da matriz de ponderação como $I_{t|t-1}^{-1/2}$, s_t tem variância unitária. No caso de $S_t = I$, Creal et al. [13] argumentam que, em algumas situações observadas, o mecanismo de atualização fica menos estável.

Blasques et al. [36, 37] apresentam uma fundamentação teórica para o uso da função *score* na equação de atualização do parâmetro variante no tempo baseado na distância de Kullback-Leibler (KL). Eles demonstram que apenas modelos que utilizam o *score* na equação de atualização são ótimos no sentido de Kullback-Leibler, ou seja, em relação à distância de Kullback-Leibler, não há nenhum outro esquema de atualização que estaria mais próximo do verdadeiro.

Creal et al. [14] mostram que muitos modelos orientados por observação, propostos na literatura estatística/econométrica, são casos particulares dos modelos GAS. Por exemplo, o modelo generalizado autorregressivo heterocedástico (GARCH), proposto por Engle [38] e Bollerslev [39], pode ser escrito como um modelo GAS utilizando-se $S_t = I_{t|t-1}^{-1}$. Os parâmetros fixos dos modelos são estimados por máxima verossimilhança, a qual possui forma fechada e pode ser obtida diretamente por meio da densidade/função de probabilidade condicional (2-5) e da equação de atualização de f_t (2-6). A função é, então, avaliada para um determinado valor θ^* de θ .

Para uma série observada y_1, \dots, y_n , a estimativa de máxima verossimilhança seria expressa como

$$\hat{\theta} = \arg \max_{\theta} \sum_{t=1}^n \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta).$$

As propriedades dos estimadores de máxima verossimilhança nos modelos GAS foram estudadas por Blasques et al. [40, 41]. Para fins de inferência, Creal et al. [14] propõem computar os erros padrões e estatísticas t para os parâmetros estimados, com base na inversa da matriz Hessiana do logaritmo da verossimilhança avaliada no ótimo ($\hat{\theta}$). Sob certas condições de regularidade, o estimador de máxima verossimilhança $\hat{\theta}$ de θ é consistente e satisfaz

$$\sqrt{n} (\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}),$$

na qual $H = \lim_{n \rightarrow \infty} E[(\partial l / \partial \theta)(\partial l / \partial \theta')] / n$ e $l = \sum_{t=1}^n \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta)$.

Nos casos em que há alguma restrição natural nos parâmetros da distribuição preditiva, é necessário utilizar uma parametrização diferente para o modelo. Como será visto na subseção 2.4.1, para a distribuição de Poisson com parâmetro λ_t (média e variância), $\lambda_t > 0$ e, assim, utilizamos $f_t = \ln(\lambda_t)$.

2.4

Modelos condicionais

2.4.1

Modelos para número de sinistros

Como mencionado em Klugman et al. [18], na maioria dos casos no setor de seguros, o número de sinistros, N_t , pode ser modelado com as distribuições de contagem Poisson e binomial negativa. A distribuição binomial negativa, por possuir dois parâmetros, apresenta maior flexibilidade para ajuste do que a distribuição Poisson e costuma ser mais apropriada para os casos em que os dados a serem modelados apresentem sobredispersão. Neste trabalho foram

usadas as distribuições Poisson e binomial negativa para modelar o número de sinistros (N_t). A formulação GAS para essas distribuições são apresentadas a seguir.

Modelo GAS Poisson:

Seja a distribuição condicional para o número de sinistros N_t

$$(N_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) \sim \text{Poisson}(\lambda_t),$$

com função de distribuição de probabilidade

$$P(N_t = n_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) = \frac{\lambda_t^{n_t} e^{-\lambda_t}}{n_t!}, \quad n_t = 0, 1, \dots, \lambda_t > 0, \quad (2-9)$$

sendo

$$E(N_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) = \text{Var}(N_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) = \lambda_t.$$

Definimos o parâmetro variante no tempo, f_t , em função da média condicional λ_t utilizando uma reparametrização que garanta que a média λ_t seja sempre positiva:

$$f_t = \ln(\lambda_t). \quad (2-10)$$

Utilizando as formulações (2-7), (2-8) e (2-10) aplicadas a (2-9), temos que

$$\nabla_t = n_t - \lambda_t, \quad I_{t|t-1} = \lambda_t$$

e

$$s_t = (n_t - \lambda_t)/\lambda_t^d, \quad \text{para } d = 0, 1 \text{ ou } 1/2.$$

O valor de “ d ” vai depender da escolha da matriz de ponderação S_t ($I, I_{t|t-1}^{-1}$ ou $I_{t|t-1}^{-1/2}$, respectivamente). Para os casos em que a série apresenta um comportamento sazonal, é possível incorporar um efeito sazonal no modelo por meio da equação de atualização de f_t (2-6). Nesse sentido, trabalharíamos com um modelo em que $p = q = 12$ mas com coeficientes A_i e B_i diferentes de zero para valores de i e j múltiplos de 12 ou próximos. Portanto, a equação de atualização de f_t assumiria a seguinte forma:

$$\begin{aligned} f_{t+1} = & w + A_1 s_t + A_2 s_{t-1} + A_3 s_{t-2} + A_{11} s_{t-10} + A_{12} s_{t-11} \\ & + B_1 f_t + B_2 f_{t-1} + B_3 f_{t-2} + B_{11} f_{t-10} + B_{12} f_{t-11}. \end{aligned}$$

Modelo GAS Binomial Negativa:

Para modelagem de N_t via distribuição binomial negativa, seguimos a parametrização apresentada em Creal et al. [13]. Nesse caso, temos que

$$(N_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) \sim \text{Bin Neg}(r, p_t),$$

com a função de distribuição de probabilidade dada por

$$P(N_t = n_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) = \frac{\Gamma(n_t + r)}{\Gamma(r)n_t!} p_t^r (1 - p_t)^{n_t}, \quad (2-11)$$

com $n_t = 0, 1, \dots, r > 0, 0 < p_t \leq 1$ e média condicional e variância condicional dadas por

$$E(N_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) = \frac{r(1 - p_t)}{p_t}$$

e

$$\text{Var}(N_t|f_t, \mathcal{F}_{t-1}; \theta^{(1)}) = \frac{r(1 - p_t)}{p_t^2}.$$

Para garantir que $0 < p_t \leq 1$ adotamos a função de ligação logit, ou seja,

$$f_t = \ln \left(\frac{p_t}{1 - p_t} \right). \quad (2-12)$$

Utilizando as formulações (2-7), (2-8) e (2-12) aplicadas a (2-11), temos que

$$\nabla_t = r(1 - p_t) - n_t p_t, \quad I_{t|t-1} = r(1 - p_t),$$

e

$$s_t = \frac{r(1 - p_t) - n_t p_t}{(r(1 - p_t))^d}, \quad d = 0, 1 \text{ ou } 1/2.$$

Assim como no modelo GAS Poisson, o valor de “ d ” vai depender da escolha da matriz de ponderação S_t ($I, I_{t|t-1}^{-1}$ ou $I_{t|t-1}^{-1/2}$, respectivamente) e é determinado empiricamente a partir de avaliação de ajuste dos modelos.

2.4.2

Modelos para severidade

Para modelar a severidade, seguindo os resultados de Creal et al. [42], assumimos que os valores de sinistros (severidade) no tempo t ($x_{i,t}$) são identicamente distribuídos e independentes condicionalmente a f_t e ao conjunto de informações já disponíveis \mathcal{F}_{t-1} . A densidade preditiva para $x_{i,t}$ é dada por

$$x_{i,t} \sim p(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}), \quad \text{para } i = 1, \dots, n_t, \quad t = 1, \dots, T,$$

na qual n_t é o número de sinistros no tempo t .

Todas as distribuições dependem do mesmo parâmetro variante no tempo f_t e parâmetros fixos $\theta^{(2)}$ e a densidade condicional $p(x_t|f_t, \mathcal{F}_{t-1}; \theta^{(2)}, N_t = n_t)$ é dada por

$$p(x_t|f_t, \mathcal{F}_{t-1}; \theta^{(2)}, N_t = n_t) = \prod_{i=1}^{n_t} p(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}).$$

Sendo assim, conforme descrito em Creal et al. [42], o logaritmo da verossimilhança, o *score* e a matriz de informação apresentam uma forma aditiva simples:

$$l = \sum_{t=1}^T \ln p(x_t|f_t, \mathcal{F}_{t-1}; \theta^{(2)}, N_t = n_t) = \sum_{t=1}^T \sum_{i=1}^{n_t} \ln p(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}), \quad (2-13)$$

$$\nabla_t = \sum_{i=1}^{n_t} \nabla_{i,t} = \sum_{i=1}^{n_t} \frac{\partial \ln p(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)})}{\partial f_t}, \quad (2-14)$$

e

$$I_{t|t-1} = E_{t-1} [\nabla_t \nabla_t'] = \sum_{i=1}^{n_t} E_{i,t-1} [\nabla_{i,t} \nabla_{i,t}'] = \sum_{i=1}^{n_t} I_{t|t-1,i}, \quad (2-15)$$

sendo n_t o número de sinistros no tempo t .

As distribuições gama e lognormal são, com frequência, escolhas para modelagem de valores de sinistros de seguro de automóveis por conta de suas características apropriadas para severidade como suporte não negativo, assimetria positiva e variância proporcional à média ao quadrado. Nesse capítulo, essas duas distribuições são utilizadas para modelar a severidade dos sinistros.

Modelo GAS Gama:

Para o modelo GAS Gama, adotamos uma parametrização que torna mais simples a identificação da média da distribuição, que seria o parâmetro de interesse variante no tempo:

$$(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) \sim \text{Gama} \left(\alpha, \frac{\mu_t}{\alpha} \right),$$

sendo a função de densidade de probabilidade dada por

$$f(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) = \frac{1}{\Gamma(\alpha)(\alpha^{-1}\mu_t)^\alpha} x_{i,t}^{\alpha-1} e^{-\frac{\alpha x_{i,t}}{\mu_t}}, \quad x_{i,t} > 0, \quad \alpha, \mu_t > 0, \quad (2-16)$$

na qual α é o parâmetro de forma da distribuição gama e μ_t/α é o parâmetro de escala. Nessa modelagem, teremos tanto a média condicional como a variância condicional variantes no tempo:

$$E(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) = \mu_t$$

e

$$\text{Var}(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) = \frac{\mu_t^2}{\alpha}.$$

Assim como no caso do modelo GAS Poisson, para garantir que a média μ_t seja sempre positiva, adotamos a função de ligação logarítmica, isto é,

$$f_t = \ln(\mu_t). \quad (2-17)$$

Conjugando as formulações (2-8), (2-16) e (2-17), temos que

$$\nabla_{i,t} = \alpha \left(\frac{x_{i,t}}{\mu_t} - 1 \right), \quad I_{t|t-1,i} = \alpha,$$

e, adotando (2-13), (2-14) e (2-15), obtemos

$$\nabla_t = \sum_{i=1}^{n_t} \nabla_{i,t} = \alpha \left(\frac{\sum_{i=1}^{n_t} x_{i,t}}{\mu_t} - n_t \right) \quad \text{e} \quad I_{t|t-1} = \sum_{i=1}^{n_t} I_{t|t-1,i} = n_t \alpha. \quad (2-18)$$

A partir das equações (2-7) e (2-18), o *score* ponderado é dado por

$$s_t = \frac{\alpha}{(\alpha n_t)^d} \left(\frac{\sum_{i=1}^{n_t} x_{t,i}}{\mu_t} - n_t \right), \text{ para } d = 0, 1 \text{ ou } 1/2,$$

sendo que o valor de “ d ” vai depender da escolha da matriz de ponderação S_t ($I, I_{t|t-1}^{-1}$ ou $I_{t|t-1}^{-1/2}$, respectivamente).

Modelo GAS Lognormal:

Para o modelo GAS Lognormal, utilizamos a parametrização tradicional, tornando o parâmetro μ_t variante no tempo. Nessa modelagem, assim como no caso do modelo GAS Gama, teremos tanto a média condicional f_t como a variância condicional variantes no tempo.

Seja a distribuição condicional para severidade de sinistro $x_{i,t}$

$$(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) \sim \text{Lognormal}(\mu_t, \sigma^2),$$

com função de densidade de probabilidade dada por

$$f(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) = \frac{1}{x_{i,t}\sigma\sqrt{2\pi}} e^{-\frac{(\ln x_{i,t} - \mu_t)^2}{2\sigma^2}}, \quad x_{i,t} > 0, \quad \sigma > 0, \quad \mu_t \in \mathbb{R}, \quad (2-19)$$

na qual μ_t é média e σ^2 é a variância da distribuição de $\ln x_{i,t}$, $(\ln x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) \sim \text{Normal}(\mu_t, \sigma^2)$. A média condicional e a variância condicional de $x_{i,t}$ são dadas por

$$E(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) = e^{\mu_t + \frac{\sigma^2}{2}}$$

e

$$\text{Var}(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) = (e^{\sigma^2} - 1)e^{2\mu_t + \sigma^2}.$$

Considerando o parâmetro variante no tempo μ_t igual a f_t , não é necessário impor restrições na modelagem, uma vez que $\mu_t \in \mathbb{R}$:

$$f_t = \mu_t. \quad (2-20)$$

Conjugando as formulações (2-8), (2-19) e (2-20), temos que

$$\nabla_{t,i} = \frac{\ln x_{i,t} - \mu_t}{\sigma^2}, \quad I_{t|t-1,i} = \frac{1}{\sigma^2},$$

e adotando (2-13), (2-14) e (2-15), obtemos

$$\nabla_t = \sum_{i=1}^{n_t} \nabla_{t,i} = \frac{\sum_{i=1}^{n_t} \ln x_{i,t} - n_t \mu_t}{\sigma^2}, \quad I_{t|t-1} = \sum_{i=1}^{n_t} I_{t|t-1,i} = \frac{n_t}{\sigma^2}. \quad (2-21)$$

A partir das equações (2-7) e (2-21), o *score* ponderado é dado por

$$s_t = \left[\frac{\sum_{i=1}^{n_t} \ln x_{i,t} - n_t \mu_t}{\sigma^2} \right] \left(\frac{\sigma^2}{n_t} \right)^d, \text{ para } d = 0, 1 \text{ ou } 1/2,$$

sendo que o valor de “ d ” vai depender da escolha da matriz de ponderação S_t ($I, I_{t|t-1}^{-1}$ ou $I_{t|t-1}^{-1/2}$, respectivamente). Nosso objetivo é a obtenção da distribuição de sinistro agregado. Nesse sentido, tendo em vista que essa distribuição não pode ser obtida de forma analítica para a grande maioria dos casos, utilizaremos os modelos para séries de números de sinistros (N_t) e para severidade (x_{it}) descritos nessas duas últimas subseções com a técnica da Transformada Rápida de Fourier, apresentada a seguir, para obtenção da distribuição alvo.

2.4.3

Transformada Rápida de Fourier

A Transformada Rápida de Fourier (FFT) é um método eficiente de calcular distribuições compostas por meio da inversão das funções características dessas distribuições. Enquanto nem todas as distribuições apresentam função geradoras de momentos, todas as distribuições possuem funções características únicas. Da mesma maneira, dada uma função característica, sempre existirá uma distribuição única, o que torna conveniente o método que utiliza função característica.

A função característica de uma variável aleatória X contínua é definida por:

$$\phi_X(z) = E[e^{izX}] = \int_{-\infty}^{+\infty} f(x)e^{izx}dx, \quad z \in \mathbb{R}.$$

No caso de uma soma aleatória, como o sinistro agregado no tempo t (S_t^*) no modelo de risco coletivo (2-2), a função característica é dada por

$$\begin{aligned} \phi_{S_t^*}(z) &= E[e^{izS_t^*}] = E[E(e^{iz(X_{1t}+X_{2t}+\dots+X_{N_t t})}|N_t)] \\ &= E[\phi_{X_{it}}(z)^{N_t}] = P_{N_t}[\phi_{X_{it}}(z)], \quad z \in \mathbb{R}, \end{aligned}$$

na qual $P_{N_t}(z) = E[z^{N_t}]$ é a função geradora de probabilidade da variável aleatória N_t no tempo t . Para variáveis aleatórias discretas (N_t), a função geradora de probabilidade é definida por:

$$P_{N_t}(z) = E(z^{N_t}) = \sum_{n=0}^{\infty} p(n)z^n. \quad (2-22)$$

Seja f_x uma função definida para valores inteiros x , $0 \leq x \leq n-1$. Para o vetor $(f_0, f_1, \dots, f_{n-1})$, a Transformada de Fourier Discreta (*Discrete Fourier Transform* - DFT) é o mapeamento definido por:

$$\tilde{f}_k = \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi i}{n}jk\right), \quad 0 \leq k \leq n-1.$$

O mapeamento inverso (IDFT), que recompoe os valores da função original, é dado por:

$$f_j = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{f}_k \exp\left(-\frac{2\pi i}{n}kj\right), \quad 0 \leq j \leq n-1.$$

A FFT é um método extremamente rápido para computar a DFT para vetores longos. Trata-se de um algoritmo que reduz o número de computações da ordem de $O(n^2)$ para $O(n \log_2 n)$. Portanto, a FFT pode ser vista como uma versão discreta da Transformada de Fourier ou função característica e pode ser

usada para inverter funções características para obtenção de distribuição de variáveis aleatórias. A existência de um algoritmo FFT para computar a DFT de forma eficiente se tornou conhecida apenas na metade da década de 60, com o trabalho de Cooley & Tukey [43]. Detalhes dessa metodologia encontram-se em Press et al. [44]. Segundo Klugman et al. [18], o uso desse método pela primeira vez para inverter funções características de distribuições remete a Bertram [45]. Robertson [19] fornece detalhes do método bem como aplicações para cálculo da distribuição de sinistro agregado.

Waller et al. [20] apresentam uma revisão e discussão da inversão de função característica como uma ferramenta para obter funções de distribuição acumuladas e Wang [21] e Bühlmann [46] também abordam o uso do método para obtenção da distribuição do sinistro agregado. Conforme mencionado por Kaas et al. [22], a FFT se apresenta como um método muito mais rápido para obtenção da distribuição do sinistro agregado do que métodos recursivos, além de lidar de maneira mais eficiente e intuitiva com convoluções no caso de avaliação de perdas agregadas de grupos. Klugman et al. [18] apresentam um passo a passo para obtenção da distribuição do sinistro agregado pela FFT, que é descrito a seguir. Essa metodologia pode ser utilizada tanto com distribuições incondicionais no tempo como com as distribuições condicionais às observações passadas, conforme feito neste capítulo:

- (a) discretizar a distribuição de severidade no tempo t (x_t), obtendo um vetor $f_{x_t}(0), f_{x_t}(1), \dots, f_{x_t}(n-1)$, onde $n = 2^r$ é o número de pontos desejados na distribuição do sinistro agregado (S_t^*). Deve-se escolher n grande o suficiente tal que $P(S_t^* < n) \approx 1$;
- (b) aplicar a FFT a esse vetor de probabilidades, obtendo a função característica da distribuição de severidade $\phi_{x_t}(z)$;
- (c) usando a função geradora de probabilidade da distribuição do número de sinistros no tempo t ($P_{N_t}(z)$, equação 2-22), obter a função característica da distribuição do sinistro agregado no tempo t : $\phi_{S_t^*}(z) = P_{N_t}[\phi_{x_t}(z)]$; e
- (d) aplicar a inversa da FFT a $\phi_{S_t^*}(z)$ para recuperar o vetor de probabilidades, obtendo a distribuição do sinistro agregado no tempo t .

2.5 Aplicação

2.5.1 Dados

A base de dados usada neste capítulo consiste em sinistros individuais de seguro de automóveis de uma seguradora brasileira, ocorridos entre 01/2006 e 02/2014, totalizando 30.959 sinistros em 98 meses. A Figura 2.1 apresenta as séries temporais dos dados (número de sinistros, severidade média e sinistro agregado). Os valores de severidade foram deflacionados usando índice de inflação de consumo IPCA¹, estando expressos em unidades-índice². A Figura 2.2 apresenta o boxplot mensal da série do número de sinistros que não aponta um comportamento sazonal significativo.

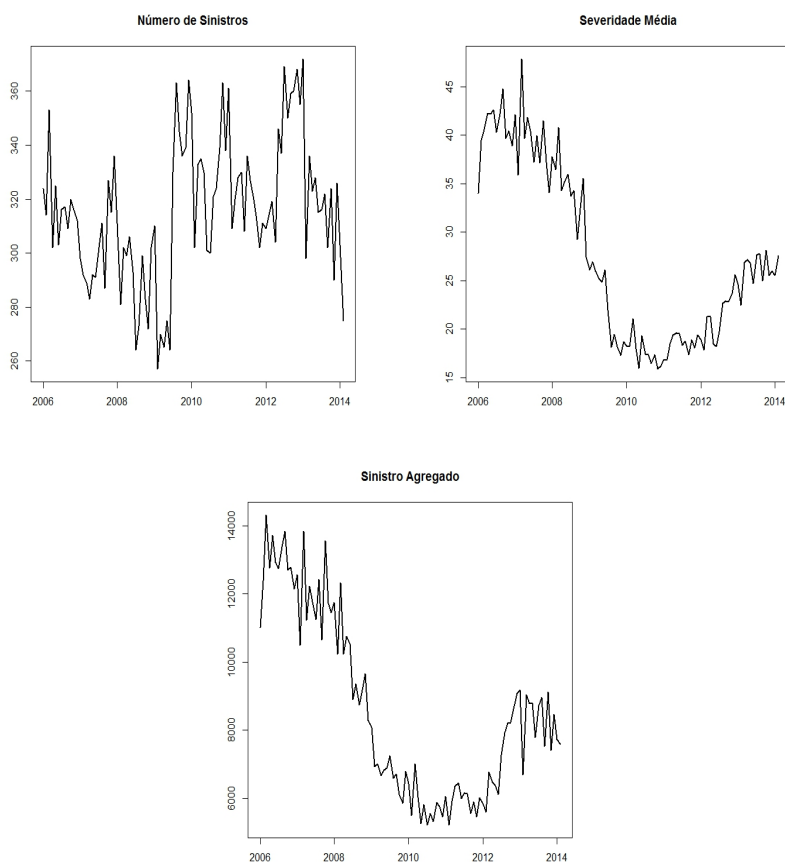


Figura 2.1: Séries temporais de números de sinistros, severidade média e sinistro agregado no período de 01/2006 a 02/2014.

¹Fonte: IBGE, Sistema Nacional de Índices de Preços ao Consumidor.

²As unidades-índice são obtidas pela divisão dos valores correntes pelo índice do mês correspondente, a partir do índice 1000 para jan/2006.

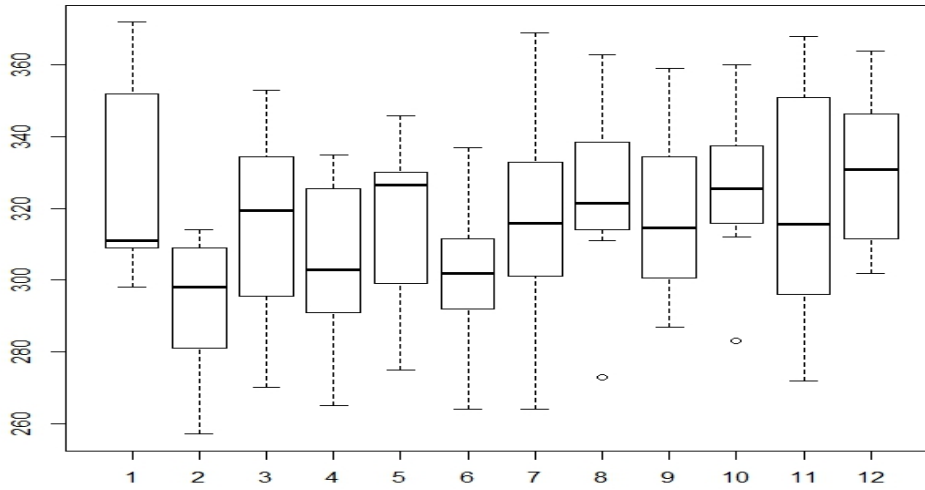


Figura 2.2: Box-plot mensal da série de número de sinistros com dados de 01/2006 a 02/2014.

2.5.2 Resultados

Os modelos apresentados nas subseções 2.4.1 e 2.4.2 foram implementados utilizando a linguagem R (R Core Team [47]). Para o ajuste dos modelos, foram utilizadas as 86 primeiras observações, deixando as 12 observações restantes para avaliação fora da amostra. Utilizamos modelos com $p = q = 1$, na dinâmica de atualização de f_t (equação 2-6), e com matriz de ponderação igual à inversa da informação de Fisher ($S_t = I_{t|t-1}^{-1}$). Os modelos também foram ajustados com matriz de ponderação igual à raiz quadrada da inversa da informação de Fisher ($S_t = I_{t|t-1}^{-1/2}$). Entretanto, como a otimização por máxima verossimilhança se mostrou mais estável com a primeira opção, adotamos $S_t = I_{t|t-1}^{-1}$. Na Tabela 2.1³, apresentamos as estimativas de máxima verossimilhança dos parâmetros fixos dos modelos, bem como seus erros padrão, obtidos a partir da matriz Hessiana calculada numericamente.

Como forma de avaliação e comparação do ajuste dos modelos, utilizamos as medidas AIC (Critério de Informação de Akaike), BIC (Critério Bayesiano de Schwarz), o valor do logaritmo da verossimilhança avaliada com valores de $\hat{\theta}$, bem como a medida de erro MAPE⁴ (*Mean Absolute Percentage Error*). Os resultados apresentados na Tabela 2.2 indicam que o modelo com distribuição gama se ajusta melhor aos dados de severidade, enquanto que os modelos

³Nessa tese, foi adotada a notação numérica americana, com o ponto como separador decimal.

⁴MAPE = $\frac{100\%}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_{t|t-1}}{y_t} \right|$

Tabela 2.1: Estimativas de máxima verossimilhança e erros padrões das estimativas referentes aos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1) para número de sinistros e aos modelos GAS Gama(1,1) e GAS Lognormal(1,1) para severidade.

Parâmetros	Número de sinistros		Severidade	
	Poisson GAS(1,1)	Bin. Neg. GAS(1,1)	Gama GAS(1,1)	Lognormal GAS(1,1)
w	1.0245*** (0.1389)	0.2187 (0.1622)	0.0463 (0.0396)	0.0270 (0.0335)
A_1	0.5170*** (0.0908)	0.5188*** (0.1038)	0.5772*** (0.0854)	0.4225*** (0.0794)
B_1	0.8219*** (0.0241)	0.8225*** (0.0888)	0.9841*** (0.0121)	0.9826*** (0.0172)
r	-	1081.05 (727.35)	-	-
α	-	-	0.4785*** (0.0034)	-
σ	-	-	-	1.9832*** (0.0085)

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

Poisson e binomial negativa para número de sinistros apresentam desempenho similares. Cabe destacar que o parâmetro de dispersão (r) da distribuição binomial negativa não se mostrou estatisticamente significativa.

Tabela 2.2: Resultados *dentro da amostra*, período de 01/2006 a 02/2013, dos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1) para número de sinistros e modelos GAS Gama(1,1) e GAS Lognormal(1,1) para severidade.

	Modelo	\log -Veross.	AIC	BIC	MAPE(%)
<i>Número de sinistros</i>					
	GAS Poisson(1,1)	-381.967	771.935	781.752	4.872
	GAS Binomial Negativa(1,1)	-380.412	770.825	783.097	4.870
<i>Severidade</i>					
	GAS Gama(1,1)	-108790.3	217590.6	217602.9	7.042
	GAS Lognormal(1,1)	-109255.4	218520.8	218533.1	5.007

As Figuras 2.3 e 2.4 ilustram os ajustes dos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1) para a série de número de sinistros dentro da amostra. Para o modelo GAS Poisson, houve comparação dos valores observados da série com valores estimados do parâmetro λ_t e, para o modelo GAS Binomial Negativa, fizemos a comparação dos valores observados da série com valores da média estimada. As Figuras 2.5 e 2.6 ilustram os ajustes dos modelos GAS Gama(1,1) e GAS Lognormal(1,1) para a série de severidade dentro da amostra. Para o modelo GAS Gama, é apresentada a comparação

dos valores estimados do parâmetro variante no tempo μ_t (que corresponde à média condicional da distribuição preditiva) com a média dos valores de severidade observados a cada instante t ; já para o modelo GAS Lognormal, é apresentada a comparação dos valores estimados do parâmetro variante no tempo μ_t com a média do logaritmo dos valores de severidade observados a cada instante t , uma vez que

$$(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) \sim \text{Gama}\left(\alpha, \frac{\mu_t}{\alpha}\right) \rightarrow E\left[\frac{1}{N_t} \sum_{i=1}^{N_t} x_{ti} | N_t\right] = \mu_t$$

e

$$(x_{i,t}|f_t, \mathcal{F}_{t-1}; \theta^{(2)}) \sim \text{Lognormal}(\mu_t, \sigma^2) \rightarrow E\left[\frac{1}{N_t} \sum_{i=1}^{N_t} \ln x_{ti} | N_t\right] = \mu_t.$$

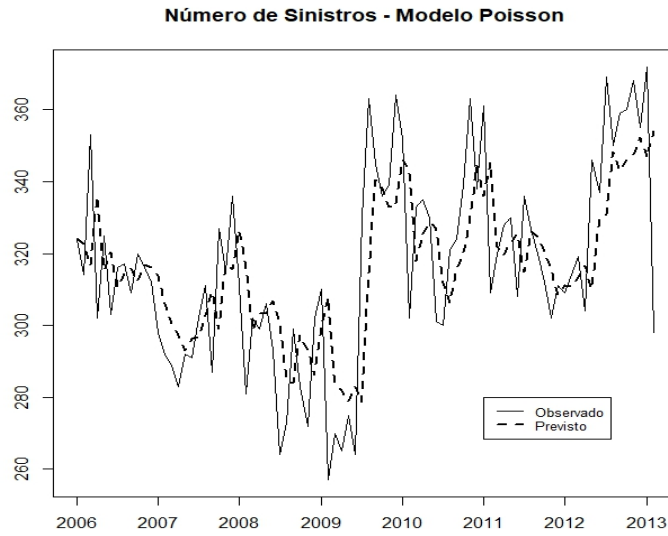


Figura 2.3: Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Poisson(1,1) no período de 01/2006 a 02/2013.

2.5.2.1 Diagnósticos

Para fins de diagnóstico e avaliação dos modelos propostos, utilizamos os resíduos quantílicos, propostos por Dunn & Smyth [48] e Kalliovirta [49] para modelos de séries temporais não lineares. Dunn & Smyth [48]

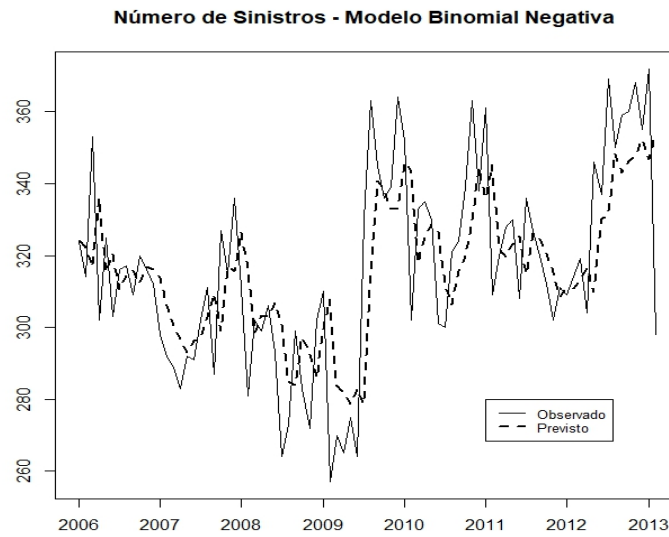


Figura 2.4: Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Binomial Negativa(1,1) no período de 01/2006 a 02/2013.

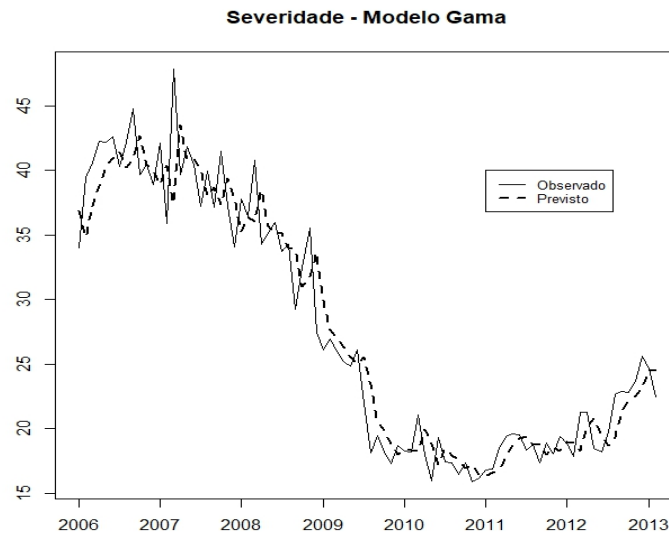


Figura 2.5: Média dos valores de severidade observados, em linha sólida, versus valores estimados do parâmetro variante no tempo μ_t , em linha pontilhada, pelo modelo GAS Gama(1,1) no período de 01/2006 a 02/2013.

apresentam formulação para os resíduos quantílicos tanto para distribuições contínuas como para distribuições discretas. Kalliovirta [49] defende o uso dos resíduos quantílicos normalizados para modelos não lineares ou com mistura de distribuições e apresenta testes estatísticos específicos. Kalliovirta [49]

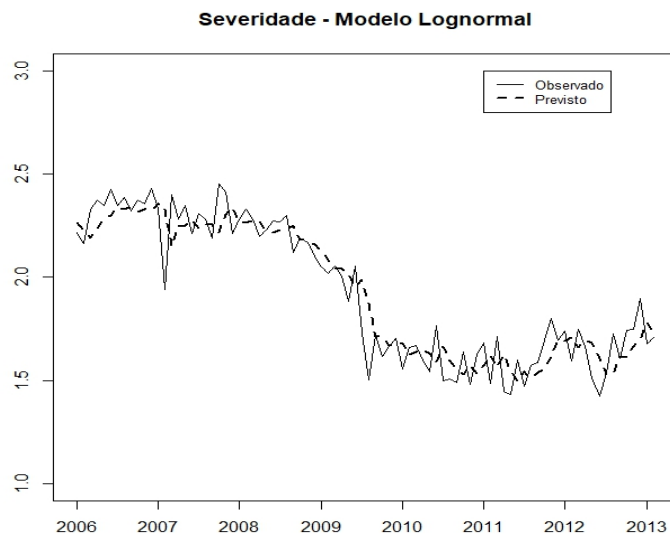


Figura 2.6: Média do logaritmo dos valores de severidade observados, em linha sólida, versus valores estimados do parâmetro variante no tempo μ_t , em linha pontilhada, pelo modelo GAS Lognormal(1,1) no período de 01/2006 a 02/2013.

ressalta que os resíduos quantílicos e os resíduos de Pearson são idênticos para vários modelos com verossimilhança Gaussiana, como modelos lineares padrão e modelos lineares e não lineares autorregressivos com erros normalmente distribuídos. Nesse sentido, os resíduos quantílicos poderiam ser considerados uma generalização dos resíduos de Pearson.

Os resíduos quantílicos utilizados nesse trabalho são definidos por duas transformações: na primeira, utiliza-se a função de distribuição acumulada estimada para transformar as observações em aproximadamente variáveis uniformes independentes; na segunda, utiliza-se a inversa da função de distribuição acumulada da Normal padrão para obter variáveis aproximadamente independentes e normalmente distribuídas. Conforme descrito por Kalliovirta [49], a partir dessas transformações, sob correta especificação do modelo, os resíduos obtidos seriam assintoticamente independentes e normalmente distribuídos, facilitando o uso de testes de diagnósticos.

De acordo com Dunn & Smyth [48], os resíduos quantílicos teóricos são definidos como

$$R_{t,\theta} = \Phi^{-1}(F_t(y_t|f_t, \mathcal{F}_{t-1}; \theta)),$$

e os resíduos quantílicos observados como $r_{t,\hat{\theta}} = \Phi^{-1}(F_t(y_t|f_t, \mathcal{F}_{t-1}; \hat{\theta}))$, onde $\Phi^{-1}(\cdot)$ é a inversa da função de distribuição acumulada da Normal padrão, $F_t(\cdot)$

é a função de distribuição acumulada condicional de y_t e $\hat{\theta}$ é um estimador do vetor de parâmetros θ .

Para o caso de distribuições discretas, Dunn & Smyth [48] definem o resíduo quantílico randomizado observado como $r_{t,\hat{\theta}} = \Phi^{-1}(u_t)$, sendo u_t uma variável aleatória uniforme no intervalo $(a_t, b_t]$, com $a_t = \lim_{y \uparrow y_t} F_t(y|f_t, \mathcal{F}_{t-1}; \hat{\theta})$ e $b_t = F_t(y_t|f_t, \mathcal{F}_{t-1}; \hat{\theta})$.

Para avaliação dos resíduos quantílicos, realizamos testes de autocorrelação serial (teste Ljung-Box até o *lag* 30), heterocedasticidade condicional (teste Ljung-Box até o *lag* 30 nos resíduos quadrados) e de normalidade (teste de Jarque-Bera). Os resultados dos testes, apresentados na Tabela 2.3, foram satisfatórios para todos os modelos, com exceção do modelo GAS Lognormal para severidade, com indicação de não normalidade dos resíduos. Os gráficos apresentados nas Figuras A.1, A.2 e A.3 (Apêndice A) também confirmam o ajuste adequado dos modelos GAS Poisson, GAS Binomial Negativa e GAS Gama.

Tabela 2.3: P-valores dos testes aplicados aos resíduos quantílicos dos modelos GAS Poisson(1,1) e GAS Binomial Negativa(1,1), para número de sinistros, e modelos GAS Gama(1,1) e GAS Lognormal(1,1) para severidade - período 01/2006 a 02/2013. Teste Ljung-Box até *lag* 30.

Modelo	Ljung-Box	Ljung-Box res ²	Jarque-Bera
<i>Número de sinistros</i>			
GAS Poisson(1,1)	0.345	0.372	0.573
GAS Binomial Negativa(1,1)	0.347	0.357	0.523
<i>Severidade</i>			
GAS Gama(1,1)	0.940	0.319	0.860
GAS Lognormal(1,1)	0.939	0.105	0.005

2.5.3 Previsão

A previsão um passo à frente é realizada de forma direta nos modelos GAS, uma vez que a evolução do vetor do parâmetro variante no tempo f_t leva à obtenção da distribuição preditiva $p(y_{t+1}|f_{t+1}, \mathcal{F}_t; \theta)$. Para k -passos à frente, $k \geq 2$, as previsões são obtidas por simulação de Monte Carlo. Para os intervalos de confiança de previsões fora da amostra, adotamos a metodologia para intervalos de múltipla filtragem proposta por Blasques et al. [50]. Esses intervalos de confiança levam em conta tanto a incerteza do vetor de parâmetro estático θ como a incerteza da inovação. A construção do IC se inicia com a geração de M valores θ^i da distribuição de $\hat{\theta}_T$ e, em seguida requer que seja rodado o filtro (equação de atualização 2-6) de $t = 1$ a $t = T$ para cada valor

simulado θ^i , o que resulta em M cenários gerados para \hat{f}_t^i . Tendo em vista que a distribuição amostral finita de $\hat{\theta}_T$ é desconhecida, simulamos valores da distribuição assintótica

$$\hat{\theta}_T^i \sim N(\hat{\theta}_T, T^{-1}\hat{W}),$$

sendo \hat{W} obtida por meio da inversa da matriz Hessiana.

A partir de cada um dos M pares $(\hat{\theta}_T^i, \hat{f}_{T+1}^i)$, geramos S valores $\hat{y}_{T+1}^1, \dots, \hat{y}_{T+1}^S$. Utilizando a equação de atualização do parâmetro variante no tempo, obtemos $M \times S$ valores $\hat{y}_t^{i,s}$ e $\hat{f}_t^{i,s}$ a cada instante $t = T + 2, \dots, T + k$. O algoritmo para esse procedimento está descrito na seção A.2 do Apêndice A.

As observações dos últimos 12 meses foram deixadas de fora da amostra para avaliação da habilidade de previsão dos modelos propostos. As Figuras 2.7 e 2.8 apresentam as previsões e intervalos de confiança fora da amostra para os modelos de número de sinistro (GAS Poisson e GAS Binomial Negativa) e as Figuras 2.9 e 2.10 apresentam as previsões e intervalos de confiança fora da amostra para os modelos de severidade (GAS Gama e GAS Lognormal).

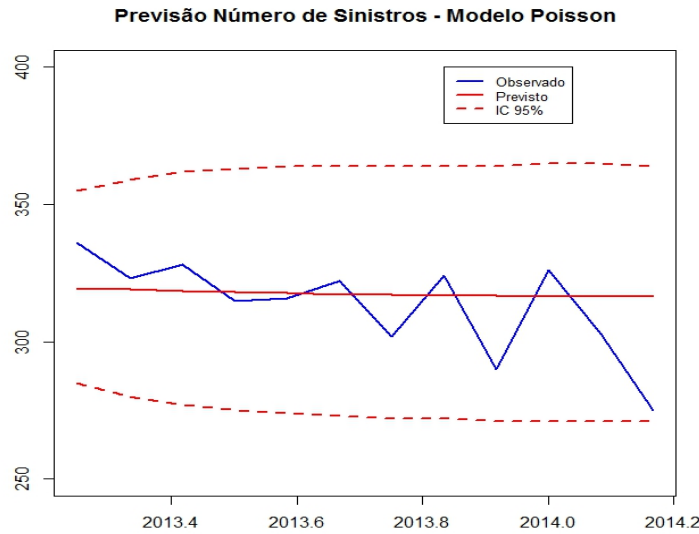


Figura 2.7: Previsão *fora da amostra* do modelo GAS Poisson(1,1) para número de sinistros no período de 03/2013 a 02/2014.

A Tabela 2.4 apresenta os valores previstos por cada um dos modelos e os valores observados para os 12 meses de previsão, bem como das medidas de erro MAPE. Assim como no ajuste dos modelos dentro da amostra, percebemos resultados muito similares para os modelos GAS Poisson e GAS

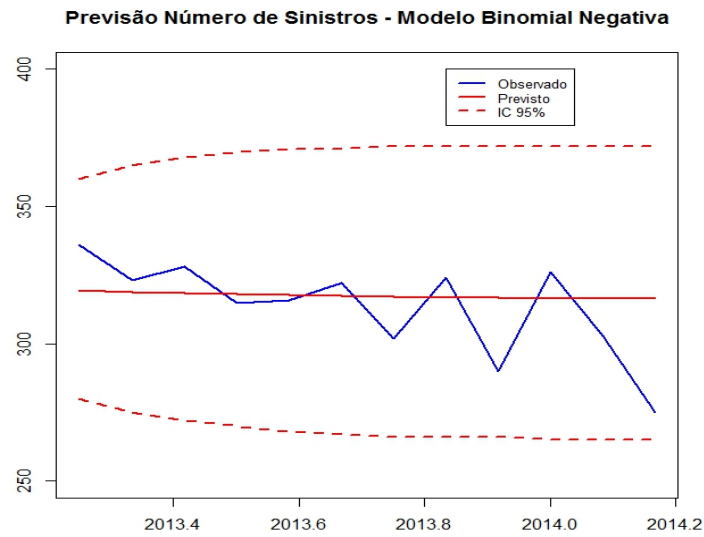


Figura 2.8: Previsão *fora da amostra* do modelo GAS Binomial Negativa(1,1) para número de sinistros no período de 03/2013 a 02/2014.

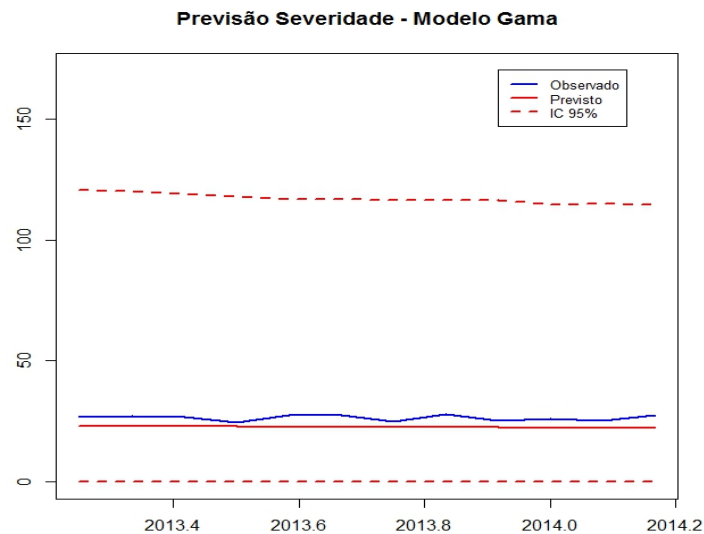


Figura 2.9: Previsão *fora da amostra* do modelo GAS Gama(1,1) para severidade no período de 03/2013 a 02/2014.

Binomial Negativa para número de sinistros, bem como um melhor desempenho do modelo GAS Gama para severidade. Em comparação com um modelo *benchmark* estático, usando as mesmas distribuições de probabilidade e estimação por máxima verossimilhança como descrito em Kaas et al. [22], nossa abordagem apresentou melhores resultados, indicando a pertinência da modelagem dinâmica proposta. Os valores previstos e MAPE referentes aos modelos

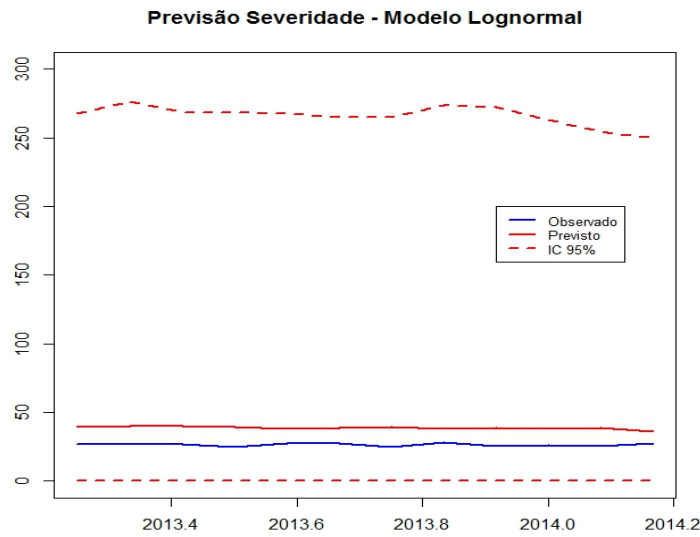


Figura 2.10: Previsão *fora da amostra* do modelo GAS Lognormal(1,1) para severidade no período de 03/2013 a 02/2014.

estáticos também podem ser vistos na Tabela 2.4.

Tendo em vista que o ajuste do modelo GAS Lognormal com $p = q = 1$ não foi satisfatório, com desvio da hipótese de normalidade dos resíduos, sob correta especificação do modelo, e baixa capacidade preditiva, ajustamos também o modelo GAS considerando $p = q = 2$ na dinâmica de atualização de f_t (equação 2-6). Entretanto, o modelo continuou com desempenho insatisfatório, também apresentando desvio da hipótese de normalidade. A partir dos resultados, verificamos que as distribuições condicionais da severidade (x_{it}) oriundas do modelo GAS Lognormal possuíam variância muito elevadas, com caudas bem longas, quando comparadas com as distribuições condicionais do modelo GAS Gama. Para essa base de dados de valores de sinistros, portanto, as distribuições com caudas mais longas não se mostraram adequadas.

Tabela 2.4: Previsão *fora da amostra* e MAPE dos modelos para número de sinistros (GAS Poisson(1,1), GAS Binomial Negativa(1,1) e modelos estáticos) e modelos para severidade (GAS Gama(1,1), GAS Lognormal(1,1) e modelos estáticos) - período de 03/2013 a 02/2014.

[illegible]

2.5.4

Distribuição do sinistro agregado

Para obtenção da distribuição do sinistro agregado 12 passos à frente, utilizamos a metodologia descrita na subseção 2.4.3, com $n = 2^{18}$. Conforme já mencionado, uma das vantagens dessa metodologia é obter a distribuição integralmente, e não apenas alguns momentos. Considerando os ajustes dos modelos propostos, o modelo GAS Poisson, por parcimônia, foi adotado para o número de sinistros, uma vez que o ajuste e o desempenho dos modelos com distribuições Poisson e binomial negativa foram muito similares. Para severidade dos sinistros, foi escolhido o modelo GAS Gama tendo em vista o desempenho ruim do modelo GAS Lognormal. Adicionalmente, com intuito de comparação, utilizamos a aproximação Normal para o sinistro agregado, descrita por Kass et al. [22] (p. 58), que consiste em ajustar uma distribuição Normal com média e variância obtidas com base nas equações 2-3 e 2-4 e nos resultados dos modelos estáticos Poisson e Gama.

A partir da distribuição do sinistro agregado foi possível calcular o VaR de forma a dimensionar perdas extremas para a seguradora e auxiliar na gestão de risco. A Tabela 2.5 apresenta os valores observados de sinistro agregado nos meses de 03/2013 a 02/2014, bem como o VaR 95% calculado para o mesmo período com os modelos GAS Poisson e GAS Gama e a técnica FFT e modelos estáticos Poisson-Gama com aproximação Normal. A distribuição do sinistro agregado obtida a partir da aproximação Normal com resultados dos modelos estáticos para número de sinistros e severidade apresentou quantis inferiores aos valores calculados com os modelos dinâmicos GAS e FFT. Percebe-se, portanto, que os modelos estáticos com aproximação Normal estimariam medidas de risco, como VaR, inferiores, o que poderia acarretar prêmios, provisões e necessidades de capital subdimensionados.

A Figura 2.11 apresenta alguns quantis, assim como as médias, das distribuições do sinistro agregado para os 12 meses à frente obtidas com o uso da FFT e os modelos GAS Poisson e GAS Gama. A Figura 2.12 apresenta as distribuições de sinistro agregado para cada um dos 12 meses fora da amostra (03/2013 a 02/2014).

2.6

Considerações Finais

Esse capítulo propõe uma nova abordagem para obtenção da distribuição de sinistro agregado condicionada no tempo. Utilizamos um novo arcabouço de modelos para séries temporais, orientados por *score* com parâmetros variantes no tempo, para modelar a distribuição preditiva do número de sinistros

Tabela 2.5: Valores observados de sinistro agregado (S_t^*), no período de 03/2013 a 02/2014, e valores de VaR 95%, para o mesmo período (fora da amostra), obtidos a partir dos modelos GAS Poisson e Gama e modelos estáticos Poisson e Gama com aproximação Normal para sinistro agregado.

Meses à frente	S_t	VaR 95% GAS	VaR 95% estático
1	9,034.14	9,504.60	8,794.61
2	8,779.50	9,500.40	8,794.61
3	8,798.25	9,345.60	8,794.61
4	7,788.88	9,332.40	8,794.61
5	8,724.52	9,238.20	8,794.61
6	8,955.08	9,198.00	8,794.61
7	7,541.92	9,133.80	8,794.61
8	9,115.49	9,118.80	8,794.61
9	7,406.60	9,139.80	8,794.61
10	8,469.20	9,130.80	8,794.61
11	7,737.99	9,099.00	8,794.61
12	7,583.84	9,066.60	8,794.61

PUC-Rio - Certificação Digital Nº 1321810/CA

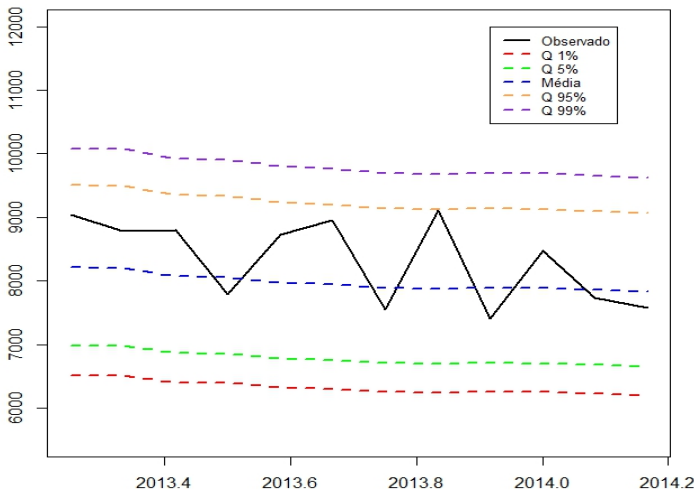


Figura 2.11: Quantis e médias das distribuições do sinistro agregado dos 12 meses de fora da amostra (03/2013 a 02/2014) obtidas a partir dos modelos GAS Poisson(1,1) e GAS Gama(1,1) e do uso da FFT.

e dos valores de sinistros (severidade). Em comparação com distribuições estáticas, as distribuições preditivas se apresentam mais apropriadas por levar em consideração a dinâmica temporal e a incerteza nos parâmetros. Os modelos orientados por *score* apresentados nesse trabalho para o mercado de seguros têm como vantagem o fato de a função de verossimilhança ter forma fechada e os parâmetros estocásticos serem perfeitamente previsíveis a partir da informação passada. Além disso, o fato de o modelo estar baseado no *score*

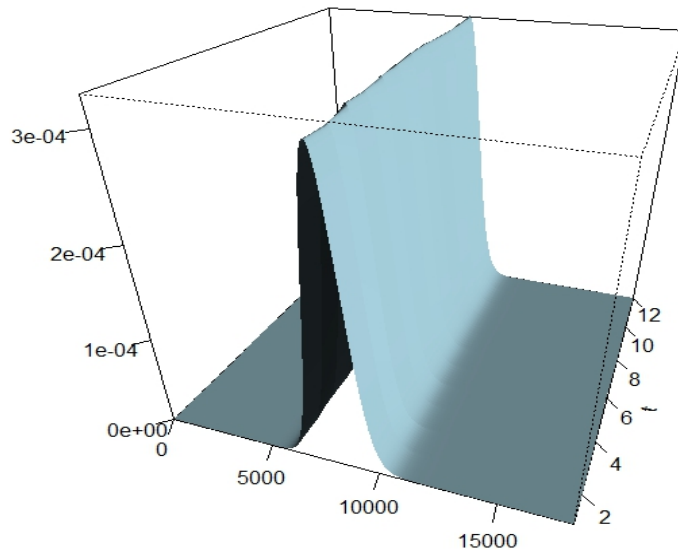


Figura 2.12: Distribuições do sinistro agregado para os 12 meses de fora da amostra (03/2013 a 02/2014), obtidas com o uso da FFT e do modelo GAS Poisson(1,1) para número de sinistros e modelo GAS Gama(1,1) para severidade.

possibilita que seja explorada toda a estrutura da densidade preditiva e não apenas alguns de seus momentos.

Os modelos propostos foram aplicados a dados de seguros de automóveis de uma seguradora brasileira. Os ajustes dos modelos foram avaliados por meio de medidas de ajustes, análise de resíduos, bem como testes de acurácia de previsão. Em comparação com um modelo *benchmark* estático, usando as mesmas distribuições de probabilidade e estimação por máxima verossimilhança, nossa abordagem dinâmica orientada por *score* apresentou melhores resultados. Após avaliação e seleção dos modelos mais apropriados para número de sinistros (modelo GAS Poisson) e severidade (modelo GAS Gama), foi utilizada a técnica da Transformada Rápida de Fourier para obtenção da distribuição de sinistro agregado para 12 meses à frente. Como resultado, obtivemos a densidade completa para a variável aleatória sinistro agregado e não somente médias e momentos superiores.

A partir dessa distribuição, é possível calcular a probabilidade de que a seguradora tenha uma perda na carteira acima de determinado valor, efetuar cálculos de prêmios, avaliação de contratos de resseguro, probabilidade de ruína da empresa, bem como obter subsídios para cálculos de provisão e capital

baseado em risco. As aplicações de nosso método vêm ao encontro de demandas dos atuais princípios de solvência e de mensuração de contratos de seguros ao possibilitar a obtenção de distribuições preditivas, condicionadas no tempo, da perda de carteiras de seguradoras. A aferição da incerteza em relação a este processo estocástico permite o exercício de práticas de gestão de risco, tanto por meio da obtenção de medidas, como VaR, TVaR ou outras, quanto para cálculo de margem de risco.

Até onde sabemos, existem poucos estudos sobre a obtenção da distribuição preditiva do sinistro agregado e a maioria deles utilizam abordagem Bayesiana, que exige um custo computacional alto, além de requerer conhecimentos específicos para implementação dos métodos de amostragem. É importante ressaltar que a modelagem proposta pode ser facilmente aplicada a outras linhas de negócio, inclusive na área de risco operacional e de crédito, assim como desenvolvida com diferentes distribuições de probabilidade que melhor se ajustem aos dados.

3

Avaliação consistente de mercado para seguros financeiros: uma abordagem multivariada usando modelos orientados por score, fatores macroeconômicos e cópulas

Nesse capítulo, uma abordagem completa é proposta para modelar número de sinistros de seguros financeiros, em linha com as recomendações internacionais para órgãos supervisores e reguladores da indústria de seguros sobre avaliação consistente de mercado. Trabalhamos com dados de duas linhas de negócios de seguros financeiros: seguro de fiança locatícia e seguro de crédito interno. Ambas apresentam como característica marcante o fato de o sinistro ser decorrente de um evento de *default*. Utilizamos variáveis macroeconômicas de forma a tornar as previsões consistentes com o cenário econômico e, uma vez que estamos lidando com uma série bivariada, avaliamos e fizemos uso de modelos de cópulas para considerar a estrutura de dependência na previsão de números de sinistros das duas séries para abordagens conjuntas.

3.1

Introdução

Avaliação consistente de mercado de provisões de seguro é um dos requisitos mais importantes no Solvência II¹ e também nos estudos do IFRS - *Insurance contracts Phase II*. Essa forma de mensurar provisões técnicas também está presente nos princípios da IAIS [16]. De acordo com esses requerimentos, é necessário aplicar a chamada “abordagem do balanço total” na preparação das demonstrações financeiras de companhias seguradoras. Esse conceito implica que ativos e passivos devam ser avaliados de uma maneira coerente e consistente, usando avaliação econômica, também conhecida como avaliação realista. Uma avaliação apropriada das demonstrações financeiras desempenham um papel fundamental na aferição da saúde financeira como um todo de uma seguradora

¹Solvência II é uma directiva da União Europeia (UE), que codifica e harmoniza a regulamentação de seguros da UE, mas que tem também servido de base para órgãos supervisores e reguladores da indústria de seguros de fora da Europa. Solvência II está baseado em três pilares: Pilar I - Requerimento de Capital / Pilar II - Governança e Supervisão / Pilar 3 - Divulgação e transparência.

Observa-se que a primeira etapa para aferição da saúde financeira de uma seguradora é a avaliação de ativos e passivos sob o ponto de vista de solvência. A fim de permitir uma aferição de solvência adequada para a empresa, através da comparação entre o capital requerido e o disponível, é necessário primeiramente calcular o capital disponível. Seu valor é obtido através da diferença entre o valor de mercado dos ativos elegíveis e o valor consistente do passivo. De acordo com Solvência II [4] e IAIS [16], a avaliação consistente significa que se deve levar em consideração as informações disponíveis nos mercados financeiros e de capital relevantes, utilizando metodologias e parâmetros consistentes com esses mercados. Em nossa pesquisa, utilizamos essas fontes de informação na modelagem dos sinistros de seguros financeiros. Nossa abordagem tem a vantagem de usar dados macroeconômicos, a fim de atender à consistência da previsão com o mercado financeiro e, uma vez que estamos lidando com série temporal multivariada (mais de uma linha de negócios), também investigamos a existência de dependência não captada pelos modelos univariados e, em caso positivo, fazemos uso de modelos cópula para considerar a estrutura de dependência ao simular números de sinistros futuros de mais de uma carteira para abordagens conjuntas.

Modelagem do número de sinistros em seguros de crédito, incluindo suas diferentes linhas de negócio, está intimamente ligada à modelagem de risco de *default*. Na verdade, um evento de *default*, se coberto por uma apólice de seguro, é caracterizado por um sinistro para uma companhia seguradora. Em geral, modelos para avaliar risco de *default* podem ser separados em duas amplas categorias: modelos estruturais e de forma reduzida. A base dos modelos estruturais é certamente Merton [51]. Baseado na teoria do valor da firma, esses modelos adotam premissas a respeito das equações diferenciais que governam o processo do valor do ativo da firma. Assume-se que uma companhia emite um débito, com vencimento em T , para financiar suas atividades. No tempo T , se a companhia não for capaz de honrar seus compromissos com os credores, é possível afirmar que a empresa está em *default*. Deste modo, o *default* pode ocorrer apenas no final de T e é acionado quando o valor da empresa é menor do que o valor de débito.

Neste trabalho, focamos na abordagem da forma reduzida, que trata o *default* como um evento aleatório que pode acometer qualquer firma ou indivíduo em qualquer tempo. Conforme apontado por Figlewski et al. [52], no modelo básico de forma reduzida, um evento de crédito corresponde a um salto em um processo de Poisson com uma determinada taxa de intensidade. Nesse caso, o “evento” pode ser definido de forma flexível como *default* ou falta de cumprimento de obrigações contratuais, ou qualquer outra alteração

bem definida do estado. A abordagem de forma reduzida tem sido amplamente utilizada para análise de risco de crédito tanto no campo acadêmico como pelo mercado - vide, por exemplo, Duffie et al. [53] e Koopman et al. [54].

Muitos trabalhos na literatura abordam a relação entre exposição ao risco de crédito e condições da macroeconomia. Bangia et al. [55] apontaram que intensidades de *upgrade*, *downgrade* e *default* diferem entre os diferentes regimes econômicos. Yan et al. [56] usaram a abordagem de forma reduzida para extrair intensidades de *default* de títulos corporativos com grau de investimento e encontrou fortes evidências de fatores comuns, sendo o mais forte o desempenho do mercado de ações. Huang & Kong [57] demonstraram que anúncios macroeconômicos têm um impacto significativo sobre os *spreads* de crédito. Diversos outros estudos mostraram que a probabilidade de *default* é sensível a fatores macroeconômicos, tais como Duffie et al. [53].

Figlewski et al. [52] avaliaram como fatores macroeconômicos afetam risco de crédito com uso de modelos de intensidade de Cox para *defaults* e transições entre *rating* de grau de investimento e grau especulativo. Figlewski et al. [52] destacaram que estudos empíricos muitas vezes introduzem variáveis macroeconômicas nos modelos como variáveis contemporâneas. Entretanto, na medida em que os macro fatores estão medindo aspectos da saúde financeira geral, não é plausível que seu efeito sobre padrões seja instantâneo. Em vez disso, espera-se que, por exemplo, altas taxas de juros ou crescimento lento na economia levem, cumulativamente, a um aumento gradual do risco de crédito. Dessa forma, relataram a importância de incluir valores defasados das variáveis macroeconômicas na especificação do modelo e, para tanto, propuseram uma estrutura de defasagem considerando uma média ponderada em uma janela fixa.

Neste contexto, um dos principais *inputs* para o mercado (e também para o Solvência II) avaliar de forma consistente os passivos de seguro de crédito é chegar a uma previsão confiável do número de *default* (sinistros) decorrentes das apólices em vigor. No nosso trabalho, utilizamos dados reais de duas linhas de negócios de seguradoras brasileiras, de importante participação no mercado de seguros financeiros. As linhas de negócios utilizadas foram o seguro de fiança locatícia e o seguro de crédito interno. Ambos apresentam como característica marcante o fato de o sinistro ser o resultado de um evento de *default*. Para o seguro de fiança locatícia, sinistros são decorrentes do *default* do inquilino, enquanto que sinistros de seguro de crédito interno resultam de *defaults* do devedor.

Tendo em vista que nosso objetivo de modelagem é uma série temporal bivariada, modelos adequados foram utilizados para esta tarefa. Nossa abor-

dagem permite o uso de preditores autorregressivos baseados em observações anteriores e/ou nos parâmetros estimados. Como variáveis exógenas, usamos variáveis macroeconômicas, como taxa de juros, inflação, PIB e taxa de câmbio. Todas as séries temporais são mensais. Para modelar a estrutura de dependência entre as duas linhas de negócio, utilizamos modelos de cópulas.

Esse capítulo está organizado da seguinte maneira: na seção 3.2, serão apresentados a metodologia e as abordagens para modelagem de séries temporais de contagem, bem como uma parte de teoria de cópulas adotada no trabalho. Na seção 3.3, será apresentada a aplicação da modelagem a dados reais, com descrição dos dados e resultados. Finalmente, na seção 3.4, apresentaremos as considerações finais e indicação de possíveis trabalhos futuros. No Apêndice B, estão os gráficos relativos aos resíduos quantílicos dos modelos trabalhados, formulações de cópulas e o algoritmo para previsão com modelos GAS com cópulas.

3.2

Metodologia

Nossa abordagem para modelar uma série de tempo bivariada faz uso de diferentes famílias de distribuições, tanto para as marginais como para as funções de cópulas. A modelagem usando cópulas permite fatorar a distribuição multivariada em suas distribuições univariadas marginais e uma cópula. Na prática, este fato facilita tanto a especificação da distribuição multivariada como a sua estimação. Seguindo o método IFM² (*Inference Functions for Margins*), introduzido por Joe [58], primeiro ajustamos as distribuições das marginais e então, em um segundo passo, ajustamos a cópula às marginais padronizadas Uniforme(0,1). Um esquema básico ilustrando a metodologia é apresentado na Figura 3.1.

A abordagem proposta foi aplicada em duas etapas. Na primeira etapa, as séries temporais bivariadas são inicialmente modeladas separadamente, com ambas usando preditores autorregressivos e regressores exógenos guiados por fatores macroeconômicos. Testaremos duas modelagens diferentes para as marginais: modelos lineares generalizados autorregressivos e modelos orientados por *score*. Na segunda etapa, é investigada a existência de dependência nos resíduos dos modelos e funções de cópulas serão utilizadas para modelar a

²IFM consiste em um método de máxima verossimilhança de 2 etapas aplicado para obter as estimativas dos parâmetros das marginais e da cópula. Joe [58] argumenta que pode-se esperar o método IFM seja bastante eficiente, uma vez que é totalmente baseado na estimação por máxima verossimilhança. O sucesso deste procedimento de estimação começa com bons ajustes para as marginais e, então, ajusta-se a cópula à série bivariada via transformação integral de probabilidade Uniforme(0,1).

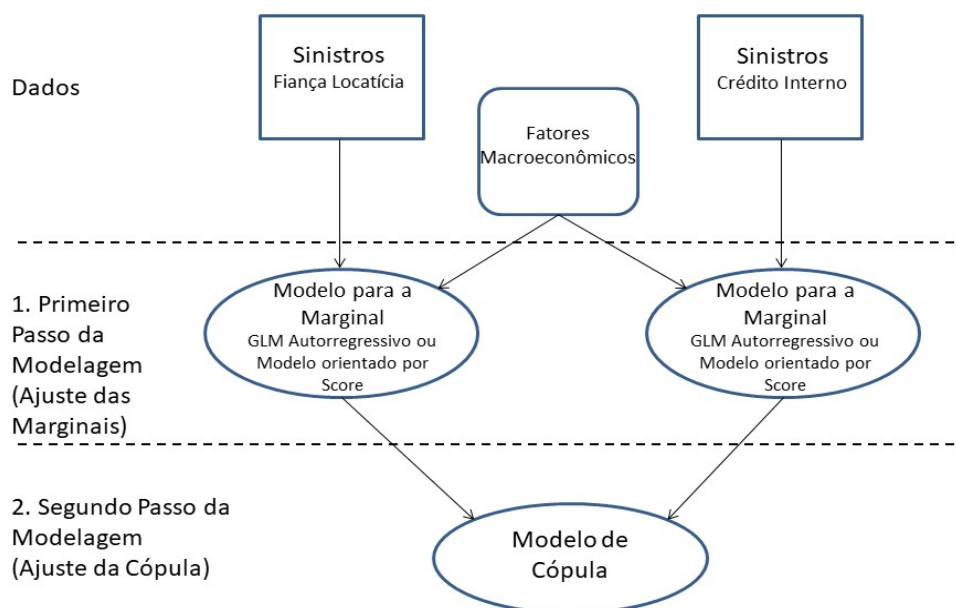


Figura 3.1: Esquema para avaliação de sinistros de seguro de riscos financeiros bivariado usando modelos orientados por *score*, GLM autorregressivo, fatores macroeconômicos e cópulas.

estrutura de dependência das séries temporais condicionado aos modelos ajustados na primeira etapa.

3.2.1 Modelagem para as distribuições marginais

A modelagem para séries temporais de contagem tem sido de grande interesse e merecido diversas publicações. Séries temporais de contagem podem ser observadas em diferentes áreas, sempre que um número de eventos possa ser observado ao longo do tempo. Como exemplo, temos o número de sinistros para uma linha de seguro de negócios de seguros, o número diário de internações hospitalares de saúde pública, o número de transações no mercado de ações por minuto ou o número de itens defeituosos por hora em um controle de qualidade industrial. Nos modelos para séries de contagem, a variável aleatória de contagem deve ser inteira não negativa e a modelagem deve capturar a dependência temporal entre as observações.

Conforme apontado por Liboschik et al. [59], uma abordagem conveniente e flexível é empregar a metodologia dos modelos lineares generalizados (GLM) (Nelder & Wedderburn [60] e McCullagh & Nelder [61]) para modelar as observações condicionalmente às informações passadas, a partir da escolha

de uma distribuição apropriada para dados de contagem e de uma função de ligação adequada. Esta abordagem é detalhada por Kedem & Fokianos [62]. Conforme apontado por Kedem & Fokianos [62], a teoria dos modelos lineares generalizados foi originalmente desenvolvida para dados independentes, mas pode ser estendida, sob algumas hipóteses, para dados dependentes. Os modelos GLM podem levar em conta a autocorrelação por meio do uso de valores defasados da variável dependente na componente sistemática do modelo.

Zeger & Qaqish [63] propuseram uma abordagem de quasi-verossimilhança para regressão de séries temporais e apresentaram diversos modelos, inclusive para séries de contagens. Davis et al. [64] discutiram modelos orientados por observação para contagens Poisson. Jung & Tremayne [65] fazem um detalhado levantamento sobre modelos de séries temporais de contagem. Fokianos [66, 67, 68] também apresenta uma visão geral aprofundada dos modelos para as séries temporais de contagem. Algumas classes de modelos para séries temporais de contagens podem ser citadas: modelos autorregressivos e médias móveis de valor inteiro não-negativo, INARMA (*Integer Autoregressive Moving Average Models*), obtidos através da substituição da multiplicação por um operador aleatório, chamado *thinning*, nos modelos autorregressivos de médias móveis (ARMA) convencionais (ver Weiß[69]); modelos autorregressivos e médias móveis generalizados, GARMA (*Generalized Autoregressive Moving Average Models* - Benjamin et al. [70]), como uma extensão dos modelos ARMA para distribuições condicionais não-Gaussianas; e modelos de espaço de estados, como proposto por Harvey & Fernandes [71].

Neste capítulo, usamos um arcabouço desenvolvido recentemente para a dinâmica de modelos não Gaussianos de séries temporais, os quais serão utilizados para séries de contagem de número de sinistros. Esses modelos são conhecidos como modelos orientados por *score*, uma vez que seus parâmetros variantes no tempo são orientados pelo *score* obtido pelo modelo probabilístico usado para descrever os dados. Os modelos foram ajustados para cada série temporal de contagem (marginais). Para fins de comparação, ajustamos também modelos lineares generalizados com componentes autorregressivos, apresentados por Fokianos & Tjøstheim [72] e Liboschik et al. [59], para cada série temporal.

3.2.1.1

Modelos Lineares Generalizados Autorregressivos

Seja $\{y_t : t \in \mathbb{N}\}$ uma série temporal de contagem e $\{X_t : t \in \mathbb{N}\}$ um vetor de dimensão r de variáveis exógenas variantes no tempo, $X_t = (x_{1t}, \dots, x_{rt})^\top$. Em nossa abordagem, esses seriam os fatores macroeconômicos. A média

condicional $E(y_t|\mathcal{F}_{t-1})$ é modelada por um processo latente $\{\lambda_t : t \in \mathbb{N}\}$, de forma que $E(y_t|\mathcal{F}_{t-1}) = \lambda_t$.

Seja \mathcal{F}_{t-1} o histórico do processo $\{y_{t-1}, \lambda_{t-1}, X_t : t \in \mathbb{N}\}$ até o tempo $t-1$ incluindo informações disponíveis das variáveis exógenas até t . Trabalharemos com a distribuição condicional de $y_t|\mathcal{F}_{t-1}$. Nos modelos lineares generalizados autorregressivos, adotamos a seguinte forma geral:

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^p \beta_k \tilde{g}(y_{t-k}) + \sum_{l=1}^q \alpha_l g(\lambda_{t-l}) + \eta^\top X_t, \quad (3-1)$$

na qual $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ é uma função de ligação e $\tilde{g} : \mathbb{R}^+ \rightarrow \mathbb{R}$ é uma função de transformação. O vetor de parâmetros $\eta_t = (\eta_1, \dots, \eta_r)^\top$ corresponde aos efeitos das variáveis exógenas e $g(\lambda_t)$ é o preditor linear. Neste modelo, o efeito das variáveis exógenas entra diretamente na dinâmica do processo e afeta as observações futuras tanto pela regressão nas observações passadas como pela regressão nas médias latentes passadas.

Liboschik et al. [59] apresentam o modelo log-linear como um caso especial do modelo 3-1:

$$\log(\lambda_t) = \beta_0 + \sum_{k=1}^p \beta_k \log(y_{t-k} + 1) + \sum_{l=1}^q \alpha_l \log(\lambda_{t-l}) + \eta^\top X_t. \quad (3-2)$$

Fokianos & Tjøstheim [72] estudaram o modelo log-linear (3-2) com $p = q = 1$ e demonstraram que a adição de uma constante em cada observação na função de transformação logarítmica de forma a evitar zeros não afeta a inferência e que o valor da constante 1 seria uma escolha razoável. O modelo log-linear implica positividade do processo de média condicional λ_t , o que é necessário em séries temporais de contagem. As propriedades assintóticas dos estimadores de máxima verossimilhança para o referido modelo foram estudadas por Fokianos & Tjøstheim [72].

Assim como utilizado por Haberman & Renshaw [73] e por Neves et al. [74] para taxas de mortalidade, na modelagem de número de sinistros pode-se usar a informação da exposição como variável *offset* no modelo. O termo *offset* é usado para um valor conhecido no componente sistemático do modelo. No caso dos modelos autorregressivos apresentados, o logaritmo da exposição entraria na formulação do preditor linear com coeficiente igual a 1. A exposição considera o número de riscos expostos no mês, portanto, se o risco ficou na carteira durante todo o mês, ele conta como uma unidade no mês; caso tenha se iniciado ou terminado no decorrer do mês, a exposição é proporcional ao número de dias em que esteve na carteira da seguradora. Com a introdução do logaritmo da exposição como variável *offset*, tiramos o efeito da variação

do número de sinistros em função da variação da exposição e acabamos por modelar a taxa de ocorrência de sinistros.

O modelo log-linear (3-2) passa a ter a seguinte formulação com a introdução da exposição no tempo t (r_t) como *offset*:

$$\log(\lambda_t) = \beta_0 + \log(r_t) + \sum_{k=1}^p \beta_k \log(y_{t-k} + 1) + \sum_{l=1}^q \alpha_l \log(\lambda_{t-l}) + \eta^\top X_t. \quad (3-3)$$

Adicionalmente, aplicaremos aos dados um modelo mais simples autor-regressivo, chamado de modelo multiplicativo por Kedem & Fokianos [62] (p. 153) como sendo uma das variações de um modelo introduzido inicialmente por Zeger & Qaqish [63]:

$$\log(\lambda_t) = \beta_0 + \beta_1 \log(\tilde{y}_{t-1}) + \eta^\top X_t,$$

sendo $\tilde{y}_{t-1} = \max(c, y_{t-l})$, $0 < c < 1$.

Com a introdução da exposição no tempo t (r_t) como *offset*, temos que

$$\log(\lambda_t) = \beta_0 + \log(r_t) + \beta_1 \log(\tilde{y}_{t-1}) + \eta^\top X_t. \quad (3-4)$$

Modelo Log-linear Poisson

Para modelagens de contagens, como número de sinistros, distribuições muito utilizadas são Poisson e a binomial negativa.

Assumindo distribuição condicional Poisson, $y_t | F_{t-1} \sim \text{Poisson}(\lambda_t)$, e o modelo log-linear (3-2), temos que:

$$p(y_t | \mathcal{F}_{t-1}) = \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!}, \quad y_t = 0, 1, \dots \quad (3-5)$$

Neste caso, $\text{Var}(y_t | \mathcal{F}_{t-1}) = \text{E}(y_t | \mathcal{F}_{t-1}) = \lambda_t$, ou seja, o processo de média latente é igual à variância condicional do processo observado.

Modelo Log-linear Binomial Negativa

A distribuição binomial negativa se apresenta como uma alternativa interessante para modelagem de contagens, uma vez que possibilita acomodar uma variância condicional maior do que λ_t .

Christou & Fokianos [75] apresentam modelo para série temporal de contagem utilizando a distribuição binomial negativa, com a estrutura prevista em 3-1 e 3-2. Assumindo que $y_t | \mathcal{F}_{t-1} \sim \text{BinNeg}(\lambda_t, \phi)$, na qual a distribuição Binomial Negativa é parametrizada em termos de sua média λ_t e de um parâmetro de dispersão $\phi \in (0, \infty)$, temos que:

$$p(y_t|\mathcal{F}_{t-1}) = \frac{\Gamma(\phi + y_t)}{\Gamma(y_t + 1) + \Gamma(\phi)} \left(\frac{\phi}{\phi + \lambda_t} \right)^\phi \left(\frac{\lambda_t}{\phi + \lambda_t} \right)^{y_t}, \quad y_t = 0, 1, \dots \quad (3-6)$$

Nesse caso, $\text{Var}(y_t|\mathcal{F}_{t-1}) = \lambda_t + \lambda_t^2/\phi$, ou seja, a variância condicional aumenta de forma quadrática com λ_t . A distribuição de Poisson é um caso limite da Binomial Negativa quando $\phi \rightarrow \infty$ (ver Klugman et al. [18], p. 79).

Ambos os modelos serão estimados pelo método de máxima verossimilhança.

3.2.1.2

Modelos orientados por score

Conforme citado anteriormente, usaremos modelos orientados por *score* com parâmetros variantes no tempo, propostos por Creal et al. [13, 14] e Harvey & Chakravarty [15]. Essa nova classe de modelos fornece um arcabouço geral para o desenvolvimento de modelos com quaisquer distribuições condicionais não Gaussianas nos quais os parâmetros variam no tempo. Esses modelos são classificados como orientados por observações na classificação proposta por Cox et al. [12]. A dinâmica do parâmetro é definida por uma equação autoregressiva orientada por um *score* ponderado, obtido da distribuição condicional ou preditiva escolhida para os dados. É também possível introduzir variáveis exógenas na equação de atualização do parâmetro variante no tempo. Nesse arcabouço, os parâmetros estocásticos são perfeitamente previsíveis dada informação passada, assim como nos modelos GARCH de Bollerslev [39].

Como um modelo orientado por observação, a distribuição um passo à frente, e a função de verossimilhança, são conhecidas por construção. As vantagens dos modelos orientados por *score* com parâmetros variantes no tempo são: simplicidade na estimação, dado que a função de verossimilhança é conhecida em sua forma fechada, permitindo estimação direta dos parâmetros fixos e desconhecidos pelo método de máxima verossimilhança; e generalidade, dado que pode ser utilizado com diferentes densidades preditivas. Diferentemente das abordagens tradicionais para a distribuição do número de sinistros, os modelos com parâmetros variantes no tempo permitem uma previsão mais acurada uma vez que o modelo segue a dinâmica do parâmetro ao longo do tempo.

A densidade preditiva do modelo e a equação de atualização associada são dadas por:

$$y_t \sim p(y_t|f_t, \mathcal{F}_{t-1}; \theta), \quad (3-7)$$

e

$$f_{t+1} = w + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1}, \quad (3-8)$$

onde $\mathcal{F}_{t-1} = \{Y^{t-1}, F^{t-1}, X_t\}$, $F^{t-1} = \{f_1, \dots, f_{t-1}\}$, $Y^{t-1} = \{y_1, \dots, y_{t-1}\}$ denota o vetor de observações passadas, $X_t = \{x_{1t}, \dots, x_{rt}\}$ é o vetor de variáveis exógenas (os fatores macroeconômicos), f_t é o vetor de parâmetros variantes no tempo, θ é o vetor de parâmetros estáticos, w é o vetor de constantes, A_i e B_j são as matrizes são matrizes de coeficientes, cujas dimensões dependem de f_t , e s_t é o vetor *score* (∇_t) ponderado pela matriz S_t ,

$$s_t = S_t \nabla_t, \quad (3-9)$$

sendo

$$\nabla_t = \frac{\partial \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t}. \quad (3-10)$$

A matriz de ponderação S_t , em geral, é escolhida entre $I_{t|t-1}^{-1}$, $I_{t|t-1}^{-1/2}$ ou I , sendo $I_{t|t-1} = E_{t-1} [\nabla_t \nabla_t']$ a matriz de informação e I a matriz identidade.

Creal et al. [14] argumentam que o uso da função *score* para atualização de f_t é intuitiva, na medida em que define a direção de subida mais acentuada para melhorar o ajuste local do modelo em termos da verossimilhança no tempo t dada a posição atual do parâmetro f_t . Portanto, o *score* forneceria uma direção natural para atualização do parâmetro. Além disso, o *score* depende da densidade completa e não apenas em momentos de primeira ou segunda ordens das observações y_t . Essa característica distingue os modelos GAS dos demais modelos orientados por observações tratados na literatura.

Blasques et al. [37] apresentam uma justificativa teórica para a optimalidade da função *score* na equação de atualização do parâmetro baseado na distância de Kullback-Leibler (KL). Eles mostram que modelos que utilizam a função *score* na equação de atualização são ótimos do ponto de vista da distância de KL, ou seja, levando em consideração a distância de Kullback-Leibler, não há nenhuma outra dinâmica de atualização que seja mais próxima do verdadeiro. Modelos orientados por *score* tornam possível o uso da densidade preditiva como um todo para fins de inferência, ao invés de apenas alguns momentos da distribuição. Por exemplo, pode-se derivar quantis condicionais, além da média condicional e intervalos de confiança.

O vetor de parâmetros fixos θ é estimado maximizando o logaritmo da função de verossimilhança, $l(\theta)$, cuja expressão pode ser obtida por:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} l(\theta) \quad \text{e} \quad l(\theta) = \sum_{t=1}^n \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta).$$

As propriedades dos estimadores de máxima verossimilhança nos modelos

GAS foram estudadas por Blasques et al. [40, 41]. Para fins de inferência, Creal et al. [14] propõem computar os erros padrões e estatísticas t para os parâmetros estimados, com base na inversa da matriz Hessiana do logaritmo da verossimilhança avaliada no ótimo ($\hat{\theta}$). Sob certas condições de regularidade, o estimador de máxima verossimilhança $\hat{\theta}$ de θ é consistente e satisfaz

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}), \quad (3-11)$$

onde $H = \lim_{n \rightarrow \infty} E[(\partial l / \partial \theta)(\partial l / \partial \theta')]/n$ e $l = \sum_{t=1}^n \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta)$.

Modelo GAS Poisson:

Utilizamos a distribuição de Poisson, apresentada em (3-5), para modelar as séries temporais de número de sinistros (y_t):

$$(y_t | f_t, \mathcal{F}_{t-1}; \theta) \sim \text{Poisson}(\lambda_t), \quad (3-12)$$

Definimos o parâmetro variante no tempo, f_t , em função da média condicional λ_t utilizando uma reparametrização que garanta que a média λ_t seja sempre positiva e, conforme descrito em Creal et al. [14] (p. 789-790), incorporamos as variáveis exógenas diretamente na expressão da média condicional λ_t :

$$f_t = \ln(\lambda_t) - \sum_{i=1}^r \eta_i x_{it}, \quad (3-13)$$

ou seja,

$$\lambda_t = e^{f_t + \sum_{i=1}^r \eta_i x_{it}}.$$

Adotando a exposição no tempo t (r_t) como variável *offset* no modelo, a formulação 3-13 passa a ser

$$f_t = \ln(\lambda_t) - \ln(r_t) - \sum_{i=1}^r \eta_i x_{it}, \quad (3-14)$$

ou seja,

$$\lambda_t = r_t e^{f_t + \sum_{i=1}^r \eta_i x_{it}}.$$

Utilizando as formulações (3-9), (3-10) e (3-13), ou (3-14), aplicadas a (3-5), temos que

$$\nabla_t = y_t - \lambda_t, \quad I_{t|t-1} = \lambda_t$$

e

$$s_t = (y_t - \lambda_t) / \lambda_t^d, \quad \text{para } d = 0, 1 \text{ ou } 1/2.$$

O valor de “ d ” vai depender da escolha da matriz de ponderação S_t ($I, I_{t|t-1}^{-1}$ ou $I_{t|t-1}^{-1/2}$, respectivamente) e é determinado empiricamente a partir de avaliação de ajuste dos modelos.

Modelo GAS Binomial Negativa:

Para o modelo GAS Binomial Negativa utilizamos a mesma parametrização da distribuição binomial negativa apresentada em (3-6):

$$(y_t|f_t, \mathcal{F}_{t-1}; \theta) \sim \text{BinNeg}(\lambda_t, \phi). \quad (3-15)$$

Definimos o parâmetro variante no tempo, f_t , em função da média condicional λ_t utilizando uma reparametrização que garanta que a média λ_t seja sempre positiva, além de incorporar as variáveis exógenas no modelo:

$$f_t = \ln(\lambda_t) - \sum_{i=1}^r \eta_i x_{it}, \quad (3-16)$$

ou seja,

$$\lambda_t = e^{f_t + \sum_{i=1}^r \eta_i x_{it}}.$$

Adotando a exposição no tempo t (r_t) como variável *offset* no modelo, a formulação 3-16 passa a ser

$$f_t = \ln(\lambda_t) - \ln(r_t) - \sum_{i=1}^r \eta_i x_{it}, \quad (3-17)$$

ou seja,

$$\lambda_t = r_t e^{f_t + \sum_{i=1}^r \eta_i x_{it}}.$$

Utilizando as formulações (3-9), (3-10) e (3-16), ou (3-17), aplicadas a (3-6), temos que

$$\nabla_t = \frac{\phi(y_t - \lambda_t)}{\lambda_t(\phi + \lambda_t)}, \quad I_{t|t-1} = \frac{\phi}{\lambda_t(\phi + \lambda_t)}$$

e

$$s_t = \frac{\phi(y_t - \lambda_t)}{(\phi + \lambda_t)} \left[\frac{(\phi + \lambda_t)}{\phi \lambda_t} \right]^d, \quad \text{para } d = 0, 1 \text{ ou } 1/2.$$

Assim como no modelo Poisson, o valor de “ d ” vai depender da escolha da matriz de ponderação S_t ($I, I_{t|t-1}^{-1}$ ou $I_{t|t-1}^{-1/2}$, respectivamente).

3.2.2

Modelagem para estrutura de dependência condicional via cópulas

3.2.2.1

Modelos de cópulas

Uma vez ajustadas as distribuições marginais a cada uma das séries temporais, usaremos funções de cópulas para modelar a estrutura de dependência condicional entre elas. Pelo método IFM (Joe [58]), o ajuste da cópula é feito na série bivariada das pseudo-observações (transformação integral de probabilidade) obtidas a partir da modelagem das marginais. Uma das vantagens das funções cópulas é a capacidade para ajuste da dependência nos extremos. Existem várias maneiras de se definir uma cópula. Por uma questão de simplicidade, faremos as definições para o caso bivariado. Sejam $I = [0, 1]$ e $I^2 = I \times I$, uma cópula bidimensional é uma função C em I^2 , assumindo valores em I , que possui as seguintes propriedades:

1. Para todo $u_1, u_2 \in I$:

$$C(u_1, 0) = 0 = C(0, u_2)$$

$$C(u_1, 1) = 1 = C(1, u_2)$$

2. Para todo $u_1, u_2, v_1, v_2 \in I$ tal que $u_1 < u_2$ e $v_1 < v_2$:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Um resultado importante, sempre referenciado e base para aplicações, é o Teorema de Sklar (Sklar [76]), que mostra como uma função cópula faz a ligação entre distribuições marginais e a distribuição conjunta.

Teorema 3.1 (Sklar) *Seja F uma função de distribuição conjunta em \mathbb{R}^2 com marginais F_i , $i = 1, 2$. Então, existe uma cópula C_F tal que para todo $y = (y_1, y_2) \in \mathbb{R}^2$:*

$$F(y_1, y_2) = C_F(F_1(y_1), F_2(y_2)). \quad (3-18)$$

Quando F'_i s são contínuas, a cópula C_F é única. Reciprocamente, se C_F é uma cópula e F'_i s são funções de distribuição, então a função F definida acima é a distribuição conjunta com marginais F_i . Dessa forma, uma cópula é uma função que agrega funções de distribuição univariadas para construir uma função de distribuição conjunta.

Em muitos casos, temos bom conhecimento sobre as distribuições marginais das variáveis individuais, mas sabe-se pouco sobre o comportamento

conjunto. Nesse sentido, a cópula permite construir distribuições conjuntas quando apenas as marginais são conhecidas com mais certeza. Uma grande vantagem das funções cópulas é que as distribuições marginais das variáveis individuais podem ser especificadas de diferentes famílias.

O resultado do Teorema de Sklar (Sklar [76]) pode ser aplicado também para os casos de marginais discretas ou mistas contínuas e discretas. Se F_i' s não forem todas contínuas, a distribuição conjunta sempre pode ser expressa como na equação 3-18, apesar de que, neste caso, a cópula não será única. Esse assunto é abordado em Schweizer & Sklar [77]. Conforme Trivedi & Zimmer [78], a falta de unicidade de uma representação de cópula para distribuições discretas é uma questão teórica que precisa ser confrontada em provas analíticas mas que não impedem aplicações empíricas. Encontrar uma cópula única requereria conhecer a forma da distribuição conjunta. Entretanto, na maioria das vezes, utiliza-se cópulas porque não se conhece a verdadeira distribuição conjunta, portanto, seja com dados discretos ou contínuos, um problema chave na modelagem é escolher uma cópula que capture adequadamente as estruturas de dependências sem sacrificar propriedade das marginais.

Genest & Nešlehová [79] fazem uma revisão sobre o uso de cópulas estáticas com marginais discretas e ressaltam que, apesar da falta de garantia de unicidade, modelos de cópulas para distribuições marginais discretas são construções válidas e muitas das propriedades de dependência estocástica permanecem. Os resultados sugerem que o problema da falta de unicidade é maior quando y_1 ou y_2 assumem um número pequeno de valores, o que não é o caso na nossa abordagem com número de sinistros. Mesmo para o caso de marginais discretas, o parâmetro θ da cópula ainda pode ser interpretado como um parâmetro de dependência.

A equação 3-18 pode ser invertida da seguinte maneira. Sejam F uma função de distribuição conjunta de um vetor aleatório $y = (y_1, y_2)$ e F_1, F_2 funções de distribuição marginais, defina $F_i(y_i) = u_i, i = 1, 2$, como a transformação integral de probabilidade (PIT - sigla em inglês para *probability integral transformation*) para a distribuição Uniforme(0,1). Então, a cópula C_F correspondente a F é denotada por:

$$C(u_1, u_2) = F \left(F_1^{-1}(u_1), F_2^{-1}(u_2) \right).$$

Para todo $(u_1, u_2) \in I^2$, onde F_i^{-1} é a função quantil $F_i, i = 1, 2$, ou seja, $F_i^{-1}(p) = \inf\{y | F_i(y) > p\}, p \in I$. Então, a cópula é uma função de distribuição multivariada com marginais Uniforme(0,1).

Patton [80] efetuou uma extensão da teoria (incondicional) de cópulas

para o caso condicional, permitindo que a teoria de cópula seja também utilizada na análise de séries temporais multivariadas, tanto com dependência condicional estática como dinâmica. Dessa forma, a versão do Teorema de Sklar para o caso condicional para aplicação em séries temporais é dada por

$$F(\mathbf{y}_t|\mathcal{F}_{t-1}) = C(F_1(y_{1t}|\mathcal{F}_{t-1}), F_2(y_{2t}|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}), \quad (3-19)$$

sendo $\mathbf{y}_t = (y_{1t}, y_{2t})$ e \mathcal{F}_{t-1} o conjunto de informações disponíveis até $t - 1$.

Com base na formulação da cópula condicional dinâmica de Patton [80], Koopman et al. [81] propuseram um modelo variante no tempo para cópulas discretas em que os parâmetros tanto dos modelos marginais como da cópula variam no tempo por meio de um esquema de atualização autorregressivo guiado por observações baseado no *score* da distribuição de probabilidade condicional com respeito aos parâmetros variantes no tempo. Para contornar a questão da não unicidade da cópula no caso discreto, Koopman et al. [81] modelaram a cópula de uma forma parcimoniosa, usando uma função de cópula paramétrica definida em todo $[0, 1]^d$, ainda que a unicidade só seja garantida para um determinado conjunto de pontos discretos. A cópula só é determinada de forma única no intervalo $RanF_1 \times \dots \times RanF_d$, onde $RanF_i$ denota a imagem da função de distribuição acumulada F_i , $i = 1, \dots, d$.

Conforme apresentado por Koopman et al. [81], no caso bivariado, a estrutura de dependência é caracterizada pela função de cópula condicional bivariada

$$C(F_1(y_{1t}|\mathcal{F}_{t-1}; \theta_{1t}^m), F_2(y_{2t}|\mathcal{F}_{t-1}; \theta_{2t}^m)|\mathcal{F}_{t-1}; \theta_t^c), \quad (3-20)$$

onde $F_i(y_{it}|\mathcal{F}_{t-1}; \theta_{it}^m)$, $i = 1, 2$, são as funções de distribuições marginais condicionais, θ_{it}^m é o vetor de parâmetros variante no tempo da i -ésima distribuição marginal e θ_t^c é o vetor de parâmetros que define a função cópula C . O fato de θ_t^c ser variante no tempo permite modelar os casos em que a estrutura de dependência muda ao longo do tempo. Ressalta-se que a função 3-20 também pode ser utilizada para modelagem de dependência estática, fazendo o parâmetro θ^c da cópula não variante no tempo.

Em nossa abordagem, ajustamos diferentes funções de cópulas, tanto da família das cópulas elípticas (Gaussiana e t -Student) como das cópulas Arquimedianas (Gumbel, Clayton, Frank e Joe). As cópulas Arquimedianas podem assumir dependência caudal assimétrica. Podemos dizer que existe dependência de cauda inferior (ou superior) quando a probabilidade de valores conjuntos mais baixos (ou mais elevados) é maior que a que poderia ser prevista a partir das distribuições marginais. As funções de cópulas Gumbel e Joe apresentam dependência de cauda superior, enquanto a cópula Clayton apresenta depen-

dência de cauda inferior. Já a cópula Frank apresenta dependência simétrica. É possível, ainda, efetuar rotações nas cópulas de forma a modelarem dependência negativa ou de outro extremo de cauda. Nelsen [11] apresenta uma lista extensa com diferentes famílias de cópulas Arquimedianas. As formulações das cópulas bivariadas estáticas utilizadas nessa tese estão apresentadas na seção B.2 do Apêndice B.

Para modelar a dependência de forma dinâmica entre as séries trabalhadas neste capítulo, considerando a formulação 3-20, ajustamos o modelo t -GAS com distribuição multivariada t -Student e matriz de correlação Σ_t variante no tempo segundo a dinâmica de atualização GAS, proposto por Creal et al. [82]. Formulações da distribuição multivariada t -Student e do esquema de atualização GAS correspondente estão apresentadas na seção B.3 do Apêndice B.

O modelo t -GAS é um modelo orientado por observações, conforme descrito na seção 3.2.1.2, que permite que parâmetros, como a matriz de covariância Σ_t varie no tempo utilizando a informação do *score* da densidade da distribuição multivariada t -Student. Para introdução do esquema de atualização GAS na distribuição multivariada t -Student, especificamos a matriz Σ_t como uma função $\Sigma(f_t)$ do vetor variante no tempo f_t .

A matriz Σ_t pode ser decomposta como

$$\Sigma_t = D_t R_t D_t, \quad (3-21)$$

na qual D_t é a matriz diagonal de desvios padrões e R_t é a matriz de correlação (simétrica). Em relação à matriz de correlação R_t , é necessário adotar uma decomposição que assegure que R_t seja positiva definida e simétrica com elementos fora da diagonal pertencentes ao intervalo $(-1, 1)$.

Creal et al. [82] apresentam algumas alternativas para essa decomposição. Nesse trabalho, foi adotada a decomposição baseada em coordenadas hiperesféricas apresentada no referido artigo. Para o caso bivariado, essa decomposição é dada por $R_t = X_t(\phi_t)' X_t(\phi_t)$, sendo que

$$X_t(\phi_t) = \begin{bmatrix} 1 & \cos \phi_t \\ 0 & \sin \phi_t \end{bmatrix},$$

e, portanto,

$$R_t = \begin{bmatrix} 1 & \cos \phi_t \\ \cos \phi_t & 1 \end{bmatrix},$$

sendo $\rho_t = \cos \phi_t$. Para esta decomposição, a função de ligação entre f_t e Σ_t é dada por $f_t' = [\text{diag}(D_t^2)', \phi_t]$.

Nosso objetivo é fazer com que a correlação seja variante no tempo. Dessa forma, conforme descrito em Creal et al. [82], assume-se variância constante e, portanto, $D_t = I$. A estrutura de variação temporal estocástica de Σ_t é dada apenas pela dinâmica de ρ_t . Nesse caso, temos que $\Sigma_t = R_t$, com $\text{vech}(\Sigma_t) = (1, \rho_t, 1)'$, e $f_t = \phi_t = \arccos \rho_t$. O esquema de atualização de f_t é dado pela equação 3-8.

Para o uso do modelo t -GAS bivariado com especificação de cópulas, a densidade t -Student multivariada (equação B-1 do Apêndice B) opera no vetor $\tilde{y}_t' = [P_\nu^{-1}(u_{1t}), P_\nu^{-1}(u_{2t})]'$, de dimensão 2×1 , sendo $P_\nu(\cdot)$ a função de distribuição univariada t -Student com ν graus de liberdade. Referenciamos o artigo de Creal et al. [82] para maiores detalhes teóricos sobre o modelo t -GAS.

3.2.2.2

Avaliação de dependência e do ajuste do modelo de cópulas

Conforme descrito em Mendes et al. [83], para medir dependência monotônica, pode-se usar a versão populacional da medida de associação conhecida como tau de Kendall (τ). O τ de Kendall não depende das distribuições marginais e é dado em termos da cópula por:

$$\tau(y_1, y_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.$$

Uma expressão similar existe para a versão populacional do ρ de Spearman, mas não para o coeficiente de correlação de Pearson.

Segundo Genest & Favre [84], ρ de Spearman e τ de Kendall são as estatísticas mais frequentemente utilizadas para medir e testar dependência. Essas medidas são úteis também para indicar o sinal da dependência, o que é importante para a escolha das possíveis funções de cópulas a serem usadas. A partir dessas medidas é possível efetuar um teste de hipótese para avaliar independência nos dados. Formulações das versões empíricas de ρ de Spearman e τ de Kendall, bem como das estatísticas do teste de independência podem ser consultadas em Genest & Favre [84].

Uma ferramenta gráfica existente para avaliação de dependência é o gráfico Kendall-plot ou K-Plot, uma adaptação do conceito de gráficos de probabilidade (como o qq-plot). Proposto por Genest & Boies [85], o K-Plot compara os quantis amostrais aos quantis teóricos sob hipótese nula de uma específica função distribuição K de uma variável aleatória associada a uma cópula C . Assim como a falta de linearidade é um sinal de não normalidade em um qq-plot, o nível de curvatura no K-plot é característico do grau de associação nos dados. Quando os quantis teóricos sob hipótese nula referem-

se à cópula de independência, o gráfico proposto mostra se há evidência de dependência entre as variáveis.

Após a seleção do modelo de cópula mais aderente aos dados, é necessário fazer avaliação da qualidade deste ajuste. Medidas de qualidade do ajuste para cópulas podem ser obtidas pelo cálculo de alguma distância entre a cópula empírica \tilde{C} e a paramétrica $\hat{C} = C_{\hat{\theta}}$ ajustada aos dados. Cópulas empíricas foram introduzidas por Deheuvels [86], e suas propriedades limites foram estudadas em Deheuvels [87, 88]. De acordo com Deheuvels [86], \tilde{C} converge para C_{θ} com aumento de n . A função de cópula empírica bivariada é definida por:

$$\tilde{C}(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(U_{i1} \leq u_1, U_{i2} \leq u_2), \quad (u_1, u_2) \in [0, 1]^2,$$

na qual \mathbf{I}_A é a função indicadora do evento A .

Para avaliação do ajuste do modelo de cópula não variante no tempo, dois testes comumente utilizados são Kolmogorov–Smirnov (KS) e Cramér–von Mises (CvM), baseados na comparação da cópula ajustada com a cópula empírica (ver Rémillard [89] para uma abordagem de testes de qualidade de ajuste de cópulas para séries temporais multivariadas). Para o caso de cópula variante no tempo, entretanto, os testes devem ser baseados na transformação de Rosenblatt, proposta por Rosenblatt [90]. Genest et al. [91] apresentam uma revisão dos diferentes testes de avaliação de ajuste para modelos de cópulas na literatura e concluem que o teste de Crámer-von Mises seria o de maior poder, tendo sido, portanto, escolhido nesse trabalho. O teste de Crámer-von Mises baseado na cópula emprírica testa a hipótese nula de que a estrutura de dependência está bem representada por uma família paramétrica específica de cópulas C_{θ} e apresenta a seguinte estatística de teste:

$$S_n = \int_{[0,1]^2} \mathcal{C}_n(u_1, u_2) d\tilde{C}(u_1, u_2), \quad (u_1, u_2) \in [0, 1]^2,$$

na qual $\mathcal{C}_n(u_1, u_2) = \sqrt{n}(\tilde{C} - C_{\theta})$.

3.3

Aplicação

3.3.1

Dados

Em nossa abordagem, para modelar uma série de tempo bivariada, usamos diferentes famílias para fatorar a distribuição multivariada em suas distribuições marginais univariadas.

Foram utilizados dados mensais de seguros de fiança locatícia e seguro de crédito interno de seguradoras brasileiras de seguros financeiros, no período de janeiro/2006 a dezembro/2015, totalizando 120 observações para cada série. Os gráficos das séries de números de sinistros para cada tipo de seguro, bem como das séries de taxa de ocorrência de sinistros³, estão nas Figuras 3.2 e 3.3.

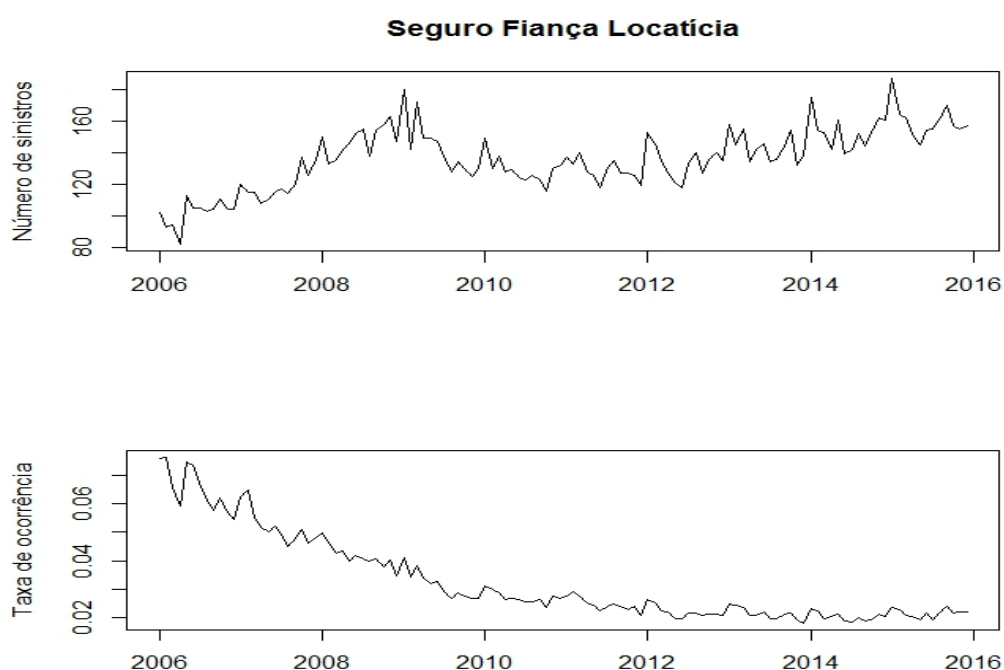


Figura 3.2: Séries temporais de número de sinistros e taxa de ocorrência do seguro de fiança locatícia no período de 01/2006 a 12/2015.

O seguro de fiança locatícia é uma modalidade que garante o pagamento de indenização, ao segurado/proprietário, dos prejuízos que venha a sofrer em decorrência do inadimplemento do locatário em relação à locação do imóvel urbano mencionado na apólice. É, portanto, uma forma de seguro em que, por uma taxa anual relativamente pequena (prêmio), os proprietários podem se proteger contra a perda de renda do aluguel. Normalmente, o seguro de

³Taxa de ocorrência = número de sinistros / exposição.

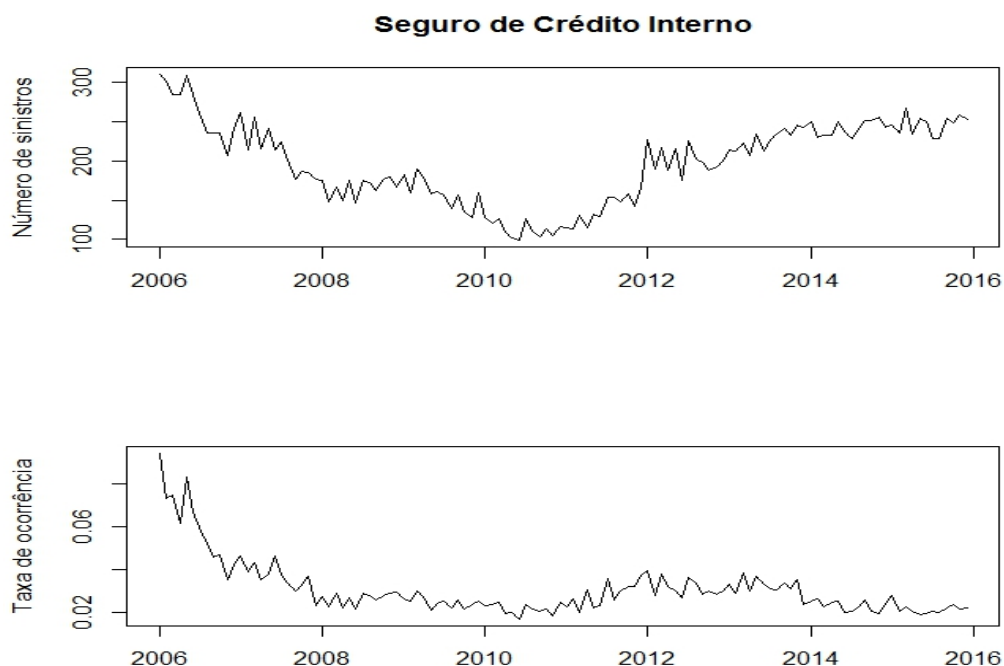


Figura 3.3: Séries temporais de número de sinistros e taxa de ocorrência do seguro de crédito interno no período de 01/2006 a 12/2015.

fiança locatícia é combinado com uma cobertura de assistência legal em que os custos judiciais para recuperação do aluguel e/ou despejar um inquilino inadimplente são cobertos. Esse é um tipo muito útil de seguro, especialmente quando os proprietários deixam sua propriedade sem um agente imobiliário, uma vez que os controles necessários envolvidos em geral tendem a dissuadir os inquilinos desonestos que, de outra forma, buscariam propriedades sendo alugadas privativamente.

Já o seguro de crédito interno objetiva ressarcir o segurado (credor), nas operações de crédito realizadas com clientes domiciliados no país, das perdas líquidas definitivas causadas por devedor insolvente. Perda líquida definitiva corresponde ao total do crédito sinistrado acrescido das despesas de sua recuperação e deduzido das quantias efetivamente recebidas, relativas a esse crédito. Este seguro é geralmente contratado por empresas que realizam operações de crédito em suas vendas, tanto para pessoa física como para pessoa jurídica, ou intermediários de operações de crédito, financiamento e investimento; consórcios, empresas de factoring, etc. caracterizados desta forma como segurados das operações de crédito. Os segurados também são os responsáveis pelo pagamento do prêmio de seguro. Os contratantes da operação de crédito, ou seja, os devedores são denominados garantidos, e é sobre eles que incide o risco de inadimplência. Nesse tipo de seguro, é caracterizada

a insolvência, ou *default*, quando for declarada judicialmente a falência do devedor, também denominado garantido; quando for deferido judicialmente o processamento da concordata preventiva do garantido; ou quando for concluído um acordo particular do garantido com a totalidade dos seus credores, com a interveniência da seguradora, para pagamento de todas as dívidas com redução dos débitos. Na modalidade “Quebra de Garantia”, a insolvência também se caracteriza quando, na cobrança judicial ou extrajudicial da dívida, os bens dados em garantia ou os bens do garantido revelem-se insuficientes ou fique evidenciada a impossibilidade de busca e apreensão, reintegração, arresto ou penhora desses bens.

Como variáveis exógenas, obtivemos dados mensais de diversas variáveis macroeconômicas disponibilizadas pelo Banco Central do Brasil por meio da base de dados Quandl ⁴: taxas de juros Selic (acumulada no mês em termos anuais, em %), taxa de inflação IPCA, PIB mensal⁵ (em R\$ milhões) e taxa de câmbio de dólar (PTAX).

Foram escolhidas variáveis macroeconômicas que representassem condições gerais macroeconômicas, direcionadores da economia e do mercado financeiro. Todas as séries foram trabalhadas com sua variação mensal. No caso do PIB, foi apurada a taxa de crescimento real mensal. Para taxa de juros, foi escolhida a Selic por ser a taxa básica de juros da economia no Brasil, utilizada no mercado interbancário para financiamento de operações com duração diária, lastreadas em títulos públicos federais. A taxa SELIC é considerada individualmente o principal instrumento de política monetária no sistema de metas de inflação. Como índice de inflação, selecionamos o IPCA, por ser o índice adotado pelo Banco Central do Brasil (Bacen) para a fixação das metas de inflação do país, acordadas entre o governo brasileiro e o Fundo Monetário Internacional (FMI). Em relação à taxa de câmbio, adotamos a taxa PTAX que corresponde à taxa média praticada no mercado interbancário, coletada e divulgada pelo Bacen.

Na Tabela 3.1, são apresentadas estatísticas descritivas das séries macroeconômicas utilizadas no trabalho e na Tabela 3.2 são apresentadas as medidas de correlação entre essas séries.

⁴Disponível em URL: <https://www.quandl.com/>. Séries BCB/4189, BCB/433, BCB/4380 e BCB/1.

⁵O PIB mensal é um indicador mensal produzido pelo Banco Central do Brasil para utilização no cálculo da relação entre agregados econômicos mensais (como dívida pública, saldo em transações correntes e saldo de crédito) e o PIB. Uma alternativa indicada para caracterização do ritmo da economia brasileira ao longo dos meses seria a adoção da série do índice IBC-Br (Índice de Atividade Econômica do Banco Central) com ajuste sazonal.

Tabela 3.1: Estatísticas descritivas das variáveis exógenas no período de 01/2006 a 12/2015.

Série	Mínimo	Q 5%	Mediana	Média	Q 95%	Máximo
Selic	-0.086	-0.054	0.000	-0.002	0.042	0.078
IPCA	-0.002	0.000	0.005	0.005	0.009	0.013
PIB	-0.083	-0.055	0.001	0.004	0.075	0.086
Câmbio	-0.094	-0.059	-0.003	0.005	0.095	0.171
	DP	Assimetria	Curtose	r(1)	r(2)	r(12)
Selic	0.030	0.163	2.951	0.816	0.722	-0.270
IPCA	0.003	0.454	3.645	0.680	0.404	0.352
PIB	0.038	0.097	2.582	-0.194	-0.164	0.673
Câmbio	0.048	0.923	4.270	0.070	0.196	-0.077

Tabela 3.2: Correlação entre as variáveis exógenas.

	Selic	IPCA	PIB	Câmbio
Selic	1.000	0.233	-0.006	0.164
IPCA	0.233	1.000	-0.124	0.068
PIB	-0.006	-0.124	1.000	0.019
Câmbio	0.164	0.068	0.019	1.000

3.3.2 Resultados

Os modelos propostos foram implementados utilizando a linguagem R (R Core Team [47]). Assim como apontado em Figlewski et al. [52], o efeito das variáveis macroeconômicas no número de sinistros de seguros financeiros provavelmente apresenta alguma defasagem. Dessa forma, propusemos algumas variações das séries originais para avaliar as séries que melhor se adequariam aos modelos para números de sinistros, incluindo séries com médias de alguns meses. As séries de variáveis macroeconômicas foram denominadas: “sem defasagem”, “com 1 mês de defasagem”, “com 3 meses de defasagem”, “média de 3 meses” (média dos últimos três meses), “média de 6 meses” (média dos últimos seis meses), “média de 3 meses def. em 1 mês” (média dos três meses anteriores, considerando um mês de defasagem), “média de 6 meses def. em 1 mês” (média dos seis meses anteriores, considerando um mês de defasagem), “média de 3 meses def. em 3 meses” (média dos três meses anteriores, considerando três meses de defasagem) e “média de 6 meses def. em 3 meses” (média dos seis meses anteriores, considerando três meses de defasagem).

Utilizando as primeiras 108 observações mensais, de forma a deixar 12 observações para avaliação fora da amostra, foram ajustados diferentes modelos

às séries de número de sinistros de seguro fiança locatícia e crédito interno, considerando também as variações de defasagens das séries macroeconômicas já citadas. A variável exposição (r_t) foi adotada como variável *offset* em todos os modelos. Para os modelos GAS Poisson e GAS Binomial Negativa, apresentados na seção 3.2.1.2, utilizamos $p = q = 1$ na dinâmica de atualização de f_t (equação 3-8) e matriz de ponderação igual à inversa da informação de Fisher ($S_t = I_{t|t-1}^{-1}$). Modelos também foram ajustados com matriz de ponderação igual à raiz quadrada da inversa da informação de Fisher ($S_t = I_{t|t-1}^{-1/2}$). Entretanto, como a otimização por máxima verossimilhança se mostrou mais estável com a primeira opção, adotamos $S_t = I_{t|t-1}^{-1}$.

3.3.2.1

Modelagem da série de seguro de fiança locatícia

Para o seguro de fiança locatícia, foram ajustados modelos GAS (seção 3.2.1.2), GLM autorregressivo simples (denominado “GLM reg.” - formulação 3-4) e modelos GLM autorregressivos propostos por Fokianos & Tjøstheim [72] (denominado “TSGLM” - formulação (3-3)), tanto com distribuição Poisson (3-5) como com distribuição binomial negativa (3-6). No caso dos modelos GAS e TSGLM, os resultados apresentados são para modelos que adotam $p = q = 1$ na equação de atualização do parâmetro variante no tempo (equações (3-8) e (3-3), respectivamente). Os resultados das medidas de AIC (Critério de Informação de Akaike) para todos os modelos ajustados estão apresentados nas Tabelas 3.3 e 3.4.

Tabela 3.3: AIC dos modelos com distribuição Poisson ajustados à série de sinistros de fiança locatícia.

Séries exógenas	GAS(1,1)	GLM reg.	TSGLM(1,1)
sem defasagem	814.12	2188.4	2000.95
com 1 mês de defasagem	809.69	2215.8	880.90
com 3 meses de defasagem	813.78	2240.2	893.77
média de 3 meses	799.85	2064.5	885.53
média de 6 meses	809.87	2014.1	891.94
média de 3 meses def. em 1 mês	812.63	2185.4	891.72
média de 6 meses def. em 1 mês	816.72	2093.3	894.32
média de 3 meses def. em 3 meses	819.92	2226.5	902.11
média de 6 meses def. em 3 meses	812.55	2134.6	918.10

Analisando os resultados apresentados nas Tabelas 3.3 e 3.4, com base nos valores de AIC, identificamos o modelo GAS Poisson(1,1) com série exógena “média de 3 meses” como potencialmente mais adequado para a série de fiança

Tabela 3.4: AIC dos modelos com distribuição binomial negativa ajustados à série de sinistros de fiança locatícia.

Séries exógenas	GAS(1,1)	GLM reg.	TSGLM(1,1)
sem defasagem	816.12	1133.8	868.94
com 1 mês de defasagem	811.69	1137.8	852.77
com 3 meses de defasagem	815.78	1138.5	850.33
média de 3 meses	801.85	1127.6	862.06
média de 6 meses	811.87	1122.8	863.78
média de 3 meses def. em 1 mês	814.63	1135.9	848.20
média de 6 meses def. em 1 mês	818.72	1129.8	852.96
média de 3 meses def. em 3 meses	821.92	1138.1	848.36
média de 6 meses def. em 3 meses	814.55	1132.4	859.39

locatícia e partimos dele para o processo de seleção de variáveis e avaliação de ajustes.

O modelo GAS Binomial Negativa com série exógena “média de 3 meses” apresentou valor de AIC próximo ($AIC = 801.85$), mas ainda superior, ao do modelo GAS Poisson com série exógena “média de 3 meses” ($AIC = 799.85$). Entretanto, a estimativa do parâmetro ϕ (dispersão) era da ordem de $1e+08$, fazendo com que o modelo convirja para um modelo Poisson e indicando que não haveria sobredispersão na série. Sendo assim, a distribuição binomial negativa não seria a mais adequada para a série de número de sinistros de fiança locatícia.

Seguindo o procedimento descrito por Kedem & Fokianos [62] e adotado por Figlewski et al. [52], utilizamos uma seleção *backward* partindo do modelo completo com as quatro variáveis exógenas e reduzindo gradativamente a partir da exclusão, a cada passo, da variável exógena cujo coeficiente se mostrou menos significativo. Como forma de avaliação e comparação do ajuste dos modelos, a cada passo, além da significância das estimativas, avaliamos também a medida AIC e o valor do logaritmo da função de verossimilhança avaliada com valores de $\hat{\theta}$. Esse procedimento de seleção foi repetido até que todos os coeficientes das variáveis macroeconômicas fossem significativos a um nível de, pelo menos, 10%. A Tabela 3.5 apresenta os resultados desse procedimento, que resultou na escolha do modelo GAS Poisson(1,1) “média de 3 meses” com variáveis IPCA e PIB, como o modelo de melhor ajuste aos dados de seguro de fiança locatícia.

A Tabela 3.6 apresenta as estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de fiança locatícia via modelos GAS Poisson(1,1) sem variáveis exógenas, com quatro variáveis exógenas e com apenas as variáveis IPCA e PIB. São apresentados também as medidas de aderência AIC e BIC, o valor do logaritmo da verossimilhança

Tabela 3.5: Medidas de aderência dentro da amostra dos modelos GAS Poisson(1,1), com série exógena “média de 3 meses” (média dos últimos três meses), para seguro de fiança locatícia no período de 01/2006 a 12/2014. A partir do modelo completo com as quatro variáveis exógenas (Selic, IPCA, PIB, Câmbio), excluímos, uma a uma, a variável cujo coeficiente se mostrou menos significativo.

Variáveis exógenas	<i>log-Veross.</i>	AIC
Selic, IPCA, PIB, Câmbio	-391.926	799.853
Selic, IPCA, PIB	-392.187	798.374
IPCA, PIB	-392.725	797.450
Sem exógenas	-403.747	815.494

e a medida de erro MAPE⁶. A partir da análise dessas informações, podemos concluir que a inclusão de variáveis exógenas no modelo trouxe ganhos em aderência dentro da amostra.

Realizamos o teste de razão de verossimilhança entre o modelo final escolhido (com IPCA e PIB) e o modelo reduzido sem variável exógena macroeconômica e o p-valor foi igual a 1.6e-05, indicando a rejeição da hipótese nula de que o modelo sem variável macroeconômica seria melhor. As medidas de ajuste (AIC e log-verossimilhança) também indicaram que a inclusão de variáveis exógenas trouxe melhorias para o modelo.

Para fins de diagnóstico e avaliação dos modelos, utilizamos os resíduos quantílicos, propostos por Dunn & Smyth [48] e Kalliovirta [49] para modelos de séries temporais não lineares. Dunn & Smyth [48] apresentam formulação para os resíduos quantílicos tanto para distribuições contínuas como para distribuições discretas.

Dunn & Smyth [48] definem os resíduos quantílicos teóricos como:

$$R_{t,\theta} = \Phi^{-1}(F_t(y_t|f_t, \mathcal{F}_{t-1}; \theta)),$$

e os resíduos quantílicos observados como $r_{t,\hat{\theta}} = \Phi^{-1}(F_t(y_t|f_t, \mathcal{F}_{t-1}; \hat{\theta}))$, onde $\Phi^{-1}(\cdot)$ é a inversa da função de distribuição acumulada da Normal padrão, $F_t(\cdot)$ é a função de distribuição acumulada condicional de y_t e $\hat{\theta}$ é um estimador do vetor de parâmetros θ . Conforme descrito por Kalliovirta [49], os resíduos quantílicos observados são assintoticamente independentes e normalmente distribuídos.

Para o caso de distribuições discretas, Dunn & Smyth [48] definem o resíduo quantílico randomizado observado como $r_{t,\hat{\theta}} = \Phi^{-1}(u_t)$, sendo u_t uma variável aleatória uniforme no intervalo $(a_t, b_t]$, com $a_t = \lim_{y \uparrow y_t} F_t(y|f_t, \mathcal{F}_{t-1}; \hat{\theta})$ e

⁶MAPE = $\frac{100\%}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_{t|t-1}}{y_t} \right|$

Tabela 3.6: Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de fiança locatícia via modelos GAS Poisson(1,1) sem variáveis exógenas, com quatro variáveis exógenas e com apenas as variáveis IPCA e PIB. São apresentados, também, as medidas de aderência AIC e BIC, o valor do logaritmo da verossimilhança e a medida de erro MAPE.

	Modelo sem exógenas	Modelo com todas as variáveis	Modelo com IPCA e PIB
w	-0.0777*** (0.0189)	-0.0711*** (0.0153)	-0.0702*** (0.01783)
A_1	0.2040* (0.1197)	0.1650* (0.0937)	0.2007** (0.0977)
B_1	0.9814*** (0.0054)	0.9834*** (0.0043)	0.9836*** (0.0050)
Selic		0.3229 (0.3877)	
IPCA		12.2805*** (5.2843)	11.146** (5.1322)
PIB		-1.4851*** (0.4685)	-1.5751*** (0.4573)
Câmbio		0.2311 (0.3179)	
$\log\text{-Veross.}$	-403.747	-391.926	-392.725
AIC	815.494	799.853	797.450
BIC	826.223	821.310	813.543
MAPE(%)	6.082	5.034	5.113

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

$$b_t = F_t(y_t|f_t, \mathcal{F}_{t-1}; \hat{\theta}).$$

Para avaliação dos resíduos quantílicos, realizamos testes de autocorrelação serial (teste Ljung-Box até o *lag* 30), heterocedasticidade condicional (teste Ljung-Box até o *lag* 30 nos resíduos quadrados) e normalidade (Jarque-Bera). Os resultados dos testes para os modelos completo (com as quatro variáveis exógenas), modelo selecionado (com IPCA e PIB) e modelo sem variável exógena estão apresentados na Tabela 3.7. Nota-se a presença de autocorrelação serial e heterocedasticidade para o modelo sem variável exógenas, além de um possível desvio da hipótese de normalidade para o modelo com variáveis exógenas, que será abordado a seguir.

Avaliando os resíduos do modelo selecionado GAS Poisson(1,1) com variáveis IPCA e PIB (ver Tabela 3.7), observamos a existência de um resíduo quantílico com valor absoluto maior que 2.5 na observação '5' referente a maio/2006. Essa observação correspondeu a um aumento substancial no número de sinistros. Dessa forma, foi inserida uma variável de intervenção (*dummy*) no modelo, tornando melhor o ajuste aos dados.

Tabela 3.7: P-valores dos testes aplicados aos resíduos quantílicos dos modelos GAS Poisson(1,1) com série exógena “média de 3 meses” (média dos últimos três meses) para o seguro de fiança locatícia no período 01/2006 a 12/2014.

Modelo	Ljung-Box	Ljung-Box res ²	Jarque-Bera
Completo	0.685	0.632	0.279
Com IPCA e PIB	0.809	0.863	0.108
Sem exógenas	1.5e-05	0.002	0.188

Os resultados das medidas aderência e dos testes nos resíduos quantílicos do novo modelo selecionado com *dummy*, a partir de agora denominado apenas de “GAS Poisson(1,1)”, estão apresentados na Tabela 3.8. Os resultados dos testes nos resíduos quantílicos foram satisfatórios para o modelo final escolhido para seguro de fiança locatícia, sem indicação de autocorrelação serial, heterocedasticidade ou desvio da hipótese de normalidade. Os gráficos de diagnósticos dos resíduos quantílicos apresentados na Figura B.1 (Apêndice B) também confirmam o ajuste adequado do referido modelo.

Tabela 3.8: Medidas de aderência e p-valores dos testes aplicados aos resíduos quantílicos do modelo final selecionado GAS Poisson(1,1) com variáveis exógenas IPCA e PIB para o seguro de fiança locatícia no período 01/2006 a 12/2014.

	<i>log-Veross.</i>	AIC	BIC	MAPE(%)
GAS Poisson(1,1)	-390.802	795.605	814.379	4.965
	Ljung-Box	Ljung-Box res ²	Jarque-Bera	
	0.653	0.971	0.310	

Na Tabela 3.9, apresentamos os valores das estimativas dos parâmetros do modelo final selecionado para a série de número de sinistros de seguro de fiança locatícia, com os respectivos erros padrões. Nota-se que todos os coeficientes são estatisticamente significantes. A Figura 3.4 apresenta os valores observados da série de número de sinistros de seguro de fiança locatícia e os valores ajustados pelo modelo final selecionado GAS Poisson(1,1), com série exógena “média de 3 meses” e uma variável *dummy* (na observação de maio/2006), enquanto a Figura 3.5 apresenta comparação entre a taxa de ocorrência observada e a taxa de ocorrência considerando os valores ajustados pelo citado modelo.

Avaliando os coeficientes estimados para as variáveis exógenas, perceberemos efeitos em linha com o esperado, ou seja, existência de uma relação negativa entre a inadimplência de operações de crédito e a fase do ciclo econômico. Variações positivas na inflação (IPCA) acarretariam aumento do número

Tabela 3.9: Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de fiança locatícia via modelo GAS Poisson(1,1) com variáveis exógenas IPCA e PIB.

Fiança Locatícia	
Parâmetros	GAS Poisson(1,1)
w	-0.0672*** (0.0199)
A_1	0.2265** (0.0966)
B_1	0.9844*** (0.0056)
$D_{mai/06}$	0.2003** (0.1003)
IPCA	10.7906** (5.1938)
PIB	-1.7692*** (0.4645)

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

/ taxa de ocorrência de sinistros de fiança locatícia. Enquanto que aumento do PIB acarretaria diminuição do número / taxa de ocorrência de sinistros.

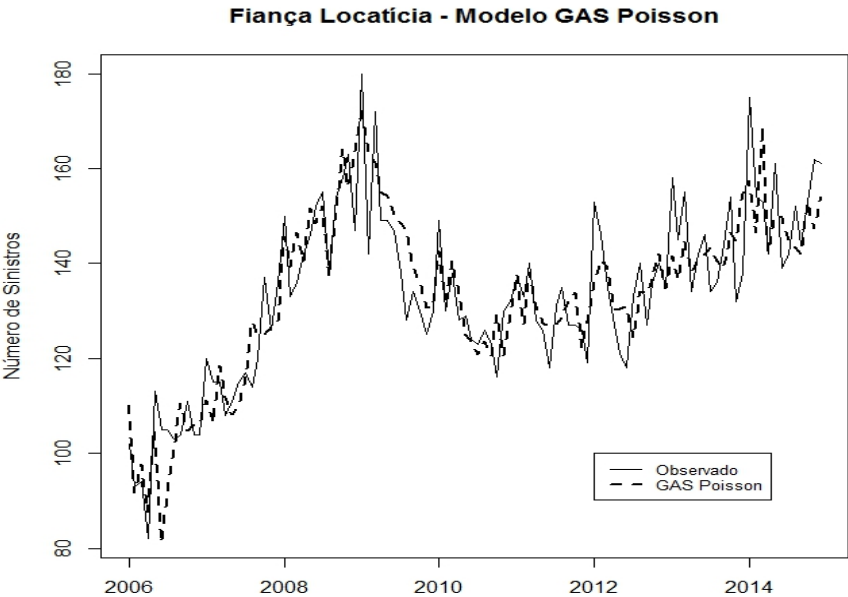


Figura 3.4: Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Poisson(1,1) para o seguro de fiança locatícia no período de 01/2006 a 12/2014, com variáveis exógenas IPCA e PIB.

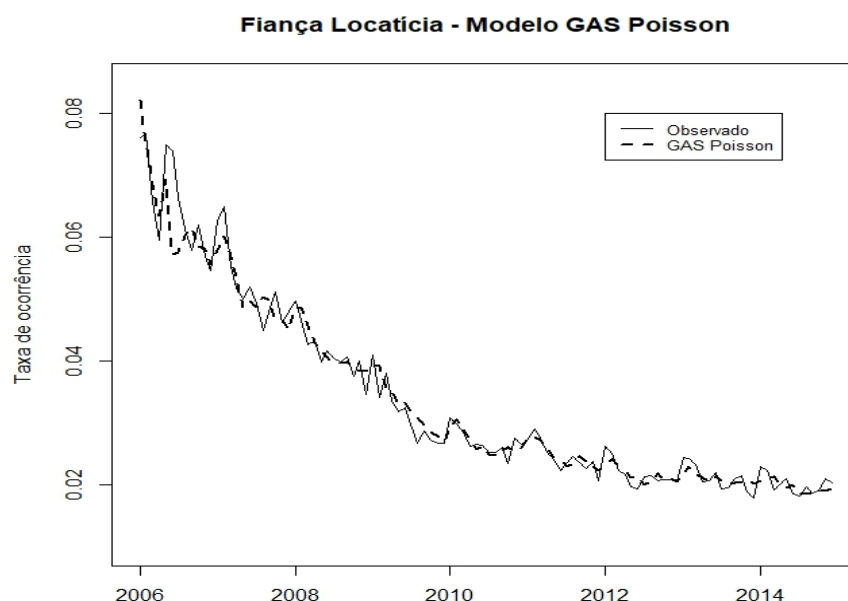


Figura 3.5: Taxa de ocorrência de sinistro observada, em linha sólida, versus taxa de ocorrência estimada, em linha pontilhada, pelo modelo GAS Poisson(1,1) para o seguro de fiança locatícia no período de 01/2006 a 12/2014, com variáveis exógenas IPCA e PIB.

3.3.2.2

Modelagem da série de seguro de crédito interno

Assim como no caso do seguro de fiança locatícia, para o seguro de crédito interno, foram ajustados modelos GAS (seção 3.2.1.2), GLM autorregressivo simples (denominado “GLM reg.” - formulação 3-4) e modelos GLM autorregressivos propostos por Fokianos & Tjøstheim [72] (denominado “TSGLM” - formulação (3-3)), tanto com distribuição Poisson (3-5) como com distribuição binomial negativa (3-6). No caso dos modelos GAS e TSGLM, os resultados apresentados são para modelos que adotam $p = q = 1$ na equação de atualização do parâmetro variante no tempo (equações (3-8) e (3-3), respectivamente). Os resultados das medidas de AIC para todos os modelos ajustados estão apresentados nas Tabelas 3.10 e 3.11.

Analizando os resultados apresentados nas Tabelas 3.10 e 3.11, com base nos valores de AIC, identificamos o modelo GAS Binomial Negativa(1,1) com série exógena “média de 6 meses” como potencialmente mais adequado para a série de número de sinistros de seguro de crédito interno e partimos dele para o processo de seleção de variáveis e avaliação de ajustes.

Os modelos GLM reg e TSGLM se mostraram instáveis quando adotada a distribuição Poisson. Esse comportamento pode ser explicado pela confirmação

Tabela 3.10: AIC dos modelos com distribuição Poisson ajustados à série de sinistros de crédito interno.

Séries exógenas	GAS(1,1)	GLM reg.	TSGLM(1,1)
sem defasagem	1147.78	2217.6	4260.49
com 1 mês de defasagem	1154.55	2222.4	4211.89
com 3 meses de defasagem	1161.74	2298.0	4470.76
média de 3 meses	1151.76	2124.9	3936.18
média de 6 meses	1139.01	1990.7	3525.62
média de 3 meses def. em 1 mês	1162.90	2196.5	4141.69
média de 6 meses def. em 1 mês	1161.04	2079.3	3908.94
média de 3 meses def. em 3 meses	1155.33	2241.0	4361.26
média de 6 meses def. em 3 meses	1159.35	2103.8	4213.89

Tabela 3.11: AIC dos modelos com distribuição binomial negativa ajustados à série de sinistros de crédito interno

Séries exógenas	GAS(1,1)	GLM reg.	TSGLM(1,1)
sem defasagem	1020.36	1145.2	1073.60
com 1 mês de defasagem	1021.95	1146.9	1076.57
com 3 meses de defasagem	1023.66	1152.6	1078.39
média de 3 meses	1020.97	1141.7	1071.27
média de 6 meses	1017.01	1132.4	1060.19
média de 3 meses def. em 1 mês	1024.49	1146.6	1075.34
média de 6 meses def. em 1 mês	1023.72	1139.7	1069.13
média de 3 meses def. em 3 meses	1022.78	1150.8	1071.82
média de 6 meses def. em 3 meses	1023.69	1143.3	1067.69

de que os dados apresentavam sobre-dispersão e, por esta razão, a distribuição binomial negativa seria mais adequada para a modelagem. Os mesmos modelos com distribuição binomial negativa ficaram estáveis, porém com desempenho inferior ao modelo GAS Binomial Negativa, conforme pode ser verificado nos resultados apresentados na Tabela 3.11.

Adotamos o mesmo procedimento realizado com a série de fiança locatícia para seleção do modelo com variáveis exógenas que se mostrassem estatisticamente significante. Partindo do modelo completo com as quatro variáveis exógenas, reduzimos gradativamente a partir da exclusão, a cada passo, da variável exógena cujo coeficiente se mostrou menos significante. Como forma de avaliação e comparação do ajuste dos modelos, a cada passo, além da significância das estimativas, avaliamos também a medida AIC e o valor do logaritmo da função de verossimilhança avaliada com valores de $\hat{\theta}$. Esse procedimento de seleção foi repetido até que todos os coeficientes das variáveis macroeconômicas fossem significantes a um nível de, pelo menos, 10%. A Tabela 3.12 apresenta os resultados desse procedimento, que resultou na escolha do modelo GAS Binomial Negativa(1,1) “média de 6 meses” com variáveis PIB e câmbio, como o

modelo de melhor ajuste aos dados de seguro de crédito interno.

Tabela 3.12: Medidas de aderência dentro da amostra dos modelos GAS Binomial Negativa(1,1), com série exógena “média de 6 meses” (média dos últimos seis meses), para seguro de crédito interno no período de 01/2006 a 12/2014. A partir do modelo completo com as quatro variáveis exógenas (Selic, IPCA, PIB, Câmbio), excluímos, uma a uma, a variável cujo coeficiente se mostrou menos significante.

Variáveis exógenas	<i>log-Veross.</i>	AIC
Selic, IPCA, PIB, Câmbio	-499.507	1017.01
Selic, PIB, Câmbio	-499.708	1015.42
PIB, Câmbio	-499.958	1013.92
Sem exógenas	-503.654	1017.31

A Tabela 3.13 apresenta as estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de crédito interno via modelos GAS Binomial Negativa(1,1) sem variáveis exógenas, com quatro variáveis exógenas e com apenas as variáveis PIB e câmbio. São apresentados também as medidas de aderência AIC e BIC, o valor do logaritmo da verossimilhança e a medida de erro MAPE. A partir da análise dessas informações, podemos inferir que a inclusão de variáveis exógenas no modelo trouxe ganhos em aderência dentro da amostra

Realizamos o teste de razão de verossimilhança entre o modelo selecionado (com PIB e câmbio) e o modelo reduzido sem variável exógena macroeconômica e o p-valor foi igual a 0.0248, indicando a rejeição da hipótese nula de que o modelo sem variável macroeconômica seria melhor. As medidas de ajuste (AIC e log-verossimilhança) também indicaram que a inclusão de variáveis exógenas trouxe melhorias para o modelo.

Para fins de diagnóstico e avaliação dos modelos, assim como no caso do seguro de fiança locatícia, utilizamos os resíduos quantílicos e realizamos testes de autocorrelação serial (teste Ljung-Box até o *lag* 30), heterocedasticidade condicional (teste Ljung-Box até o *lag* 30 nos resíduos quadrados) e normalidade (Jarque-Bera). Os resultados dos testes para os modelos completo (com as quatro variáveis exógenas), modelo selecionado (com PIB e câmbio) e modelo sem variável exógena estão apresentados na Tabela 3.14 e não apontam desvios às hipóteses de não autocorrelação serial, homocedasticidade e normalidade dos resíduos.

Avaliando os resíduos do modelo selecionado GAS Binomial Negativa(1,1) com variáveis PIB e câmbio, verificamos a existência de algumas observações mal ajustadas pelo modelo selecionado, com resíduos quantílicos correspondentes em valor absoluto superior a 2.5, e introduzimos variáveis *dummies* para

Tabela 3.13: Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de crédito interno via modelos GAS Binomial Negativa(1,1) sem variáveis exógenas, com quatro variáveis exógenas e com apenas as variáveis PIB e câmbio. São apresentados, também, as medidas de aderência AIC e BIC, o valor do logaritmo da verossimilhança e a medida de erro MAPE.

	Modelo sem exógenas	Modelo com todas as variáveis	Modelo com PIB e Câmbio
w	-0.2663*** (0.0653)	-0.3538*** (0.0747)	-0.3430*** (0.0730)
A_1	0.3059*** (0.0602)	0.3003*** (0.0637)	0.3078*** (0.0629)
B_1	0.9279*** (0.0186)	0.9052*** (0.0207)	0.9070*** (0.0205)
ϕ	73.6972*** (14.0540)	82.1743*** (16.1668)	81.3720*** (15.9822)
Selic		-0.8708 (1.0736)	
IPCA		9.5302 (14.9769)	
PIB		2.9896** (1.4444)	2.6534* (1.3554)
Câmbio		2.1147 (0.9617)**	1.6818* (0.8599)
\log -Veross.	-503.654	-499.507	-499.958
AIC	1017.31	1017.01	1013.92
BIC	1030.72	1041.15	1032.69
MAPE(%)	11.067	10.641	10.782

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

Tabela 3.14: Testes aplicados nos resíduos quantílicos dos modelos GAS Binomial Negativa(1,1), série “média de 6 meses” (média dos últimos seis meses), para seguro de crédito interno no período de 01/2006 a 12/2014.

Modelo	Ljung-Box	Ljung-Box res ²	Jarque-Bera
Completo	0.821	0.824	0.408
Com PIB, Câmbio	0.837	0.839	0.387
Sem exógenas	0.855	0.681	0.587

acomodá-las. As observações foram identificadas em dois períodos com maior variação na série de exposição.

Os resultados das medidas de aderência e dos testes nos resíduos quantílicos do novo modelo selecionado com *dummies*, a partir de agora denominado apenas de “GAS Binomial Negativa(1,1)”, estão apresentados na Tabela 3.15. Os resultados dos testes nos resíduos quantílicos foram satisfatórios para o modelo final escolhido para seguro de crédito interno, sem indicação de auto-

correlação serial, heterocedasticidade ou desvio da hipótese de normalidade. Os gráficos de diagnósticos dos resíduos quantílicos apresentados na Figura B.2 (Apêndice B) também confirmam o ajuste adequado do referido modelo.

Tabela 3.15: Medidas de aderência e p-valores dos testes aplicados aos resíduos quantílicos do modelo final selecionado GAS Binomial Negativa(1,1) com variáveis exógenas PIB e câmbio para o seguro de crédito interno no período 01/2006 a 12/2014.

	<i>log-Veross.</i>	AIC	BIC	MAPE(%)
GAS BinNeg(1,1)	-494.879	1007.76	1031.89	10.438
	Ljung-Box	Ljung-Box res ²	Jarque-Bera	
	0.992	0.358	0.265	

Em relação ao ajuste dentro da amostra do modelo GAS Binomial Negativa(1,1) selecionado com as *dummies*, verificamos que apesar de não apresentar mais resíduos quantílicos em valores absolutos superiores a 2.5, há algumas distorções no gráfico de número de sinistros na Figura 3.6. Essas variações se justificam por períodos de maior variação na série de exposição. Entratanto, ainda assim, os resultados de previsão dentro da amostra e fora da amostra se mostram satisfatórios. O bom ajuste do modelo pode ser verificado também no gráfico de taxa de ocorrência na Figura 3.7. Avaliamos a introdução de novas variáveis *dummies* mas a pequena melhora nas medidas MAPE dentro da amostra e fora da amostra não justificariam as novas intervenções dado que o modelo selecionado apresentou bons resultados na análise de resíduos. Sendo assim, optamos por seguir com o modelo GAS Binomial Negativa(1,1), com série exógena “média de 6 meses” e variáveis *dummies* nas observações ‘24’ e ‘67’, referentes a dez/2007 e jul/2011.

Na Tabela 3.16, apresentamos os valores das estimativas dos parâmetros do modelo final selecionado para a série de número de sinistros de seguro de crédito interno com os respectivos erros padrões. Nota-se que todos os coeficientes são estatisticamente significantes. A Figura 3.6 apresenta os valores observados da série de número de sinistros de seguro de crédito interno e os valores ajustados pelo modelo final selecionado GAS Binomial Negativa(1,1), com série exógena “média de 6 meses”, e *dummies* nas observações de dez/2007 e jul/2011, enquanto a Figura 3.7 apresenta como comparação entre a taxa de ocorrência observada e a taxa de ocorrência considerando os valores ajustados pelo citado modelo.

Avaliando os coeficientes estimados para as variáveis exógenas, percebemos que variações positivas na taxa de câmbio acarretariam aumento do número / taxa de ocorrência de sinistros de crédito interno, o que se justi-

Tabela 3.16: Estimativas de máxima verossimilhança e erros padrões para a série de número de sinistros de seguro de crédito interno via modelo GAS Binomial Negativa(1,1), no período de 01/2006 a 12/2014, com variáveis exógenas PIB e Câmbio, série “média de 6 meses” (média dos últimos seis meses).

Crédito Interno	
Parâmetros	GAS Binomial Negativa(1,1)
w	-0.3411*** (0.0735)
A_1	0.3351*** (0.0686)
B_1	0.9076*** (0.0207)
ϕ	92.617*** (18.838)
$D_{\text{dez}/07}$	-0.2565** (0.1171)
$D_{\text{jul}/11}$	0.3036** (0.1253)
PIB	2.4428* (1.3203)
Câmbio	1.8383** (0.8475)

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

fica pelo fato do seguro de crédito interno englobar operações de crédito com pessoas jurídicas, que podem ter posições e/ou negócios em dólar. Da mesma forma, um aumento do PIB acarretaria aumento do número / taxa de ocorrência de sinistros. Esse último efeito, apesar de não ser muito direto, pode ser explicado por uma expansão da oferta de crédito, talvez com menor controle e/ou qualidade do crédito. Intuitivamente, esperava-se que a inadimplência no crédito interno, ou seja, o número de sinistros do seguro de crédito interno, fosse negativamente afetada pela variação positiva da taxa de crescimento real do PIB, mas o efeito desta variação pode ser ambíguo, pois o aumento da oferta de crédito pode desafogar os tomadores de crédito (efeito negativo na inadimplência), mas também pode sinalizar aos tomadores de recursos leniência dos credores (efeito positivo na inadimplência). Jimenez & Saurina [92] relataram que um aumento rápido nas carteiras de empréstimos, por exemplo, decorrente de uma expansão da economia, estaria positivamente associado a um aumento nos índices de empréstimos inadimplentes mais tarde. Além disso, os empréstimos concedidos durante os períodos de crescimento teriam uma maior probabilidade de inadimplência do que aqueles concedidos durante períodos de crescimento lento do crédito. Destaca-se, ainda, que a série exógena

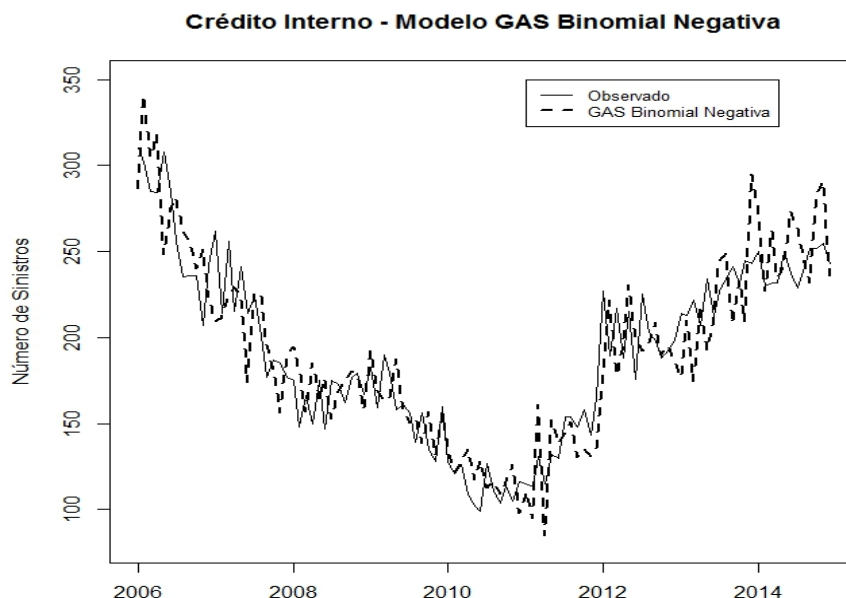


Figura 3.6: Número de sinistros observado, em linha sólida, versus número de sinistros estimado, em linha pontilhada, pelo modelo GAS Binomial Negativa(1,1) para o seguro de crédito interno no período de 01/2006 a 12/2014, com variáveis exógenas PIB e câmbio.

que mostrou mais aderente aos dados foi a “média de 6 meses” que contém a média das últimas seis observações. Dessa forma, o efeito da variação da taxa de crescimento do PIB é suavizado e um valor positivo nesta média não necessariamente significa valores positivos nas últimas observações.

3.3.2.3

Modelagem da estrutura de dependência via cópulas

Os resultados apresentados nas subseções anteriores indicaram um bom ajuste dos modelos selecionados para as séries univariadas de número de sinistros de seguro de fiança locatícia e de seguro de crédito interno. Entretanto, é possível que haja alguma dependência estocástica entre as séries. Dessa forma, inicialmente investigamos a existência de dependência nos resíduos dos modelos ajustados às séries.

Conforme descrito na subseção 3.2.2.1, o ajuste de cópulas é realizado na série das pseudo-observações obtidas por meio da transformação integral de probabilidade para a distribuição Uniforme(0,1). Nas subseções 3.3.2.1 e 3.3.2.2, os resíduos quantílicos dos modelos ajustados foram checados e as hipóteses de normalidade, não autocorrelação serial e homocedasticidade não foram rejeitadas. Adicionalmente, aplicamos nos resíduos quantílicos o teste

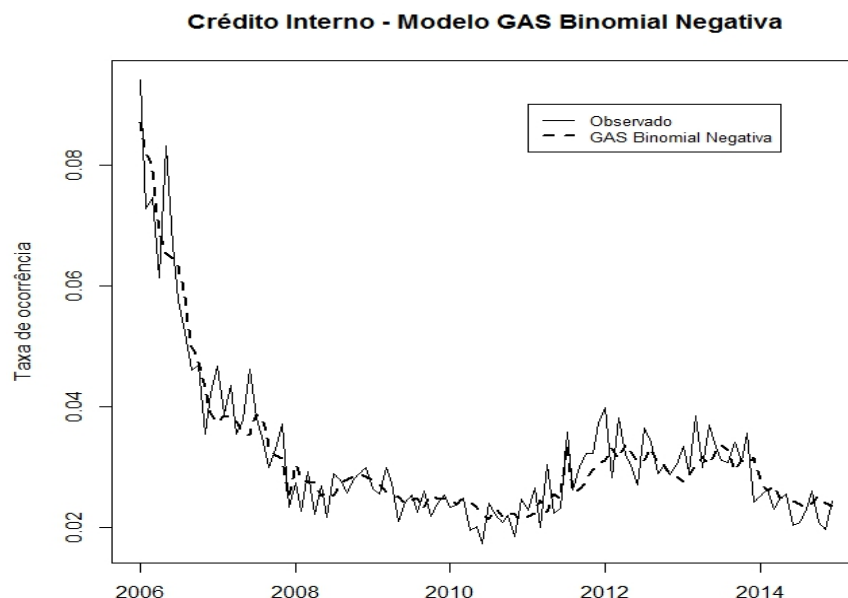


Figura 3.7: Taxa de ocorrência de sinistros observada, em linha sólida, versus taxa de ocorrência estimada, em linha pontilhada, pelo modelo GAS Binomial Negativa(1,1) para o seguro de crédito interno no período de 01/2006 a 12/2014, com variáveis exógenas PIB e câmbio.

BDS (ver Brock et al. [93]) com hipótese nula de que a série é i.i.d., não tendo sido rejeitada a hipótese nula para nenhuma das duas séries. Dessa forma, por construção, as pseudo-observações das séries podem ser consideradas independentes e uniformemente distribuídas.

Na Figura 3.8, apresentamos o gráfico de dispersão das séries de pseudo-observações de seguro de fiança locatícia e crédito interno. A partir da análise desse gráfico não é possível avaliar um padrão de dependência entre as séries. Entretanto, o Kendall-plot sob hipótese nula de independência, apresentado na Figura 3.9, indica uma possível dependência positiva entre as séries. A Tabela 3.17 apresenta as medidas de dependência ρ de Spearman e τ de Kendall, bem como o p-valor do teste para aferir independência com base em cada uma delas. As medidas indicam uma dependência positiva mas não muito forte. Tomando por base o resultado com a medida τ de Kendall, com um nível de significância de 5%, podemos rejeitar a hipótese nula de independência.

Tabela 3.17: Medidas e testes para dependência.

	Estimativa	p-valor
τ de Kendall	0.1585	0.015
ρ de Spearman	0.2226	0.021

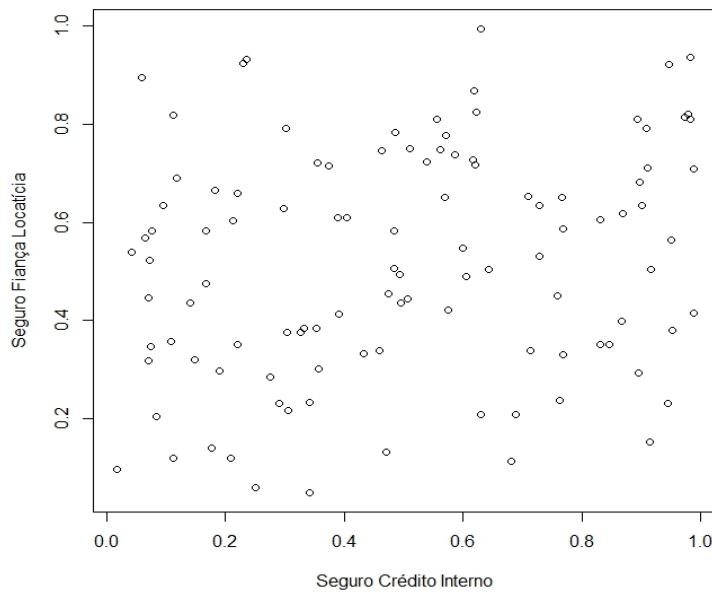


Figura 3.8: Gráfico de dispersão das séries de pseudo-observações de seguro de fiança locatícia e crédito interno.

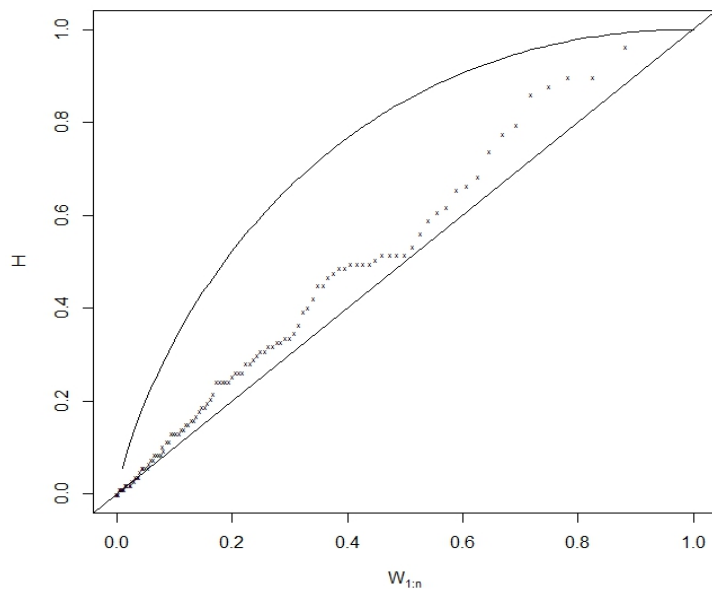


Figura 3.9: Kendall-plot sob hipótese nula de independência para as séries de pseudo-observações de seguro de fiança locatícia e de crédito interno.

Na busca pelo modelo de cópulas mais adequado, ajustamos diferentes cópulas bivariadas da família das cópulas elípticas (Normal e t -Student) bem como da família das cópulas Arquimedianas: Frank, para dependência

simétrica; Gumbel e Joe, para dependência positiva de cauda superior; e Clayton, para dependência positiva de cauda inferior. Como as estimativas dos coeficientes de dependência foram positivas, só foram ajustadas, dentre as cópulas Arquimedianas, cópulas que acomodassem dependência positiva. Adicionalmente, ajustamos cópulas com rotação de 180 graus: Survival Gumbel e Survival Joe, para dependência positiva de cauda inferior, e Survival Clayton, para dependência positiva de cauda superior. Em termos práticos, se u e v possuem estrutura de dependência de uma cópula Gumbel, então $1 - u$ e $1 - v$ possuem estrutura de dependência de uma Survival Gumbel, que ao invés de dependência na cauda superior, apresenta dependência na cauda inferior. Os resultados estão apresentados na Tabela 3.18.

Tabela 3.18: Estimativas dos parâmetros e os respectivos erros padroes, bem como os valores do logaritmo da verossimilhança e das medida AIC e BIC, para os modelos de cópulas.

cópula	$\hat{\theta}_1$	$\hat{\theta}_2$	log-Lik	AIC	BIC
Normal	0.301*** (0.105)	-	2.967	-3.933	-1.251
t -Student	0.301*** (0.105)	1.4e+04 (18264)	2.967	-1.933	3.431
Frank	1.586** (0.689)	-	2.567	-3.133	-0.451
Gumbel	1.159*** (0.093)	-	1.895	-1.791	0.891
Joe	1.175*** (0.126)	-	1.224	-0.448	2.233
Clayton	0.454** (0.191)	-	2.831	-3.661	-0.979
Survival Gumbel	1.234*** (0.106)	-	2.992	-3.984	-1.302
Survival Joe	1.342*** (0.167)	-	2.500	-2.999	-0.317
Survival Clayton	0.269** (0.110)	-	1.925	-1.185	0.832

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

Conforme citado por Manner [94], a medida AIC é um dos critérios mais utilizados para selecionar o modelo de cópulas, em uma comparação entre vários modelos ajustados. Após estudos de simulação, Manner [94] aponta que o AIC deve ser considerado primeiro quando se estiver na busca pela cópula que captaria a estrutura de dependência presente nos dados. Sendo assim, levando em consideração a medida AIC, bem como o valor do logaritmo da verossimilhança e a medida BIC, a cópula que apresentou melhor ajuste aos dados foi a copula Survival Gumbel com parâmetro estimado igual a 1.234, que

corresponde a uma cópula Gumbel com rotação de 180 graus, caracterizando uma dependência positiva de cauda inferior. A Figura 3.10 apresenta gráfico da densidade de uma cópula Survival Gumbel com parâmetro $\theta = 1.234$.

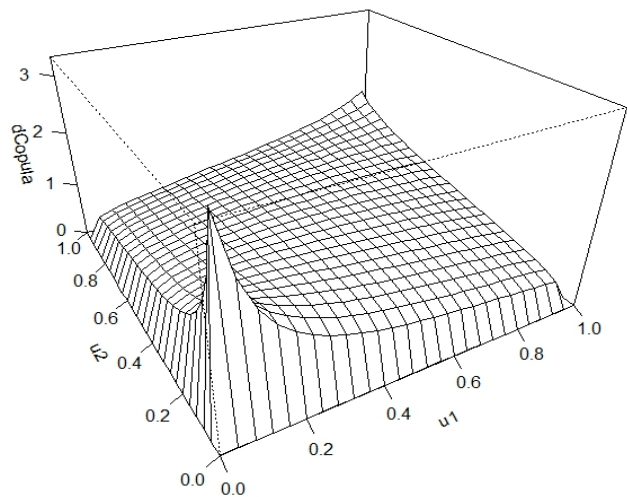


Figura 3.10: Densidade de uma cópula Survival Gumbel com parâmetro $\theta = 1.234$.

Para avaliação do ajuste do modelo de cópula não variante no tempo, conforme descrito na seção 3.2.2.2, realizamos o teste de Cramér-von Mises (CvM) com base na cópula empírica. O resultado do teste para a cópula Survival Gumbel, que apresentou melhor ajuste aos dados, está apresentado na Tabela 3.19, indicando a não rejeição da hipótese nula de que a estrutura de dependência dos dados seguiria uma cópula Survival Gumbel.

Tabela 3.19: Teste Cramer-von Mises para a cópula Survival Gumbel.

	Estimativa	p-valor
CvM com cópula empírica	0.1584	0.3004

Adicionalmente aos modelos de cópulas estáticas, efetuamos o ajuste do modelo de cópula dinâmica citado na seção 3.2.2.1 com distribuição t -Student bivariada ($k = 2$), assumindo matriz de covariância Σ_t variante no tempo, com variâncias constantes e apenas o parâmetro ρ_t variante no tempo. Os resultados do modelo t -GAS(1,1) estão apresentados na Tabela 3.20 e pode ser constatado que os parâmetros A_1 e B_1 , responsáveis pela dinâmica do modelo GAS, não são estatisticamente significativos e a estimativa de A_1 é igual a 0.000099.

Sendo assim, conforme mencionado em Patton [95], a dinâmica do coeficiente de correlação entre as variáveis pode ser adotada como estática. A Figura 3.11 apresenta a evolução da estimativa do parâmetro de correlação ρ_t e corrobora essa constatação: a variação da estimativa do parâmetro ρ_t é mínima, variando apenas de 0.2314 a 0.2320. Cabe destacar que o fato de dispormos de apenas 108 observações mensais para essa análise prejudica a detecção de dependência dinâmica.

Tabela 3.20: Estimativas de máxima verossimilhança e erros padrões para o modelo de cópula t -GAS(1,1).

Parâmetros	t -GAS(1,1)
w	0.381 (8.269)
A_1	9.9e-05 (0.002)
B_1	0.715 (6.185)
ν	49.005*** (0.056)

Significância aos níveis de 10%, 5%, e 1% é denotada por *, **, e ***, respectivamente.

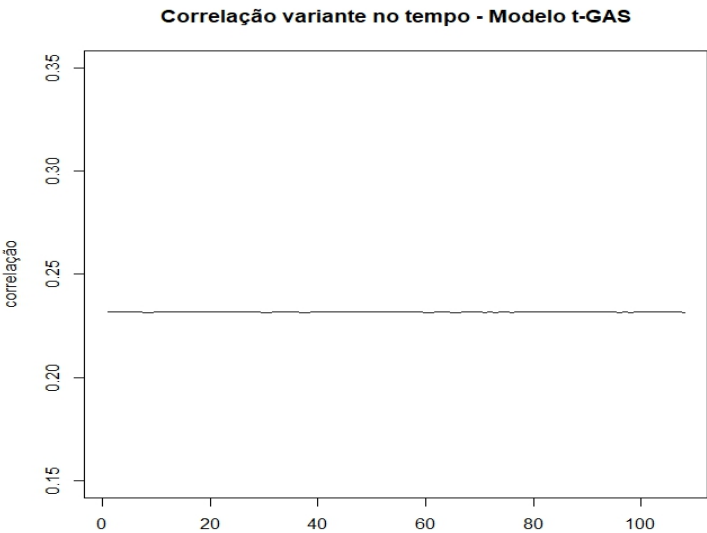


Figura 3.11: Evolução da estimativa do parâmetro ρ_t para o modelo de cópula t -GAS(1,1).

3.3.3 Previsão

Considerando os resultados das subseções 3.3.2.1, 3.3.2.2 e 3.3.2.3, é possível efetuar previsões para número de sinistros e taxa de ocorrência de sinistro nos meses adiantes.

Nos modelos GAS, a previsão um passo à frente é realizada de forma direta, uma vez que a evolução do vetor do parâmetro variante no tempo f_t leva à obtenção da distribuição preditiva $p(y_{t+1}|f_{t+1}, \mathcal{F}_t; \theta)$. Para k -passos à frente, $k \geq 2$, as previsões são obtidas por simulação de Monte Carlo. Para os intervalos de confiança de previsões fora da amostra, adotamos a metodologia para intervalos de múltipla filtragem proposta por Blasques et al. [50]. Esses intervalos levam em conta tanto a incerteza do vetor de parâmetro estático θ como a incerteza da inovação. A construção do IC se inicia com a geração de M valores θ^i da distribuição de $\hat{\theta}_T$ e, em seguida requer que seja rodado o filtro (equação de atualização 3-8) de $t = 1$ a $t = T$ para cada valor simulado θ^i , o que resulta em M cenários gerados para \hat{f}_t^i . Tendo em vista que a distribuição amostral finita de $\hat{\theta}_T$ é desconhecida, simulamos valores da distribuição assintótica

$$\hat{\theta}_T^i \sim N(\hat{\theta}_T, T^{-1}\hat{W}),$$

sendo \hat{W} obtida por meio da inversa da matriz Hessiana.

A partir de cada um dos M pares $(\hat{\theta}_T^i, \hat{f}_{T+1}^i)$, geramos S valores $\hat{y}_{T+1}^1, \dots, \hat{y}_{T+1}^S$. Utilizando a equação de atualização do parâmetro variante no tempo, obtemos $M \times S$ valores $\hat{y}_t^{i,s}$ e $\hat{f}_t^{i,s}$ a cada instante $t = T + 2, \dots, T + k$. O algoritmo para esse procedimento está descrito na seção A.2 do Apêndice A. Deixamos as observações dos últimos 12 meses de fora da amostra para avaliação da previsão dos modelos propostos.

As Figuras 3.12 e 3.13 apresentam as previsões fora da amostra para número de sinistros e taxa de ocorrência, respectivamente, para o seguro de fiança locatícia, bem como o IC 95%, considerando o modelo selecionado na subseção 3.3.2.1. As Figuras 3.14 e 3.15 apresentam os mesmos gráficos de previsão e IC para o seguro de crédito interno, considerando o modelo selecionado na subseção 3.3.2.2.

A Tabela 3.21 apresenta os valores previstos e os valores observados de número de sinistros e de taxa de ocorrências de sinistros das duas linhas de negócio para os 12 meses de previsão, bem como das medidas de erro MAPE, considerando os modelos selecionados nas subseções 3.3.2.1 e 3.3.2.2.

Para incorporar a dependência nas previsões para os instantes $t =$

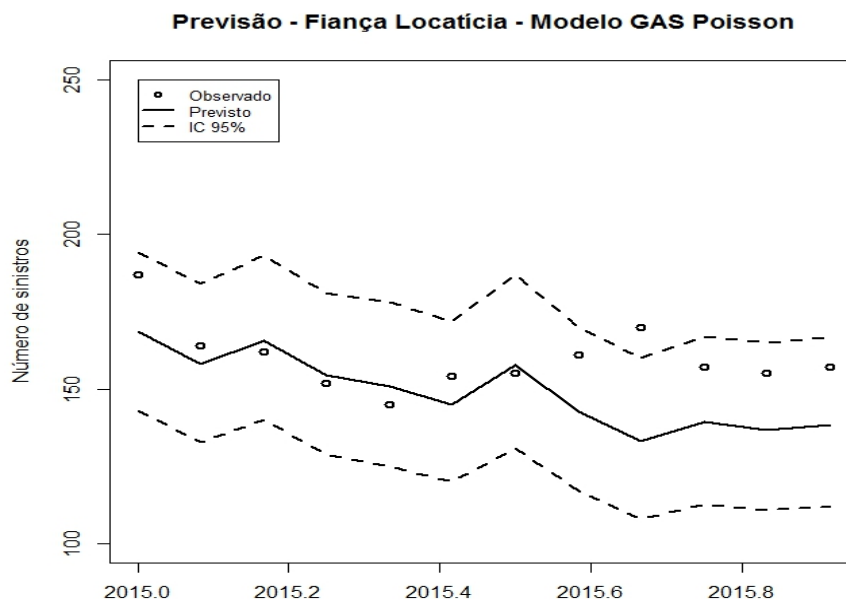


Figura 3.12: Previsão *fora da amostra* para número de sinistro de seguro de fiança locatícia, no período de 01/2015 a 12/2015, com modelo GAS Poisson(1,1) com variáveis exógenas IPCA e PIB.

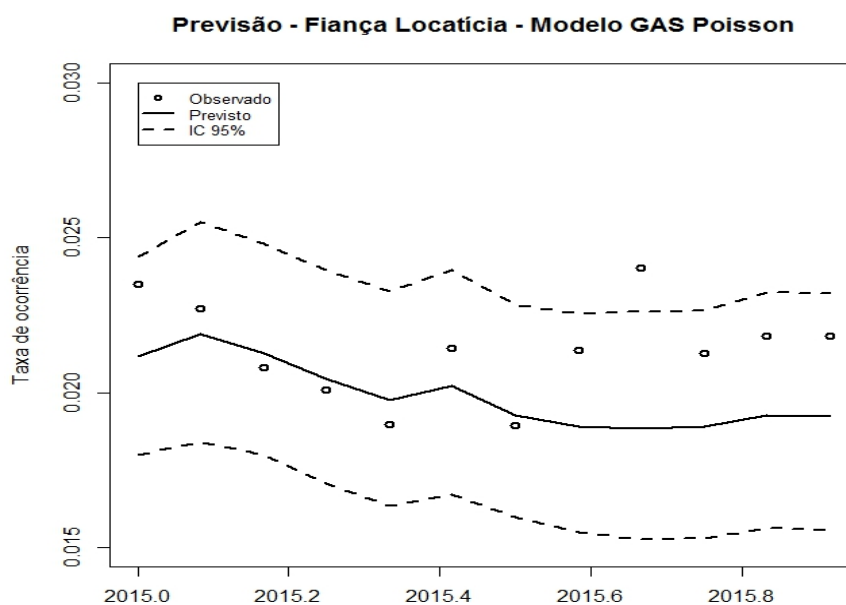


Figura 3.13: Previsão *fora da amostra* para taxa de ocorrência de sinistro de seguro de fiança locatícia, no período de 01/2015 a 12/2015, com modelo GAS Poisson(1,1) com variáveis exógenas IPCA e PIB.

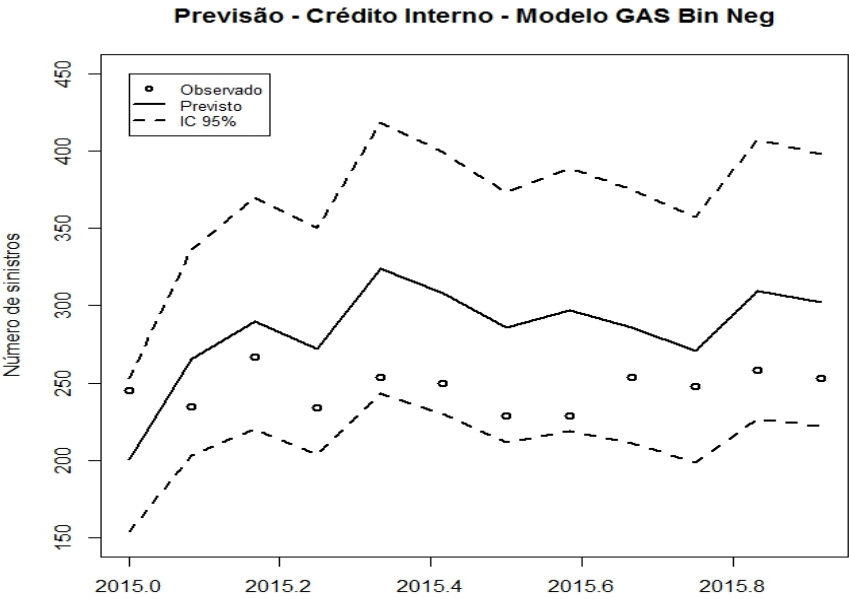


Figura 3.14: Previsão *fora da amostra* para número de sinistro de seguro de crédito interno, no período de 01/2015 a 12/2015, com modelo GAS Binomial Negativa(1,1), com variáveis exógenas PIB e câmbio.

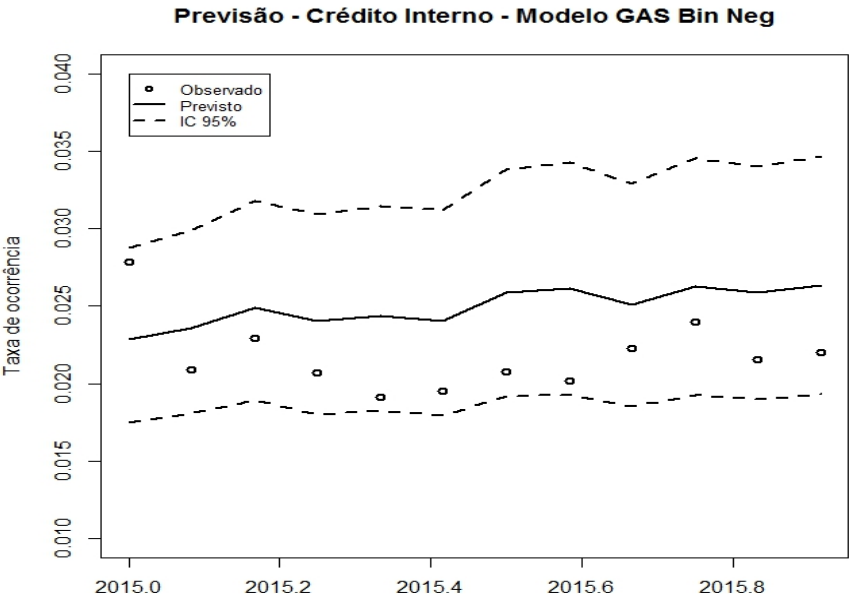


Figura 3.15: Previsão *fora da amostra* para taxa de ocorrência de sinistro de seguro de crédito interno, no período de 01/2015 a 12/2015, com modelo GAS Binomial Negativa(1,1), com variáveis exógenas PIB e câmbio.

$T + 1, \dots, T + k$, faz-se necessário gerar pares correlatados (\hat{u}_t, \hat{v}_t) a partir da cópula com parâmetro estimado $\hat{\theta}_{cop}$. A partir de cada um dos M pares $(\hat{\theta}_{j,T}^i, \hat{f}_{j,T+1}^i)$, $j = 1$ ou 2 (representando cada série) e $i = 1, \dots, M$, obtidos quando da elaboração do intervalo de confiança das previsões univariadas, geramos S pares $(\hat{u}_{j,T+1}^s, \hat{v}_{j,T+1}^s)$, $s = 1, \dots, S$. Dessa forma, podemos obter as $M \times S$ previsões $\hat{y}_{j,T+1}^{i,s}$ por meio da função quantil das distribuições marginais condicionais, utilizando os valores gerados de $(\hat{u}_{T+1}^{i,s}, \hat{v}_{T+1}^{i,s})$ e as estimativas dos parâmetros das distribuições marginais condicionais obtidos no passo anterior. Para $t = T + k$, sendo $k \geq 2$, utilizando a equação de atualização do parâmetro variante no tempo, a cada instante $t = T + 2, \dots, T + k$, obtemos $M \times S$ valores de $\hat{f}_{j,t}^{i,s}$, geramos novamente pares $(\hat{u}_t^{i,s}, \hat{v}_t^{i,s})$ e obtemos, então, as previsões $\hat{y}_{j,t}^{i,s}$, $j = 1, 2$.

Para considerar também a incerteza do parâmetro da cópula nas previsões, podemos usar o método *bootstrap* (para maiores detalhes veja Efron & Tibshirani [96]), que baseia-se em reamostragens a partir da amostra original. Dessa forma, ao invés de gerar os pares correlatados (\hat{u}_t, \hat{v}_t) sempre a partir da cópula com parâmetro estimado $\hat{\theta}_{cop}$, conforme relatado no parágrafo anterior, utilizamos previamente o método de *bootstrap* na série bivariada de pseudo-observações (u_t, v_t) , com B reamostragens, e ajustamos a função de cópula obtendo B estimativas de θ_{cop} . Assim, a cada passo de geração de novos pares correlatados (\hat{u}_t, \hat{v}_t) , sorteamos um valor no vetor de estimativas de θ_{cop} $(\hat{\theta}_{cop}^1, \dots, \hat{\theta}_{cop}^B)$ a ser usado. O algoritmo para esse procedimento está descrito na seção B.4 do Apêndice B.

Tabela 3.21: Previsão fora da amostra e MAPE, no período de 01/2015 a 12/2015, dos modelos GAS Poisson(1,1) com variáveis exógenas IPCA e PIB, para seguro de fiança locatícia, e GAS Binomial Negativa(1,1), com variáveis exógenas PIB e câmbio, para seguro de crédito interno.

	1	2	3	4	5	6	7	8	9	10	11	12	MAPE(%)
<i>meses à frente</i>													
Fiança Locatícia													
<i>Número de sinistros</i>													
Valores observados	187	164	162	152	145	154	155	161	170	157	155	157	
GAS Poisson(1,1)	168.457	158.018	165.523	154.553	151.011	145.072	157.789	142.613	133.259	139.420	136.894	138.397	8.076
<i>Taxa de ocorrência*</i>													
Valores observados	0.02351	0.02271	0.02082	0.02010	0.01896	0.02143	0.01894	0.02135	0.02403	0.02128	0.02182	0.02181	
GAS Poisson(1,1)	0.02118	0.02188	0.02127	0.02044	0.01975	0.02020	0.01928	0.01891	0.01884	0.01889	0.01927	0.01922	8.076
Crédito Interno													
<i>Número de sinistros</i>													
Valores observados	245	235	267	234	254	250	229	229	254	248	258	253	
GAS BinNeg(1,1)	200.852	265.220	289.507	271.800	323.665	307.799	285.683	297.0339	286.1156	271.0881	309.4566	302.5386	18.495
<i>Taxa de ocorrência</i>													
Valores observados	0.02786	0.02091	0.02294	0.02069	0.01912	0.01953	0.02074	0.02016	0.02229	0.02400	0.02156	0.02204	
GAS BinNeg(1,1)	0.02284	0.02360	0.02488	0.02403	0.02437	0.02404	0.02587	0.02615	0.02511	0.02624	0.02587	0.02636	18.495

*A taxa de ocorrência corresponde ao número de sinistros no tempo t dividido pela exposição no tempo t . Sendo assim, por construção, cada modelo apresentará medidas MAPE iguais para número de sinistros e para taxa de ocorrência.

A incorporação da dependência nas previsões faz sentido quando trabalhamos com as duas variáveis aleatórias de forma conjunta. Por exemplo, podemos estar interessados em estimar a despesa total com sinistros (S_t^*) da seguradora considerando as duas carteiras. Nesse sentido, para demonstrar a incorporação de dependência, podemos fazer um exercício empírico para apurar o valor de $\hat{S}_t^* = k_1 \hat{y}_{1,t} + k_2 \hat{y}_{2,t}$, onde k_1 é o valor médio de sinistros da carteira de seguro de fiança locatícia e k_2 é o valor médio de sinistros da carteira de seguro de crédito interno, uma vez detectada dependência entre as séries, devemos incorporá-la na previsão. Em relação à previsão de cada série isoladamente, a introdução de cópulas não traz alterações, a menos de flutuações estatísticas por conta da simulação.

Como o objetivo desse trabalho é modelar número de sinistros, sem adentrar na questão de modelagem da severidade, para que possamos apresentar o cálculo de S_t^* , supomos um valor constante e arbitrário para a severidade de cada linha de negócio⁷. Dessa forma, utilizando simulação de Monte Carlo, podemos obter as distribuições simuladas de S_t^* que são de suma importância para cálculo de prêmio de seguro, provisões e de necessidade de capital. Podemos, também, calcular medidas de risco, como o VaR (valor em risco) que possibilita avaliar o risco de perdas extremas para a seguradora.

Inicialmente, efetuamos cálculo da previsão dentro da amostra um passo à frente para S_t^* ($t = 1, \dots, 108$), por meio de simulação, tendo por base os vetores de média condicional variante no tempo já estimados, bem como a estimativa do parâmetro estático $\hat{\phi}$ da distribuição binomial negativa para o caso do seguro de crédito interno. A partir dos valores simulados, para a modelagem independente e a modelagem com dependência, calculamos o VaR 95% a cada instante e efetuamos o teste de Kupiec (ver Kupiec [97]) considerando os valores “observados” de S_t^* (calculados com base nos valores observados de número de sinistros e nos valores arbitrados para severidade). No teste de Kupiec, utiliza-se um processo de aferição retroativa da proporção de vezes em que o nível pré-especificado para VaR é ultrapassado, após a comparação entre a estimativa de perda potencial e o valor efetivamente observado. Nos dois casos, independente e com dependência, o resultado do teste, com p-valor de 0.2825 e 0.498, respectivamente, indica a não rejeição da hipótese nula de que a proporção verdadeira de falhas do modelo é igual ao nível nominal pré-especificado para o VaR. Entende-se por falha a ocorrência de perdas (despesa total) que exceda à estimativa de VaR para o mês em questão. Ressaltamos que a pouca quantidade de dados faz com que o efeito da dependência na simulação de S_t^* não seja percebido com tanta intensidade.

⁷Em nossa aplicação, supomos $k_1 = \text{R}\$500.00$ e $k_2 = \text{R}\$800.00$.

Cabe destacar, também, que o poder do teste de Kupiec, que representa a capacidade de rejeitar um modelo não adequado, diminui com poucas observações. A Figura 3.16 apresenta os valores “observados” de S_t^* e valores de VaR 95%, dentro da amostra (jan/2006 a dez/2014).

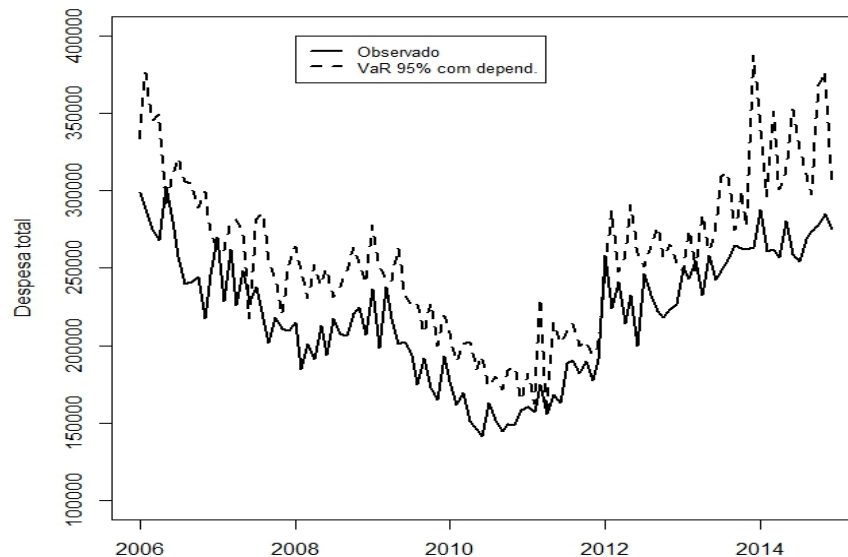


Figura 3.16: Valores “observados” (calculados com base nos valores observados de número de sinistros e nos valores arbitrados para severidade) de despesa total (S_t^*), em preto, e valores de VaR 95%, dentro da amostra (jan/2006 a dez/2014), um passo à frente, considerando modelagem com dependência, em azul.

Posteriormente, utilizando a metodologia já descrita de incorporação da dependência nas previsões, foram obtidas, via simulação, as distribuições de S_t^* para o período fora da amostra (jan/2015 a dez/2015). A Figura 3.17 apresenta a previsão fora da amostra e o intervalo de confiança 95% para S_t^* considerando a modelagem independente e a modelagem com dependência. Percebe-se, como esperado, um intervalo de confiança mais largo para a modelagem com dependência. A Figura 3.18 apresenta alguns outros quantis das distribuições simuladas de S_t^* para o período fora da amostra, onde também pode ser verificado intervalos mais largos para modelagem com dependência. Na Tabela 3.22 constam os valores de VaR 95%, fora da amostra (jan/2015 a dez/2015), para despesa total (S_t^*) também considerando modelagem independente e modelagem com dependência.

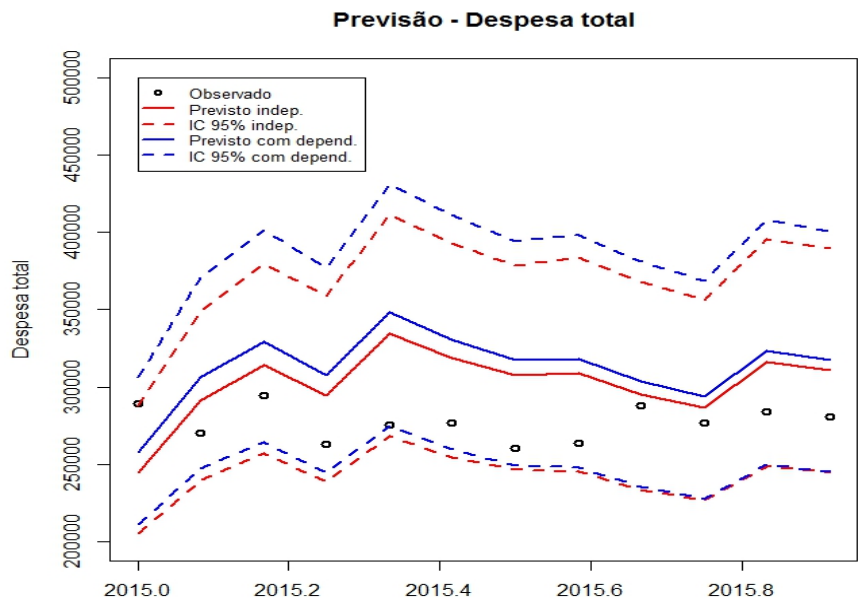


Figura 3.17: Previsão fora da amostra para para despesa total (S_t^*) e intervalo de confiança 95% considerando modelagem independente e modelagem com dependência para jan/2015 a dez/2015.

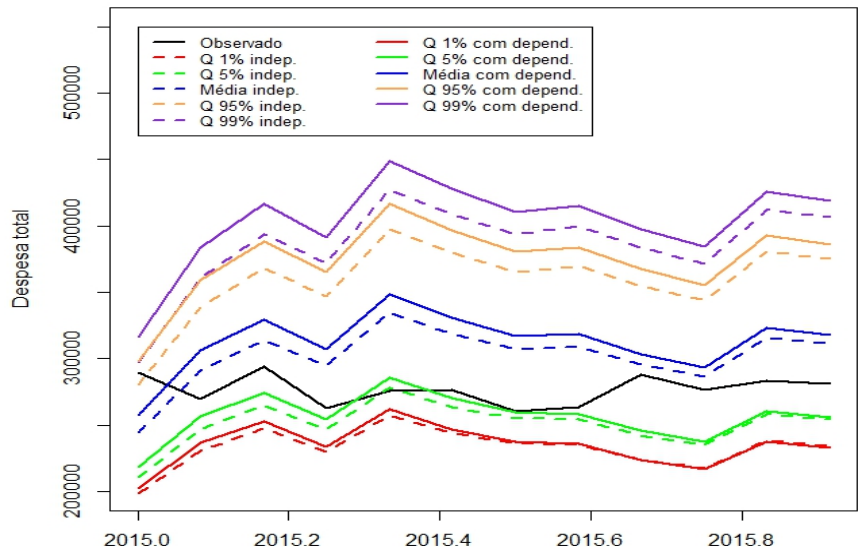


Figura 3.18: Quantis e médias das distribuições simuladas para despesa total (S_t^*), nos 12 meses à frente (01/2015 a 12/2015), considerando modelagem independente e modelagem com dependência.

Tabela 3.22: Valores “observados” de despesa total (S_t^*), em Reais, e valores de VaR 95%, fora da amostra, considerando modelagem independente e modelagem com dependência para jan/2015 a dez/2015.

Meses à frente	S_t	VaR 95% indep.	VaR 95% com depend.
1	289,500.00	280,800.00	298,200.00
2	270,000.00	338,900.00	359,300.00
3	294,600.00	368,200.00	388,700.00
4	263,200.00	347,100.00	365,200.00
5	275,700.00	397,300.00	416,500.00
6	277,000.00	379,700.00	396,800.00
7	260,700.00	365,600.00	380,700.00
8	263,700.00	369,900.00	383,900.00
9	288,200.00	355,000.00	367,600.00
10	276,900.00	343,700.00	355,200.00
11	283,900.00	380,900.00	392,500.00
12	280,900.00	375,200.00	385,900.00

3.4

Considerações Finais

Esse capítulo propõe uma nova abordagem para modelagem e previsão de número de sinistros de seguros financeiros com vistas a atender aos requisitos internacionais da IAIS [16] sobre avaliação consistente de mercado, que significa levar em consideração as informações disponíveis nos mercados financeiros e de capital relevantes, utilizando metodologias e parâmetros consistentes com esses mercados.

A modelagem do número de sinistros em seguros financeiros, incluindo suas diferentes linhas de negócio, está intimamente ligada à modelagem de risco de *default*. As linhas de negócios utilizadas foram o seguro de fiança locatícia e o seguro de crédito interno, sendo que ambos apresentam como característica marcante o fato de o sinistro ser o resultado de um evento de *default*. Muitos trabalhos na literatura abordam a relação entre exposição ao risco de crédito e condições da macroeconomia. Nesta linha, utilizamos variáveis macroeconômicas nos modelos de forma a tornar as análises e previsões consistentes com o cenário econômico.

A abordagem proposta foi aplicada em duas etapas. Inicialmente, as séries temporais bivariadas foram modeladas separadamente, com três tipos de modelos autorregressivos e com séries de variáveis exógenas com diferentes defasagens. Após identificação dos modelos e das séries exógenas que melhor se ajustaram aos dados, foi realizado um refinamento dos modelos, mantendo apenas as variáveis exógenas cujos coeficientes eram estatisticamente significantes. Na segunda etapa, foi investigada a existência de dependência entre os

resíduos dos modelos de cada série e utilizadas diferentes funções de cópulas para modelar a estrutura de dependência dos dados condicionado aos modelos ajustados na primeira etapa.

Dentre os modelos autorregressivos avaliados, tanto para seguro de fiança locatícia como para o seguro de crédito interno, o modelo GAS se mostrou com melhor ajuste. Para o seguro de fiança locatícia, a distribuição Poisson se mostrou mais adequada, com variáveis exógenas IPCA e PIB; enquanto que para o seguro de crédito interno, a distribuição Binomial Negativa obteve um melhor ajuste, com variáveis exógenas PIB e câmbio.

Para o ajuste de cópulas foi adotado o método IFM (*Inference Functions for Margins*), no qual primeiro ajustamos as distribuições das marginais e então, em um segundo passo, ajustamos a cópula às pseudo-observações das marginais. A cópula resume a estrutura de dependência da distribuição conjunta, independentemente da especificação das distribuições marginais, sendo, portanto, muito conveniente para o estudo da estrutura de dependência quando os dados se ajustam a diferentes distribuições, como a aplicação apresentada nesse trabalho. Foram testadas diferentes funções de cópulas estáticas, tendo a cópula Survival Gumbel apresentado melhor ajuste segundo o critério de AIC. Adicionalmente, foi avaliado modelo de cópula dinâmica, entretanto, não foi identificada dinâmica temporal na estrutura de dependência dos dados.

A partir da modelagem das marginais com modelos dinâmicos orientados por *score* e do uso de cópulas para modelar a estrutura de dependência foi possível obter a distribuição conjunta das perdas (despesa total) das duas linhas de negócio. A estrutura de dependência então modelada interfere diretamente na distribuição da soma total das perdas da sociedade seguradora. Por meio de avaliação dentro e fora da amostra, foi verificada a relevância de utilizar modelos dinâmicos que capturem evolução condicionada no tempo, bem como modelo de estrutura de dependência, não linear, a fim de tratar da situação de dependência relevante nas caudas.

Esse trabalho traz uma importante contribuição para modelagem e previsão de número de sinistros e para gestão e mensuração de riscos das seguradoras. Uma possível extensão desse estudo é a utilização da metodologia proposta em conjunto com modelagem para a severidade dos sinistros. É importante ressaltar, também, que a modelagem proposta pode ser facilmente aplicada a outras linhas de negócio, assim como desenvolvida com diferentes distribuições de probabilidade que melhor se ajustem aos dados.

Os artigos desenvolvidos nesta tese apresentam aplicações de modelagem estatística/atuarial ao mercado segurador. Tal mercado, em função da vigência de inúmeros princípios de regulação de solvência, deve ou tem demandado ferramentas para gestão, avaliação e mensuração de riscos. Neste sentido, modelos matemáticos, estatísticos e atuariais têm sido utilizados para modelar a dinâmica dos riscos aos quais sociedades seguradoras estão expostas. Como estes requisitos de solvência têm cobrado maior aderência destes modelos à experiência observada das companhias, tem-se observado o uso de técnicas estatísticas cada vez mais apuradas para a captura de fatores determinantes no perfil de risco das operações securitárias.

Essa tese apresenta duas importantes contribuições para modelagem dinâmica de número de sinistros, severidade e sinistro agregado, que são a base para estudos atuariais no mercado de seguros. A primeira se refere à dinâmica temporal dos fatores de risco que costumeiramente seguradoras enfrentam. Ao modelar estes riscos, a premissa de que o comportamento de variáveis é o mesmo ao longo do tempo é forte e de difícil aceitação ou comprovação. Desta forma, a introdução de preditores que se alteram com o passar do tempo é natural e, mais, um requisito para capturar a evolução de variáveis condicionadas no tempo. O objetivo foi apresentar novas alternativas para modelagens de variáveis aleatórias no setor de seguros, que sensibilizam a distribuição de perdas totais de uma seguradora, utilizando como base o arcabouço dos modelos orientados por *score* com parâmetros variantes no tempo.

A segunda contribuição engloba modelagem dinâmica nas distribuições marginais de um problema bivariado na linha de negócio de seguro de crédito, com preditores relacionados à situação macroeconômica do país. Além disso, para obtenção da distribuição conjunta foi utilizado modelo de cópulas. Tal opção foi tomada devido ao fato da geralmente assumida premissa de independência ser bastante perigosa sob uma realidade de integração entre mercados, eventos e práticas de negócios (Breymann et al. [5], Jouanin et al. [6] e Embrechts et al. [7]). Embora independência torne os cálculos bem mais simples e diretos, para companhias de seguros, que vendem apólices de curto, médio e

longo prazos, tal premissa pode ser desastrosa (Frees & Valdez [8], Venter [9] e Frees & Wang [10]). Neste sentido, é fundamental que o estudo de dependência entre riscos seja componente importante na efetiva gestão de riscos.

Os modelos utilizados nesta tese, modelos GAS, também chamados de modelos DCS (*Dynamic Conditional Score*), são uma nova classe de modelos orientados por observações (segundo a classificação de Cox et al. [12]) para séries temporais proposto por Creal et al. [13, 14] e Harvey & Chakravarty [15]. Essa nova classe fornece um arcabouço geral para modelos com parâmetros variantes no tempo. Diferentemente da modelagem tradicional para a distribuição do número de sinistros e severidade, a utilização de modelos com parâmetros variantes no tempo permite, a princípio, que a previsão seja mais precisa, visto que o modelo vai acompanhando a evolução dos parâmetros ao longo do tempo.

Neste sentido, o primeiro artigo (capítulo 2) visou a obtenção da distribuição do sinistro agregado, que corresponde à soma aleatória dos valores de sinistros (severidade) em determinado período de tempo. Trata-se de um problema clássico na teoria do risco e fundamental para precificação de seguros, cálculo de provisões e de probabilidade de ruína. A efetiva obtenção da expressão analítica para essa distribuição de probabilidade é uma tarefa impossível na grande maioria dos casos. Neste trabalho, uma nova abordagem é proposta para a modelagem da distribuição preditiva do sinistro agregado. Combinamos os modelos orientados por *score* para especificar as distribuições não-Gaussianas tanto para o número de sinistros como para severidade. Com o uso da Transformada Rápida de Fourier, obtemos então numericamente a distribuição do sinistro agregado. Para aplicação da modelagem proposta, foram utilizados dados da carteira de seguro de automóveis de uma seguradora brasileira.

Utilizamos este novo arcabouço de modelos para séries temporais, orientados por *score* com parâmetros variantes no tempo, para modelar a distribuição preditiva do número de sinistros e dos valores de sinistros (severidade). Nossos resultados indicaram que, em comparação com distribuições estáticas, as distribuições preditivas se apresentaram mais apropriadas, possivelmente por levar em consideração a dinâmica temporal e a incerteza nos parâmetros. As aplicações do método apresentado vêm ao encontro de demandas dos atuais princípios de solvência, citados acima, e de mensuração de contratos de seguros, pois possibilitou a obtenção de distribuições preditivas, condicionadas no tempo, da perda de carteiras de seguradoras.

Até onde sabemos, poucos estudos sobre a obtenção da distribuição preditiva do sinistro agregado, com modelos dinâmicos, estão disponíveis na

literatura. Além disso, destes poucos, a maioria utiliza abordagem Bayesiana, que exige um custo computacional alto, além de requerer conhecimentos específicos para implementação dos métodos de amostragem. É importante ressaltar que a modelagem proposta pode ser facilmente aplicada a outras linhas de negócio, ou até mesmo na área de risco operacional e de crédito. Nestes casos, o sinistro passa a ser a perda operacional (para o caso de risco operacional) ou o *default* (para o caso de risco de crédito). Outra flexibilização que pode ser adotada neste tipo de abordagem é a utilização de diferentes distribuições de probabilidade que melhor se ajustem aos dados.

Já o segundo artigo (capítulo 3) propôs uma avaliação consistente de mercado para os seguros financeiros. A metodologia aplicada está em linha com práticas internacionais, afinal a demandada avaliação consistente de obrigações significa levar em consideração todas as informações disponíveis nos mercados financeiros e de capital relevantes, utilizando métodos e parâmetros consistentes com esses mercados. Em nossa pesquisa, utilizamos essas fontes de informação na modelagem dos sinistros de seguros financeiros. Um dos principais *inputs* para o mercado avaliar de forma consistente os passivos de seguro de crédito é chegar a uma previsão confiável do número de *default* (sinistros) decorrentes das apólices em vigor. Utilizamos dados reais de duas linhas de negócios de seguradoras brasileiras, com importante participação no mercado de seguros financeiros. As linhas de negócios utilizadas foram o seguro de fiança locatícia e o seguro de crédito interno. Ambos apresentam como característica marcante o fato de o sinistro ser o resultado de um evento de *default*. Para o seguro de fiança locatícia, sinistros são decorrentes do *default* do inquilino, enquanto que sinistros de seguro de crédito interno resultam de não cumprimento de contratos financeiros.

Nossos resultados neste artigo indicaram que as variáveis exógenas macroeconômicas são relevantes na modelagem e previsão de séries temporais de números de sinistros de seguros financeiros. Após a modelagem das marginais com modelos dinâmicos orientados por *score*, foram utilizadas cópulas para se obter a distribuição conjunta das perdas das linhas de negócio. A estrutura de dependência então modelada interfere diretamente na distribuição da soma total das perdas da sociedade seguradora. Por meio de avaliação dentro e fora da amostra, foi verificada a relevância de utilizar (i) modelos dinâmicos que capturem evolução condicionada no tempo e (ii) modelo de estrutura de dependência, não linear, a fim de tratar da situação de dependência relevante nas caudas

Os artigos desta tese apresentam inovações acadêmicas relevantes que podem ser aplicadas na prática por seguradores (de diversos nichos) para mo-

delar suas perdas de subscrição, ou ainda por bancos e seguradores para gerir risco operacional e/ou de crédito. Devido às contribuições apresentadas nos trabalhos, acreditamos que estes modelos serão usados como base para futuros estudos nesta e em outras direções na modelagem de riscos de subscrição, operacional e crédito e na ciência atuarial de forma geral.

Como possíveis extensões dessa tese, em relação ao primeiro artigo, que buscou a obtenção da distribuição preditiva do sinistro agregado a partir de modelos dinâmicos orientados por *score* para número de sinistros e severidade, identificamos a adaptação dos referidos modelos de forma a incorporar características de apólices sinistradas, criando uma modelagem por classes. Outra possível extensão seria a adaptação dos modelos para modelagem de sinistros de catástrofe, que são eventos raros e de alta severidade, com o uso de distribuições que consigam lidar com excesso de zeros para número de sinistros, bem como o uso de distribuições que acomodem bem valores extremos, para severidade. Em relação ao segundo artigo, caso se disponha de séries mais longas, seria possível explorar a abordagem de cópulas dinâmicas, com outras funções de cópulas não utilizadas na tese, como cópulas Arquimedianas dinâmicas. Outras possível extensão é a conjugação da abordagem do segundo artigo para número de sinistros com modelagem para severidade.

Referências bibliográficas

- [1] BIS. **International Convergence of Capital Measurement and Capital Standards.** Basel Committee on Banking Supervision, Bank for International Settlements, July 1988. Available at <http://www.bis.org/publ/bcbs04a.pdf>, 1988.
- [2] BIS. **International Convergence of Capital Measurement and Capital Standards: a Revised Framework.** Basel Committee on Banking Supervision, Bank for International Settlements, June 2004. Available at <http://www.bis.org/publ/bcbs107.pdf>, 2004.
- [3] BIS. **Basel III: International framework for liquidity risk measurement, standards and monitoring.** Basel Committee on Banking Supervision, Bank for International Settlements, December 2010. Available at <http://www.bis.org/publ/bcbs165.pdf>, 2010.
- [4] SANDSTRÖM, A.. **Handbook of solvency for actuaries and risk managers: theory and practice.** CRC Press, 2016.
- [5] BREYMAN, W.; DIAS, A. ; EMBRECHTS, P.. **Dependence structures for multivariate high-frequency data in finance.** Quantitative Finance, 3(1):1–14, 2003.
- [6] JOUANIN, J.-F.; RIBOULET, G. ; RONCALLI, T.. **Financial applications of copula functions.** RISK MEASURES FOR THE 21ST CENTURY, Par Giorgio Szego, Available at SSRN 1032588, 2004.
- [7] EMBRECHTS, P.; LINDSKOG, F. ; MCNEIL, A.. **Modelling dependence with copulas.** Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich, 2001.
- [8] FREES, E. W.; VALDEZ, E. A.. **Understanding relationships using copulas.** North American actuarial journal, 2(1):1–25, 1998.
- [9] VENTER, G. G.. **Quantifying correlated reinsurance exposures with copulas.** In: CASUALTY ACTUARIAL SOCIETY FORUM, volumen 2003, p. 215–229, 2003.

- [10] FREES, E. W.; WANG, P.. **Copula credibility for aggregate loss models**. Insurance: Mathematics and Economics, 38(2):360–373, 2006.
- [11] NELSEN, R. B.. **An introduction to copulas**, volumen 139. Springer Science & Business Media, 2013.
- [12] COX, D. R.; GUDMUNDSSON, G.; LINDGREN, G.; BONDESSON, L.; HARSAAE, E.; LAAKE, P.; JUSELIUS, K. ; LAURITZEN, S. L.. **Statistical analysis of time series: some recent developments [with discussion and reply]**. Scandinavian Journal of Statistics, p. 93–115, 1981.
- [13] CREAL, D.; KOOPMAN, S. J. ; LUCAS, A.. **A general framework for observation driven time-varying parameter models**. Tinbergen Institute Discussion Paper 08-108/4. Available at SSRN 1297183, 2008.
- [14] CREAL, D.; KOOPMAN, S. J. ; LUCAS, A.. **Generalized autoregressive score models with applications**. Journal of Applied Econometrics, 28(5):777–795, 2013.
- [15] HARVEY, A. C.; CHAKRAVARTY, T.. **Beta-t(e) garch**. University of Cambridge, Faculty of Economics, 2008.
- [16] IAIS. **Insurance core principles, standards, guidance and assessment methodology**. Basel, October 2011, 2011.
- [17] IASB. **Framework for the preparation and presentation of financial statements**. London: IASB. Available at: <www.iasb.org/IFRSs/IFRS.htm>, 2010.
- [18] KLUGMAN, S. A.; PANJER, H. H. ; WILLMOT, G. E.. **Loss models: from data to decisions**, volumen 715. John Wiley & Sons, 2012.
- [19] ROBERTSON, J. P.. **The computation of aggregate loss distributions**. In: PROCEEDINGS OF THE CASUALTY ACTUARIAL SOCIETY, volumen 79, p. 57–133, 1992.
- [20] WALLER, L. A.; TURNBULL, B. W. ; HARDIN, J. M.. **Obtaining distribution functions by numerical inversion of characteristic functions with applications**. The American Statistician, 49(4):346–350, 1995.
- [21] WANG, S.. **Aggregation of correlated risk portfolios: models and algorithms**. In: PROCEEDINGS OF THE CASUALTY ACTUARIAL SOCIETY, volumen 85, p. 848–939, 1998.

- [22] KAAS, R.; GOOVAERTS, M.; DHAENE, J. ; DENUIT, M.. **Modern actuarial risk theory: using R**, volumen 128. Springer Science & Business Media, 2008.
- [23] FRACHOT, A.; GEORGES, P. ; RONCALLI, T.. **Loss distribution approach for operational risk**. Available at SSRN 1032523, 2001.
- [24] SUISSE, C.. **Creditrisk+: A credit risk management framework**. Credit Suisse Financial Products, 1997.
- [25] HEWITT JR, C. C.. **Loss ratio distributions-a model**. PCAS LIV, p. 70, 1967.
- [26] BOWERS, N. L. N. L.; OTHERS. **Actuarial mathematics**. The society of actuaries, 1997.
- [27] PANJER, H. H.. **Recursive evaluation of a family of compound distributions**. Astin Bulletin, 12(01):22–26, 1981.
- [28] CHAUBEY, Y. P.; TRUDEAU, S. ; OTHERS. **On the computation of aggregate claims distributions: some new approximations**. Insurance: Mathematics and Economics, 23(3):215–230, 1998.
- [29] LAU, H.-S.. **An effective approach for estimating the aggregate loss of an insurance portfolio**. Journal of Risk and Insurance, p. 20–30, 1984.
- [30] SUNDT, B.. **Recursive evaluation of aggregate claims distributions**. Insurance: Mathematics and Economics, 30(3):297–322, 2002.
- [31] HECKMAN, P. E.; MEYERS, G. G.. **The calculation of aggregate loss distributions from claim severity and claim count distributions**. In: PROCEEDINGS OF THE CASUALTY ACTUARIAL SOCIETY, volumen 70, 1983.
- [32] SHEVCHENKO, P. V.. **Calculation of aggregate loss distributions**. arXiv preprint arXiv:1008.1108, 2010.
- [33] HARVEY, A.; FERNANDES, C.. **Time series models for insurance claims**. Journal of the Institute of Actuaries, 116(03):513–528, 1989.
- [34] DICKSON, D. C.; TEDESCO, L. M. ; ZEHNWIRTH, B.. **Predictive aggregate claims distributions**. Journal of Risk and Insurance, p. 689–709, 1998.

- [35] MEYERS, G. G.. **Estimating predictive distributions for loss reserve models.** *Variance*, 1(2):248–272, 2007.
- [36] BLASQUES, F.; KOOPMAN, S. J. ; LUCAS, A.. **Information theoretic optimality of observation driven time series models.** Tinbergen Institute Discussion paper 14-046/III, 2014.
- [37] BLASQUES, F.; KOOPMAN, S. ; LUCAS, A.. **Information-theoretic optimality of observation-driven time series models for continuous responses.** *Biometrika*, 102(2):325–343, 2015.
- [38] ENGLE, R. F.. **Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation.** *Econometrica: Journal of the Econometric Society*, p. 987–1007, 1982.
- [39] BOLLERSLEV, T.. **Generalized autoregressive conditional heteroskedasticity.** *Journal of econometrics*, 31(3):307–327, 1986.
- [40] BLASQUES, F.; KOOPMAN, S. J. ; LUCAS, A.. **Maximum likelihood estimation for generalized autoregressive score models.** Tinbergen Institute Discussion Paper 14-029/III. Available at SSRN 2404276, 2014.
- [41] BLASQUES, F.; KOOPMAN, S. J. ; OTHERS. **Maximum likelihood estimation for correctly specified generalized autoregressive score models: Feedback effects, contraction conditions and asymptotic properties.** Tinbergen Institute Discussion Paper 14-074/III, 2014.
- [42] CREAL, D.; SCHWAAB, B.; KOOPMAN, S. J. ; LUCAS, A.. **Observation-driven mixed-measurement dynamic factor models with an application to credit risk.** *Review of Economics and Statistics*, 96(5):898–915, 2014.
- [43] COOLEY, J. W.; TUKEY, J. W.. **An algorithm for the machine calculation of complex fourier series.** *Mathematics of computation*, 19(90):297–301, 1965.
- [44] PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T. ; FLANNERY, B. P.. **Numerical recipes in C**, volumen 2. Cambridge university press Cambridge, 1996.
- [45] BERTRAM, J.. **Numerische berechnung von gesamtschadenverteilungen.** *Blätter der DGVFM*, 15(2):175–194, 1981.

- [46] BÜHLMANN, H.. Numerical evaluation of the compound poisson distribution: recursion or fast fourier transform? *Scandinavian Actuarial Journal*, 1984(2):116–126, 1984.
- [47] R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [48] DUNN, P. K.; SMYTH, G. K.. **Randomized quantile residuals**. *Journal of Computational and Graphical Statistics*, 5(3):236–244, 1996.
- [49] KALLIOVIRTA, L.. **Misspecification tests based on quantile residuals**. *The Econometrics Journal*, 15(2):358–393, 2012.
- [50] BLASQUES, F.; KOOPMAN, S. J.; ŁASAK, K. ; LUCAS, A.. **In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models**. *International Journal of Forecasting*, 32(3):875–887, 2016.
- [51] MERTON, R. C.. **On the pricing of corporate debt: The risk structure of interest rates**. *The Journal of finance*, 29(2):449–470, 1974.
- [52] FIGLEWSKI, S.; FRYDMAN, H. ; LIANG, W.. **Modeling the effect of macroeconomic factors on corporate default and credit rating transitions**. *International Review of Economics & Finance*, 21(1):87–105, 2012.
- [53] DUFFIE, D.; SAITA, L. ; WANG, K.. **Multi-period corporate default prediction with stochastic covariates**. *Journal of Financial Economics*, 83(3):635–665, 2007.
- [54] KOOPMAN, S. J.; LUCAS, A. ; MONTEIRO, A.. **The multi-state latent factor intensity model for credit rating transitions**. *Journal of Econometrics*, 142(1):399–424, 2008.
- [55] BANGIA, A.; DIEBOLD, F. X.; KRONIMUS, A.; SCHAGEN, C. ; SCHURMANN, T.. **Ratings migration and the business cycle, with application to credit portfolio stress testing**. *Journal of banking & finance*, 26(2):445–474, 2002.
- [56] YAN, A. X.; SHI, J. ; WU, C.. **Do macroeconomic variables matter for pricing default risk?** *International Review of Economics & Finance*, 17(2):279–291, 2008.

- [57] HUANG, J.-Z.; KONG, W.. **Macroeconomic news announcements and corporate bond credit spreads**. Available at SSRN 693341, 2005.
- [58] JOE, H.. **Multivariate models and multivariate dependence concepts**. CRC Press, 1997.
- [59] LIBOSCHIK, T.; FOKIANOS, K. ; FRIED, R.. **“tscount: An r package for analysis of count time series following generalized linear models**. To appear in *Journal of Statistical Software*, 2016.
- [60] NELDER, J. A.; WEDDERBURN, R. W. M.. **Generalized linear models**. *Journal of the Royal Statistical Society*, 135:370–384, 1972.
- [61] MCCULLAGH, P.; NELDER, J. A.. **Generalized linear models**, volumen 37. CRC press, 1989.
- [62] KEDEM, B.; FOKIANOS, K.. **Regression models for time series analysis**. *Wiley Series in Probability and Statistics*. John Wiley & Sons, 2002.
- [63] ZEGER, S. L.; QAQISH, B.. **Markov regression models for time series: a quasi-likelihood approach**. *Biometrics*, p. 1019–1031, 1988.
- [64] DAVIS, R. A.; DUNSMUIR, W. T. ; STREETT, S. B.. **Observation-driven models for poisson counts**. *Biometrika*, 90(4):777–790, 2003.
- [65] JUNG, R. C.; TREMAYNE, A.. **Useful models for time series of counts or simply wrong ones?** *AStA Advances in Statistical Analysis*, 95(1):59–91, 2011.
- [66] FOKIANOS, K.. **Some recent progress in count time series**. *Statistics*, 45(1):49–58, 2011.
- [67] FOKIANOS, K.. **Count time series models**. *Time Series–Methods and Applications, Handbook of Statistics*, 30:315–347, 2012.
- [68] FOKIANOS, K.. **Statistical analysis of count time series models: A glm perspective**. *Handbook of Discrete-Valued Time Series*, p. 3–28, 2015.
- [69] WEISS, C. H.. **Thinning operations for modeling time series of counts—a survey**. *AStA Advances in Statistical Analysis*, 92(3):319–341, 2008.

- [70] BENJAMIN, M. A.; RIGBY, R. A. ; STASINOPOULOS, D. M.. **Generalized autoregressive moving average models**. Journal of the American Statistical association, 98(461):214–223, 2003.
- [71] HARVEY, A. C.; FERNANDES, C.. **Time series models for count or qualitative observations**. Journal of Business & Economic Statistics, 7(4):407–417, 1989.
- [72] FOKIANOS, K.; TJØSTHEIM, D.. **Log-linear poisson autoregression**. Journal of Multivariate Analysis, 102(3):563–578, 2011.
- [73] HABERMAN, S.; RENSHAW, A. E.. **Generalized linear models and actuarial science**. The Statistician, p. 407–436, 1996.
- [74] NEVES, C.; FERNANDES, C. ; HOELTGEBAUM, H.. **Five different distributions for the lee–carter model of mortality forecasting: A comparison using gas models**. Insurance: Mathematics and Economics, 75:48–57, 2017.
- [75] CHRISTOU, V.; FOKIANOS, K.. **Quasi-likelihood inference for negative binomial time series models**. Journal of Time Series Analysis, 35(1):55–78, 2014.
- [76] SKLAR, M.. **Fonctions de répartition à n dimensions et leurs marges**. Université Paris 8, 1959.
- [77] SCHWEIZER, B.; SKLAR, A.. **Probability metric spaces**. North-Holland, New York. Science, 222:1293, 1983.
- [78] TRIVEDI, P. K.; ZIMMER, D. M.. **Copula modeling: an introduction for practitioners**. Now Publishers Inc, 2007.
- [79] GENEST, C.; NEŠLEHOVÁ, J.. **A primer on copulas for count data**. ASTIN Bulletin: The Journal of the IAA, 37(2):475–515, 2007.
- [80] PATTON, A. J.. **Modelling asymmetric exchange rate dependence**. International economic review, 47(2):527–556, 2006.
- [81] KOOPMAN, S. J.; LIT, R. ; LUCAS, A.. **Intraday stock price dependence using dynamic discrete copula distributions**. Tinbergen Institute Discussion Paper 15-037/III/DSF90. Available at SSRN 2580840, 2015.

- [82] CREAL, D.; KOOPMAN, S. J. ; LUCAS, A.. **A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations.** *Journal of Business & Economic Statistics*, 29(4):552–563, 2011.
- [83] MENDES, B. V. M.; DE MELO, E. F. L. ; NELSEN, R. B.. **Robust fits for copula models.** *Communications in Statistics - Simulation and Computation*, 36(5):997–1017, 2007.
- [84] GENEST, C.; FAVRE, A.-C.. **Everything you always wanted to know about copula modeling but were afraid to ask.** *Journal of hydrologic engineering*, 12(4):347–368, 2007.
- [85] GENEST, C.; BOIES, J.-C.. **Detecting dependence with kendall plots.** *The American Statistician*, 57(4):275–284, 2003.
- [86] DEHEUVELS, P.. **La fonction de dépendance empirique et ses propriétés.** *académie royale de belgique. Bulletin de la Classe des Sciences*, 65(5):274–292, 1979.
- [87] DEHEUVELS, P.. **A kolmogorov-smirnov type test for independence and multivariate samples.** *Revue roumaine de mathématiques pures et appliquées*, 26(2):213–226, 1981.
- [88] DEHEUVELS, P.. **A non parametric test of independence.** *Publications de l'ISUP*, 26:29–50, 1981.
- [89] RÉMILLARD, B.. **Goodness-of-fit tests for copulas of multivariate time series.** *Working Paper.* Available at SSRN 1729982, 2010.
- [90] ROSENBLATT, M.. **Remarks on a multivariate transformation.** *The annals of mathematical statistics*, 23(3):470–472, 1952.
- [91] GENEST, C.; RÉMILLARD, B. ; BEAUDOIN, D.. **Goodness-of-fit tests for copulas: A review and a power study.** *Insurance: Mathematics and economics*, 44(2):199–213, 2009.
- [92] JIMENEZ, G.; SAURINA, J.. **Credit cycles, credit risk, and prudential regulation.** *International Journal of Central Banking*, 2(2):65–98, 2006.
- [93] BROOCK, W.; SCHEINKMAN, J. A.; DECHERT, W. D. ; LEBARON, B.. **A test for independence based on the correlation dimension.** *Econometric reviews*, 15(3):197–235, 1996.

- [94] MANNER, H.. Estimation and model selection of copulas with an application to exchange rates. METEOR research memorandum 07/056, Maastricht University, 2007.
- [95] PATTON, A. J.. Copula-based models for financial time series. In: HANDBOOK OF FINANCIAL TIME SERIES, p. 767–785. Springer, 2009.
- [96] EFRON, B.; TIBSHIRANI, R. J.. An introduction to the bootstrap. CRC press, 1994.
- [97] KUPIEC, P. H.. Techniques for verifying the accuracy of risk measurement models. The journal of Derivatives, 3(2):73–84, 1995.
- [98] FRANK, M. J.. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes mathematicae, 19(1):194–226, 1979.
- [99] NELSEN, R. B.. Properties of a one-parameter family of bivariate distributions with specified marginals. Communications in statistics-Theory and methods, 15(11):3277–3285, 1986.
- [100] GENEST, C.. Frank's family of bivariate distributions. Biometrika, 74(3):549–555, 1987.
- [101] GUMBEL, E. J.. Distributions des valeurs extrêmes en plusieurs dimensions. Publ. Inst. Statist. Univ. Paris, 9:171–173, 1960.
- [102] JOE, H.. Parametric families of multivariate distributions with given margins. Journal of multivariate analysis, 46(2):262–282, 1993.
- [103] CLAYTON, D. G.. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65(1):141–151, 1978.
- [104] GENEST, C.; RIVEST, L.-P.. Statistical inference procedures for bivariate archimedean copulas. Journal of the American statistical Association, 88(423):1034–1043, 1993.

A

Apêndice - Capítulo 2

A.1

Resíduos quantílicos

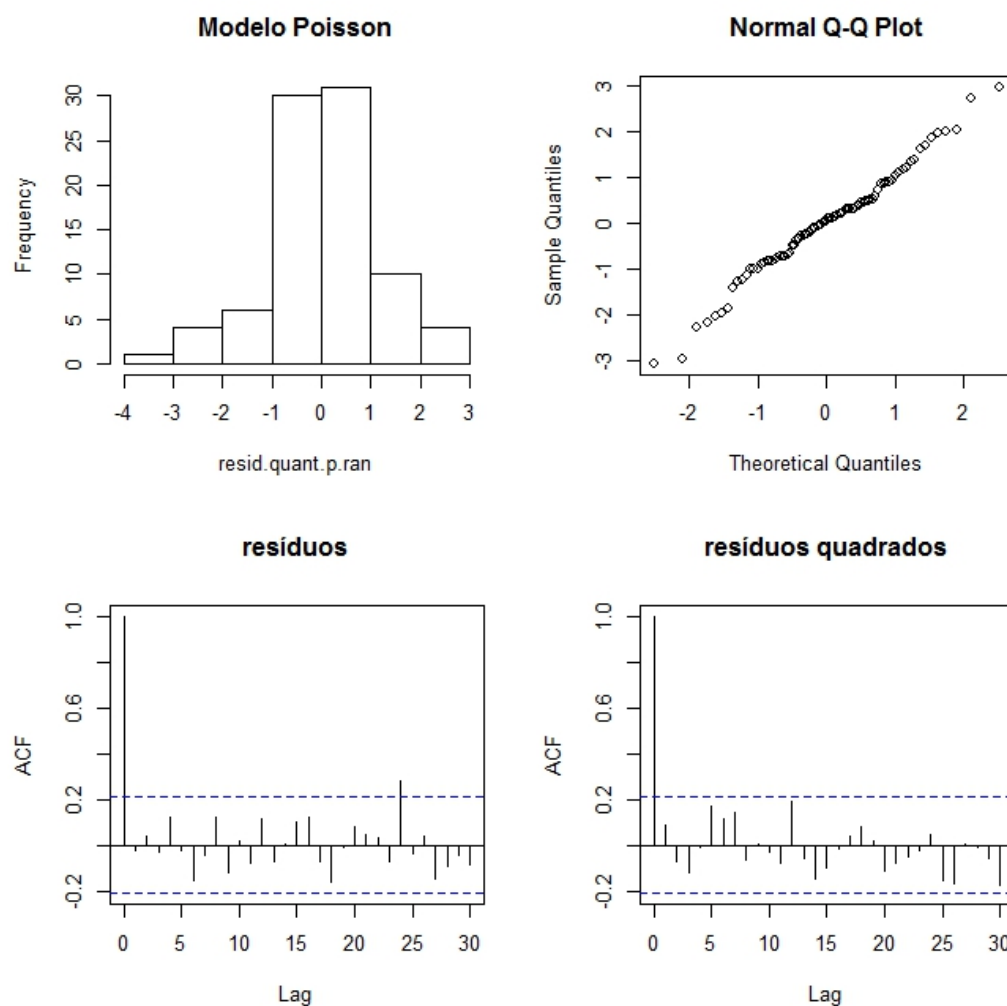


Figura A.1: Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Poisson(1,1) para número de sinistros no período 01/2006 a 02/2013.

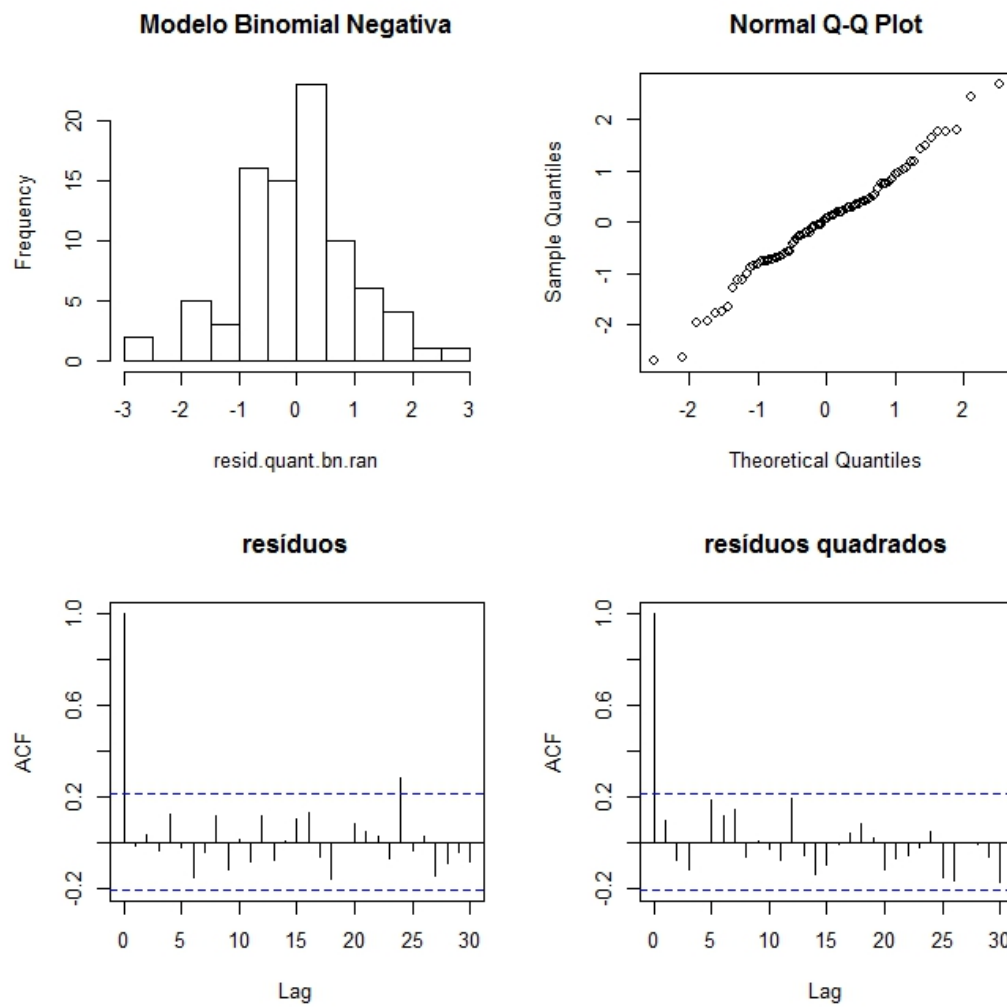


Figura A.2: Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Binomial Negativa(1,1) para número de sinistros no período 01/2006 a 02/2013.

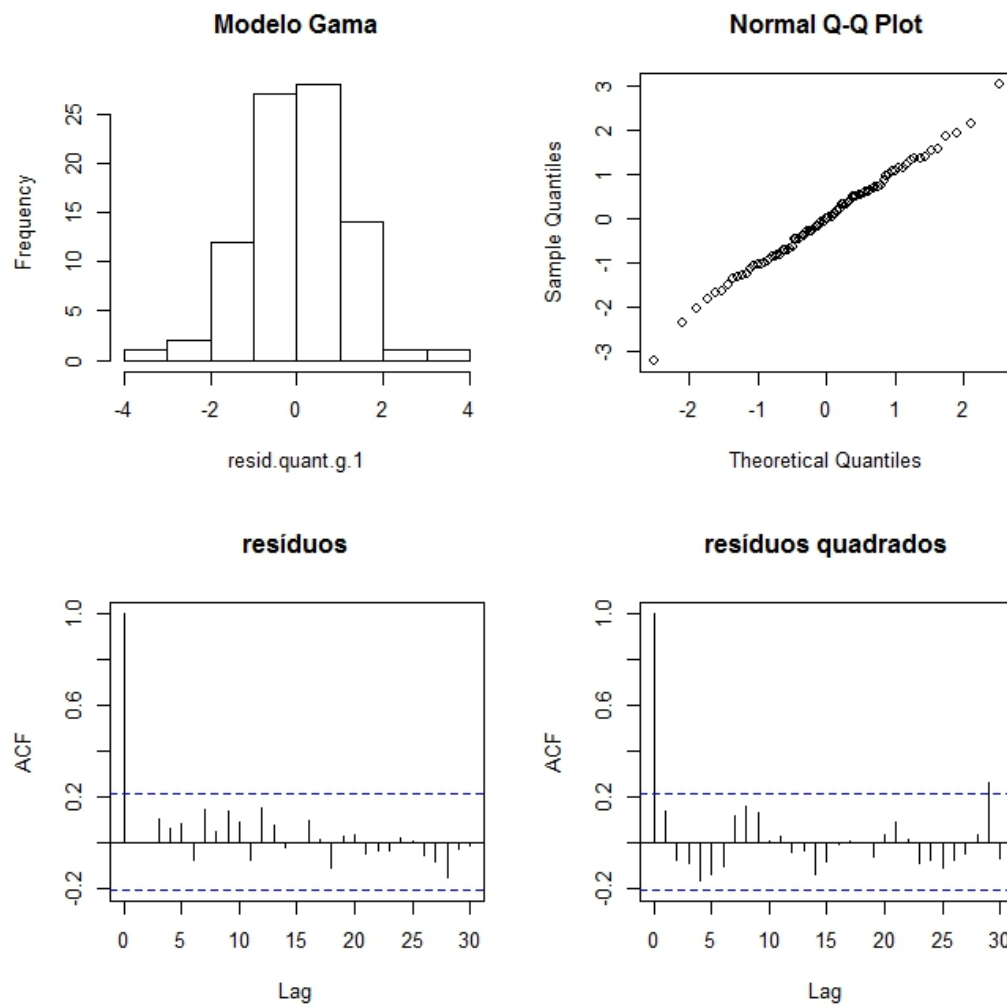


Figura A.3: Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Gama(1,1) para severidade no período 01/2006 a 02/2013.

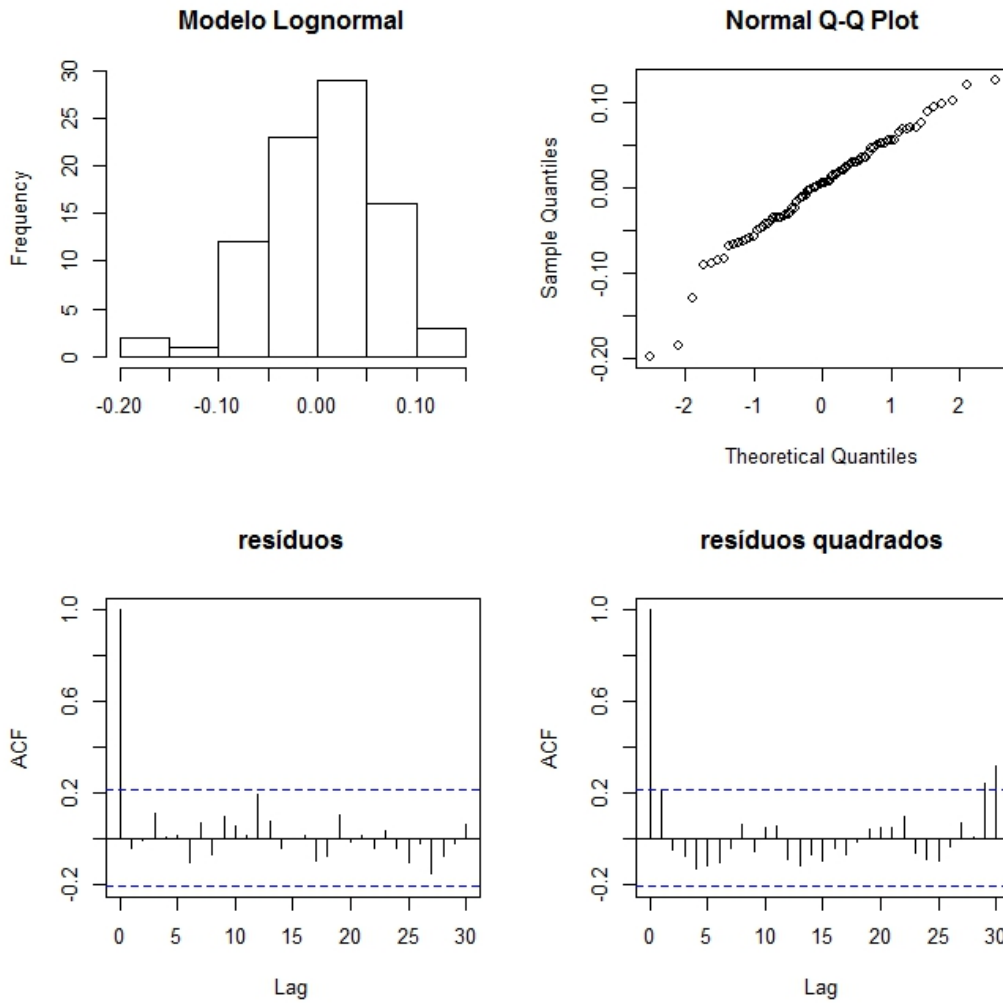


Figura A.4: Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Lognormal(1,1) para severidade no período 01/2006 a 02/2013.

A.2

Algoritmo para previsão no modelo GAS

Algoritmo para obtenção de intervalo de confiança para previsão no modelo GAS, seguindo metodologia de múltipla filtragem proposta por Blasques et al. [50].

1. Gere M valores $\hat{\theta}_T^i$, $i = 1, \dots, M$, a partir da sua distribuição assintótica

$$\hat{\theta}_T^i \sim N\left(\hat{\theta}_T, T^{-1}\hat{W}\right), \quad i = 1, \dots, M;$$

2. Para cada $\hat{\theta}_T^i$, use os dados observados y_1, \dots, y_T e a equação de atualização de f_t e rode o filtro de $t = 1, \dots, T$ para obter \hat{f}_{T+1}^i , $i = 1, \dots, M$;

3. Para cada par $(\hat{\theta}_T^i, \hat{f}_{T+1}^i)$, $i = 1, \dots, M$, siga os passos 3.1 a 3.3 para obter $M \times S$ valores de $\hat{f}_{T+k}^{i,s}$ e $\hat{y}_{T+k}^{i,s}$:
 - 3.1. Dados $\hat{\theta}_T^i$ e \hat{f}_{T+1}^i , gere S valores de $\hat{y}_{T+1}^1, \dots, \hat{y}_{T+1}^S$ a partir da densidade condicional estimada para o instante $T + 1$:

$$y_{T+1}^s \sim p_y(y_{T+1} | \hat{f}_{T+1}^i; \hat{\theta}_T^i), \quad s = 1, \dots, S;$$
 - 3.2. Use $\hat{y}_{T+1}^1, \dots, \hat{y}_{T+1}^S$ e a equação de atualização de f_t para obter $\hat{f}_{T+2}^1, \dots, \hat{f}_{T+2}^S$ condicional a $\hat{\theta}_T^i$ e \hat{f}_{T+1}^i ;
 - 3.3. Para cada \hat{f}_{T+2}^s , $s = 1, \dots, S$, refaça os passos 3.1 e 3.2 para os períodos $T + 2, \dots, T + k$;
4. Use os $M \times S$ valores de $\hat{y}_t^{i,s}$, a cada instante $t = T + 1, \dots, T + k$, para calcular os intervalos de confiança de previsão nos percentis desejados.

B

Apêndice - Capítulo 3

B.1

Resíduos quantílicos

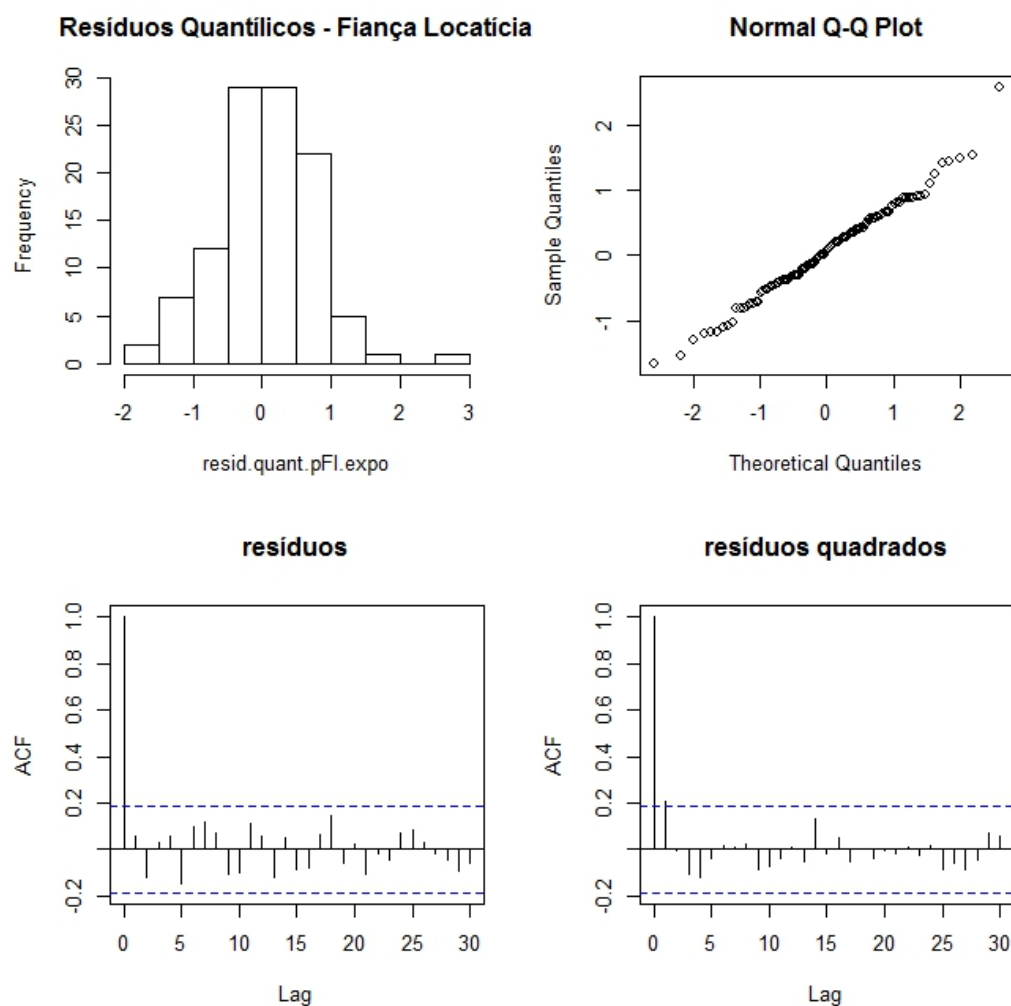


Figura B.1: Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Poisson(1,1) para o seguro fiança locatícia no período de 01/2006 a 12/2014, com variáveis exógenas IPCA e PIB.

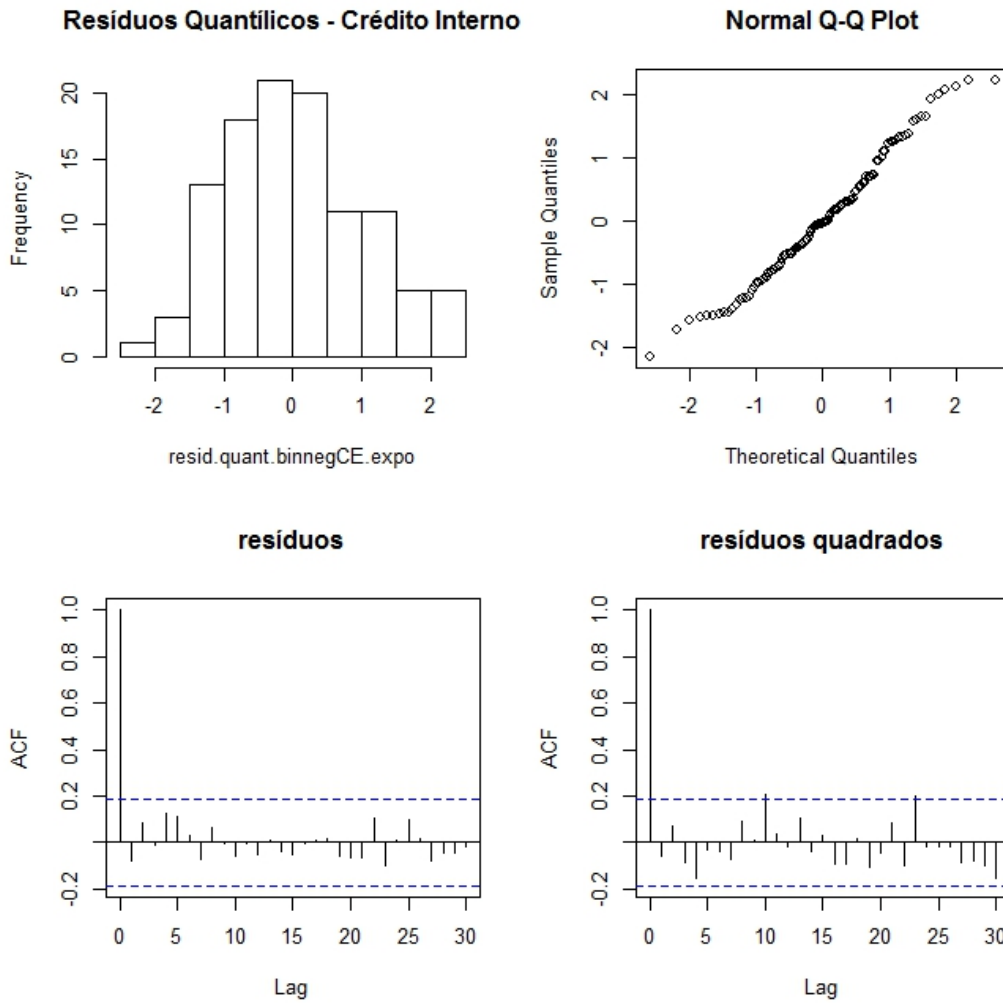


Figura B.2: Histograma, qq-plot e gráficos da ACF dos resíduos quantílicos do modelo GAS Binomial Negativa(1,1) para o seguro de credito interno no período de 01/2006 a 12/2014, com variáveis exógenas PIB e câmbio.

B.2

Cópias bivariadas

Apresentamos, a seguir, as formulações das cópias bivarias adotadas nesta tese.

Cópula Normal

$$C_R(u, v) = \Phi_R(\Phi^{-1}(u), \Phi^{-1}(v)),$$

onde $u, v \in [0, 1]^2$, $R \in [-1, 1]^2$ é a matriz de correlação, Φ é a inversa da função de distribuição acumulada da normal padrão e Φ_R é a função de distribuição acumulada conjunta da normal multivariada com vetor de média 0 e matriz de covariância igual a R .

Cópula t -Student

$$T_{\rho,\nu}(u, v) = \mathbf{t}_{\rho,\nu}(t_{\nu}^{-1}(u), t_{\nu}^{-1}(v)),$$

onde $u, v \in [0, 1]^2$ e t_{ν} é a função de distribuição t -Student univariada com ν graus de liberdade e média nula.

Cópula Frank

Família de cópula apresentada pela primeira vez em Frank [98] num contexto não estatístico. Algumas das propriedades estatísticas dessa família foram discutidas em Nelsen [99] e Genest [100].

$$C_{\theta}(u, v) = -\frac{1}{\theta} \log \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right],$$

sendo $u, v \in (0, 1)$ e $\theta \in \mathbb{R} \setminus \{0\}$.

Cópula Gumbel

Família de cópulas discutida pela primeira vez por Gumbel [101], também referenciada como família Gumbel-Hougaard.

$$C_{\theta}(u, v) = \exp \left[- \left((-\log(u))^{\theta} + (-\log(v))^{\theta} \right)^{\frac{1}{\theta}} \right],$$

sendo $u, v \in (0, 1)$ e $\theta \in [1, \infty)$.

Cópula Joe

Família de cópulas apresentada por Joe [58, 102].

$$C_{\theta}(u, v) = 1 - \left[(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta} (1-v)^{\theta} \right]^{\frac{1}{\theta}},$$

sendo $u, v \in (0, 1)$ and $\theta \in [1, \infty)$.

Cópula Clayton

Família de cópulas discutida primeiramente por Clayton [103]. Genest & Rivest [104] adotaram a nomenclatura de família Clayton.

$$C_{\theta}(u, v) = \left[\max(u^{-\theta} + v^{-\theta} - 1, 0) \right]^{-\frac{1}{\theta}},$$

sendo $u, v \in (0, 1)$ e $\theta \in [-1, \infty) \setminus \{0\}$.

B.3

Modelo t-GAS

Para modelar a dependência de forma dinâmica entre as séries trabalhadas no capítulo 3, foi adotado o modelo t -GAS com distribuição multivariada t -Student, com $k = 2$, e dinâmica de atualização GAS, proposto por Creal et al. [82].

Seja $y_t \in \mathbb{R}^k$ um vetor de observações com distribuição multivariada t -Student padronizada com ν graus de liberdade e matriz de covariância igual a Σ_t . Por simplicidade de notação, a média μ_t foi tomada como nula. A função de densidade de y_t é dada por

$$p(y_t | \Sigma_t; \nu) = \frac{\Gamma((\nu + k)/2)}{\Gamma(\nu/2)[(\nu - 2)\pi]^{k/2} |\Sigma_t|^{1/2}} \times \left[1 + \frac{y_t' \Sigma_t^{-1} y_t}{(\nu - 2)} \right]^{-\frac{(\nu + k)}{2}}, \quad (\text{B-1})$$

assumindo $\nu > 2$, de forma que a matriz de covariância exista.

Creal et al. [82] apresentam a formulação do *score* para atualização dos elementos da matriz de covariância Σ_t variante no tempo para a distribuição t -Student(B-1) ¹:

$$\nabla_t = \frac{\partial \ln p(y_t | \Sigma_t; \nu)}{\partial f_t} = \frac{1}{2} \Psi_t' \mathcal{D}_k' \Sigma_{t\otimes}^{-1} [\omega_t y_{t\otimes} - \text{vec}(\Sigma_t)],$$

$$I_{t|t-1} = E_{t-1} [\nabla_t \nabla_t'] = \frac{1}{4} \Psi_t' \mathcal{D}_k' J_{t\otimes}' [gG - \text{vec}(I) \text{vec}(I)'] J_{t\otimes} \mathcal{D}_k \Psi_t,$$

sendo $\Psi_t = \partial \text{vech}(\Sigma_t) / \partial f_t'$, $\Sigma_t = \Sigma(f_t)$, \mathcal{D}_k uma matriz de duplicação [$\mathcal{D}_k \text{vech}(A) = \text{vec}(A)$], escalar $\omega_t = (\nu + k) / (\nu - 2 + y_t' \Sigma_t^{-1} y_t)$, matriz J_t definida implicitamente por $\Sigma_t^{-1} = J_t' J_t$ e escalar $g = (\nu + k) / (\nu + 2 + k)$. Os elementos da matriz G , de dimensão $k \times k$, são dados por

$$G[(i - 1).k + l, (j - 1).k + m] = \delta_{ij} \delta_{lm} + \delta_{il} \delta_{jm} + \delta_{im} \delta_{jl},$$

para $i, j, m = 1, \dots, k$ e assumindo δ_{ij} igual a 1 se $i = j$ e zero caso contrário.

B.4

Algoritmo para previsão no modelo GAS com cópulas

Algoritmo para obtenção de intervalo de confiança para previsão no modelo GAS, seguindo metodologia de múltipla filtragem proposta por Blasques

¹Notação matricial: o produto de Kronecker é denotado por $A \otimes B$ para duas matrizes A e B . No caso em que $B = A$, temos $A_{\otimes} = A \otimes A$. O operador $\text{vec}(A)$ vetoriza a matriz A em um vetor coluna, enquanto que $\text{vech}(A)$ vetoriza a parte triangular inferior da matriz A em um vetor coluna.

et al. [50] e incorporando a dependência entre as séries nas previsões.

1. Utilize o método *bootstrap* na série bivariada de pseudo-observações (u_t, v_t) , $t = 1, \dots, T$, com B reamostragens, e estime para cada série bivariada o parâmetro θ_{cop} da cópula, obtendo, então, $\hat{\theta}_{cop}^1, \dots, \hat{\theta}_{cop}^B$;
2. Gere M valores $\hat{\theta}_{j,T}^i$, $j = 1$ ou 2 (representando cada série) e $i = 1, \dots, M$, a partir da sua respectiva distribuição assintótica

$$\hat{\theta}_{j,T}^i \sim N(\hat{\theta}_{j,T}, T^{-1}\hat{W}_j), \quad j = 1 \text{ ou } 2, i = 1, \dots, M;$$

3. Para cada $\hat{\theta}_{j,T}^i$, use os dados observados $y_{j,1}, \dots, y_{j,T}$ e a equação de atualização de $f_{j,t}$ e rode o filtro de $t = 1, \dots, T$ para obter $\hat{f}_{j,T+1}^i$, $i = 1, \dots, M$;
4. Para cada par $(\hat{\theta}_{j,T}^i, \hat{f}_{j,T+1}^i)$, $j = 1, 2$ e $i = 1, \dots, M$, siga os passos 4.1 a 4.3 para obter $M \times S$ valores de $\hat{f}_{j,T+k}^{i,s}$ e $\hat{y}_{j,T+k}^{i,s}$:
 - 4.1. Dados $\hat{\theta}_{j,T}^i$ e $\hat{f}_{j,T+1}^i$, gere S pares $(\hat{u}_{j,T+1}^s, \hat{v}_{j,T+1}^s)$, $s = 1, \dots, S$, a partir da cópula com parâmetro estimado $\hat{\theta}_{cop}^l$ (sendo l sorteado entre 1 e B) e, então, obtenha as S previsões de cada série $\hat{y}_{j,T+1}^{i,s}$, $j = 1, 2$, por meio da função quantil das distribuições marginais condicionais;
 - 4.2. Use $\hat{y}_{j,T+1}^1, \dots, \hat{y}_{j,T+1}^S$ e a equação de atualização de f_t para obter $\hat{f}_{j,T+2}^1, \dots, \hat{f}_{j,T+2}^S$ condicional a $\hat{\theta}_{j,T}^i$ e $\hat{f}_{j,T+1}^i$;
 - 4.3. Para cada $\hat{f}_{j,T+2}^s$, $j = 1, 2$ e $s = 1, \dots, S$, refaça os passos 4.1 e 4.2 para os períodos $T + 2, \dots, T + k$;
5. Use os $M \times S$ valores de $\hat{y}_{j,t}^{i,s}$ de cada série, a cada instante $t = T + 1, \dots, T + k$, para calcular os intervalos de confiança de previsão nos percentis desejados.