



Daiane Rodrigues dos Santos

Estimação de Provisões IBNR (Incurred But Not Reported) em Mercado de Seguros via Modelos com Coeficientes Variantes no Tempo

Tese de Doutorado

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

Orientador: Prof. Cristiano Augusto Coelho Fernandes

Co-orientador: Prof. Joel Maurício Correia da Rosa

Rio de Janeiro
Janeiro de 2017



Daiane Rodrigues dos Santos

Estimação de Provisões IBNR (Incurred But Not Reported) em Mercado de Seguros via Modelos com Coeficientes Variantes no Tempo

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

Prof. Cristiano Augusto Coelho Fernandes

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Joel Maurício Corrêa da Rosa

Co-Orientador

Rockefeller University

Prof. Eduardo Fraga Lima de Melo

SUSEP

Prof.^a. Jessica Quintanilha Kubrusly

UFF

Dr. Rodrigo Simões Atherino

JGP

Prof. Luciano Vereda Oliveira

UFF

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro

Técnico Científico

Rio de Janeiro, 09 de Janeiro de 2017

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

Daiane Rodrigues dos Santos

Graduou-se em Ciências Econômicas pela Universidade do Estado do Rio de Janeiro - UERJ em 2008. Fez mestrado no Departamento de Economia da Universidade Federal do Espírito Santo - UFES em 2011, com ênfase em macroeconomia. É professora substituta da Universidade Federal Rural do Rio de Janeiro - UFRRJ e trabalhou como economista na Fundação Centro de Estudos do Comércio Exterior - FUNCEX.

Ficha Catalográfica

Santos, Daiane Rodrigues dos

Estimação de provisões IBNR (Incurred But Not Reported) em mercado de seguros via modelos com coeficientes variantes no tempo / Daiane Rodrigues dos Santos; orientador: Cristiano Augusto Coelho Fernandes; co-orientador: Joel Maurício Correia da Rosa. – 2017.

164 f. : il. color. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2017.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. IBNR. 3. Modelos GAS. 4. Modelos em Espaço de Estado. 5. Modelo Atherino Multivariado. I. Fernandes, Cristiano Augusto Coelho. II. Rosa, Joel Maurício Correia da. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Agradecimentos

Em primeiro lugar, agradeço a Jesus Cristo pelo seu infinito amor e pela sua infinita misericórdia. Em Jericó, Jesus olhou para Zaqueu e o escolheu em meio à multidão, transformando seu coração e, conseqüentemente, sua vida ao chamá-lo pelo nome. Jesus igualmente fez comigo.

Eu gostaria de agradecer, em especial, ao meu esposo Tiago por ter me apoiado nesta difícil e, à primeira vista, impossível jornada. Devido ao seu amor, companheirismo, amizade, paciência e compreensão. Ao seu apoio e ao seu suporte no LaTeX e no R, este trabalho pôde ser concretizado. Obrigada por ter feito do meu sonho o nosso sonho!

Gostaria de agradecer a minha amada e gentil mãe Marineth e ao meu honrado e querido irmão Ulisses. Em toda a minha vida, eles estiveram presentes, me apoiando e amando incondicionalmente.

Em particular, quero manifestar o meu apreço ao professor Joel Rosa que, além de coorientador, nesses últimos anos tem sido um grande amigo. Eu não tenho palavras para expressar minha sincera gratidão a ele. Agradecimentos especiais ao meu orientador, professor Cristiano Fernandes, por ter me guiado durante esta caminhada com seu conhecimento.

Expresso minha gratidão ao estimado professor Alvaro Veiga por ter acreditado na minha competência e ter sido peça fundamental da minha qualificação.

Sou grata a todos os meus amigos, principalmente, a Soraida Aguilar (minha colombiana amada e querida) e a Frances Blank que, muitas das vezes, seguraram minha mão, apoiaram-me e não me deixaram fracassar.

Agradeço aos amigos Rafael Cunha, Hedmilton Cardoso, Rodrigo Branco, Gilson Matos e Manoel Pereira pelo carinho dado a mim nesse logo caminho. Agradeço a Karla, Vanessa, Carol, Barbara e Marcinha pelo apoio incondicional e pelas horas de descontração.

Agradeço de forma carinhosa a minha cunhada Vivi pelo apoio constante a mim e a minha família.

Agradeço a minha cunhada Millene e a querida amiga Adilene pelas correções textuais.

Agradeço as amigas e secretárias do departamento, Márcia Guerra, Adriane Senger e Alcina Pontes, por sempre me receberem com tanto carinho e alegria.

Também sou grata aos membros da minha banca examinadora, Dr. Eduardo Fraga Lima de Melo, Dra. Jessica Quintanilha Kubrusly, Dr. Luciano Vereda Oliveira e Dr. Rodrigo Simões Atherino, que comprometeram muito generosamente seu tempo e conhecimento para avaliar a minha tese.

Agradeço a minha paróquia Sagrado Coração de Jesus, em especial, aos meus aluninhos que sempre me receberam com muito amor e carinho na sala de aula. Muitas vezes, chegava aos nossos encontros semanais pensando que meus problemas (relacionados ao Doutorado) eram os maiores da face da terra e ao conversar com eles percebia que eu não tinha do que reclamar. Eu não tinha, de fato, problemas. Tudo poderia ser resolvido com mais horas de estudo.

Em memória, o amigo Leonardo Chieppe, por me doar um pouco do seu raro tempo aqui na terra.

Gostaria de agradecer ao CNPq pelo apoio financeiro e a PUC-Rio pela bolsa de isenção de mensalidades do doutorado.

Há muito mais a quem agradecer... A todos aqueles que, embora não nomeados, me presentearam com seus inestimáveis apoios em distintos momentos e por suas presenças afetivas, o meu reconhecido e carinhoso muito obrigada!

Resumo

Santos, Daiane Rodrigues dos; Fernandes, Cristiano Augusto Coelho (Orientador); Rosa, Joel Maurício Correia da (Co-orientador). **Estimação de Provisões IBNR (Incurred But Not Reported) em Mercado de Seguros via Modelos com Coeficientes Variantes no Tempo**. Rio de Janeiro, 2017. 164p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese apresenta duas contribuições para a modelagem e previsão de sinistros já ocorridos e ainda não avisados (*Incurred But Not Reported* - IBNR), quando organizados numa estrutura de dados conhecida como triângulo de *run-off*. Ambas as contribuições são baseadas em arcabouços gerais para a construção de modelos para séries temporais com coeficientes variantes no tempo. Em nossa primeira contribuição desenvolvemos a extensão multivariada do modelo em espaço de estado proposto por Atherino em 2008. A partir dessa extensão é possível modelar simultaneamente um ou mais triângulos de *run-off* associados às diversas coberturas de uma seguradora, levando-se em consideração a dependência entre os distintos triângulos, capturada pela estrutura da matriz variância – covariância do modelo SUTSE e a dependência entre as células de cada triângulo de *run-off*, capturada pelas componentes de nível e de periodicidade, de acordo com a proposta de Atherino et al. (2010). Em nossa segunda contribuição desenvolvemos um arcabouço geral para a modelagem univariada de triângulos de *run-off* a partir da estrutura dos modelos GAS (*Generalized Autoregressive Score*) desenvolvidos por Creal et al. (2013). Esse arcabouço, bastante flexível, permite a escolha de qualquer distribuição para as entradas do triângulo *run-off*, considerando que os seus parâmetros variem ao longo do período de origem ou de desenvolvimento. Em particular consideramos as distribuições gama e log-normal. Nossos resultados foram comparados com os obtidos através do método *chain ladder* (Mack, 1993), utilizado como *benchmark* na indústria de seguros. O teste Diebold e Mariano (1995) evidenciou que os modelos propostos geram melhores previsões, comparadas as previsões do método *chain ladder*.

Palavras-chave

IBNR; Modelos GAS; Modelos em Espaço de Estado; Modelo Atherino Multivariado; CASCO; RCFV.

Abstract

Santos, Daiane Rodrigues dos; Fernandes, Cristiano Augusto Coelho (Advisor); Rosa, Joel Maurício Correia da (Co-advisor). **Estimation of IBNR (Incurred But Not Reported) provisions in insurance via models whit time-varying coefficients**. Rio de Janeiro, 2017. 164p. PhD Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis presents two contributions to the modeling and prediction of a type of claims in the insurance industry known as IBNR (Incurred But Not Reported) when these are organized in a data structure known as the run-off triangle. Both contributions are based on general frameworks for building models for time series with time varying coefficients. In our first contribution we developed the multivariate extension of the state space model proposed by Atherino in 2008. From this extension it is possible to model simultaneously one or more run-off triangles associated with different coverages from an insurer, taking into account the dependence between different triangles, captured by the structure of the variance-covariance matrix of the SUTSE model, while the dependence between the cells of each triangle run-off is captured by the components of level and periodicity according to the model proposed by Atherino (2008). In our second contribution we developed a general framework for univariate modeling of run-off triangles using the structure of GAS models (Generalized Autoregressive Score) developed by Creal et al. (2013). This framework, very flexible, allows one to choose any distribution to the inputs of the triangle run-off, considering that its parameters can vary over the period of origin or period of development. In particular we have considered both gamma and log-normal distributions. Our results were compared with those obtained by the chain ladder method (Mack, 1993) used as a benchmark in the insurance industry. The Diebold and Mariano test (1995) showed that the proposed models produced better predictions compared to the predictions of the chain ladder method.

Keywords

IBNR; GAS Models; State Space Models; Multivariate Atherino Model; CASCO; RCFV.

Sumário

1	Introdução	19
2	Estimação de reservas IBNR via triângulo de run-off	22
2.1	Sinistros no ramo de seguros	22
2.2	SUSEP - Órgão regulador de seguros no Brasil	23
2.3	Formato dos dados de sinistros ocorridos e não avisados (IBNR)	23
2.3.1	Triângulo de run-off	23
2.3.2	Representação incremental do triângulo de run-off	24
2.3.3	Representação acumulada do triângulo de run-off	25
2.3.4	Empilhamento dos dados contidos no triângulo de run-off	25
2.3.5	Organização pela diagonal do triângulo de run-off	27
2.4	Tipos de reservas calculadas	28
2.5	Adição do efeito linha ao triângulo de run-off	29
2.6	Metodologia Univariada para o Cálculo das Reservas IBNR proposta por Mack (1993) - Chain Ladder	30
2.6.1	Chain Ladder - Notações e resultados básicos	30
2.6.2	Erro médio quadrático obtido por Mack	31
2.7	Metodologias multivariadas para o cálculo das reservas IBNR	32
2.8	Modelos em espaço de estado e reservas IBNR	33
3	Modelo Atherino bivariado	35
3.1	Uso da distribuição log-normal para a série composta pelos dados contidos no triângulo de run-off	36
3.2	Modelo Atherino univariado	36
3.2.1	Forma em espaço de estado linear e Gaussiana	38
3.2.2	Matriz de covariância condicional	39
3.2.3	Cálculo da reserva IBNR quando os dados contidos no triângulo de run-off estão expostos na forma empilhada	40
3.2.4	Reserva IBNR estimada e o Erro Médio Quadrático: Casos total e por ano	41
3.3	Modelo Atherino bivariado	41
3.3.1	Formulação SUTSE	42
3.3.2	Formulação SUTSE - nível comum	47
3.4	Análise do poder preditivo - medidas de aderência	54
4	Modelos GAS para triângulo de run-off	55
4.1	Modelos GAS - especificações e propriedades	55
4.1.1	Inicialização do vetor de parâmetros variantes no tempo	57
4.1.2	Algoritmos genéticos para geração de valores iniciais	58
4.1.3	Diagnósticos	59
4.2	Modelos GAS - adaptação para série em painel	60
4.2.1	Especificações dos modelos GAS (1,1) para dados em painel	60
4.2.2	Estimação por Máxima Verossimilhança - adaptada ao triângulo de run-off	61

4.2.3	Intervalo de confiança in sample para os parâmetros estáticos do modelo GAS(1,1)	62
4.2.3.1	Simulation-based bands	63
4.2.4	Previsão fora da amostra - efeito linha no triângulo de run-off	63
4.3	Modelos GAS e densidades condicionais	64
4.3.1	Densidade gama adaptada ao triângulo de run-off	65
4.3.1.1	Dinâmica de atualização - distribuição gama	65
4.3.1.2	Cálculo do IBNR assumindo distribuição gama	66
4.3.1.3	Cálculo do coeficiente de variação para a distribuição gama	67
4.3.1.4	Cálculo do intervalo de confiança para a distribuição gama	68
4.3.2	Densidade log-normal adaptada ao triângulo de run-off	69
4.3.2.1	Dinâmica de atualização - distribuição log-normal	69
4.3.2.2	Cálculo do IBNR assumindo distribuição log-normal	70
4.3.2.3	Cálculo do coeficiente de variação para a distribuição log-normal	70
4.3.2.4	Cálculo do intervalo de confiança para a distribuição log-normal	71
5	Aplicações	72
5.1	Resultados do modelo Atherino bivariado (formulação SUTSE)	77
5.2	Resultados do modelo GAS (1,1)	84
5.2.1	Cobertura Casco	85
5.2.1.1	Modelo GAS (1,1) com distribuição gama	85
5.2.1.2	GAS (1,1) com distribuição gama - reserva IBNR e intervalo de confiança, cobertura Casco	90
5.2.1.3	Modelo GAS (1,1) com distribuição log-normal	92
5.2.1.4	GAS (1,1) com distribuição log-normal - reserva IBNR e intervalos de confiança	97
5.2.2	Previsão fora do triângulo (efeito linha ou reservas de prêmio) - Cobertura Casco	99
5.2.3	Cobertura RCFV	101
5.2.3.1	Modelo GAS (1,1) com distribuição gama	101
5.2.3.2	GAS (1,1) com distribuição gama - reserva IBNR e intervalo de confiança	106
5.2.3.3	Cobertura RCFV - resultados do modelo GAS (1,1) com distribuição log-normal	107
5.2.3.4	GAS (1,1) com distribuição log-normal - reserva IBNR e intervalo de confiança	112
5.2.4	Previsão fora da amostra (efeito linha no triângulo de run-off)	114
5.3	Medidas de aderências - coberturas Casco e RCFV	116
5.4	Reservas IBNR estimadas - coberturas Casco e RCFV	118
5.5	Teste Diebold e Mariano dentro da amostra - coberturas Casco e RCFV	121
6	Conclusão	123
A	Abordagem univariada para caso multivariado	127
A.1	Matriz H_t não diagonal - Modelo Atherino bivariado	128
A.2	Filtragem e suavização - MEE multivariado na forma univariada	129
A.3	Missing values - MEE multivariado na forma univariada	130
A.4	Inicialização difusa exata - MEE multivariado na forma univariada	130
A.5	Verossimilhança exata - MEE multivariado na forma univariada	132

A.6	Algoritmo EM	133
A.7	MEE - diagnósticos e medidas de aderência	134
A.8	Análise dos resíduos	134
B	Teste de homogeneidade da variância	136
C	Teste Diebold e Mariano	137
D	Resultado do modelo Atherino univariado	139
E	Modelo GAS simplificado - cobertura Casco	145
F	Modelo GAS simplificado - cobertura RCFV	147
G	Cobertura Casco - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição gama)	149
H	Cobertura Casco - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição log-normal)	153
I	Cobertura RCFV - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição gama)	157
J	Cobertura RCFV - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição log-normal)	161

Lista de figuras

4.1	Algoritmo para a geração de valores iniciais.	59
5.1	Participações dos segmentos de seguros no Brasil. Fonte: SUSEP.	72
5.2	Cobertura Casco - série empilhada de sinistros ocorridos e avisados e seus logaritmos.	76
5.3	Cobertura RCFV - série empilhada de sinistros ocorridos e avisados e seus logaritmos.	76
5.4	Coberturas Casco e RCFV - logaritmo das séries IBNR e as séries suavizadas.	79
5.5	Cobertura Casco - logaritmo da série IBNR, o nível e a periodicidade suavizadas (modelo Atherino bivariado - formulação SUTSE).	80
5.6	Cobertura RCFV - logaritmo da série IBNR, o nível e a periodicidade suavizadas (modelo Atherino bivariado - formulação SUTSE).	80
5.7	Casco - inovações padronizadas e resíduos suavizados (modelo Atherino bivariado - formulação SUTSE).	81
5.8	RCFV - inovações padronizadas e resíduos suavizados (modelo Atherino bivariado - formulação SUTSE).	81
5.9	Coberturas Casco e RCFV - função de autocorrelação e função de autocorrelação parcial.	82
5.10	Coberturas Casco e RCFV - Q-Q plotes das inovações padronizadas, modelo Atherino bivariado.	83
5.11	Casco - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição gama.	88
5.12	Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição gama.	88
5.13	Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.	89
5.14	Casco - Q-Q plotes dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.	89
5.15	Cobertura Casco - reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição gama (histograma).	92
5.16	Casco - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição log-normal.	95
5.17	Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.	95
5.18	Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.	96
5.19	Casco - Q-Q plotes dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.	96
5.20	Cobertura Casco - reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição log-normal (histograma).	99
5.21	RCFV - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição gama.	103

5.22	RCFV - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição gama.	104
5.23	RCFV - função de autocorrelação e função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.	104
5.24	RCFV - Q-Q plotes dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.	105
5.25	Cobertura RCFV - reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição gama (histograma).	107
5.26	RCFV - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição log-normal.	110
5.27	RCFV - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.	110
5.28	RCFV - função de autocorrelação função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.	111
5.29	RCFV - Q-Q plotes dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.	111
5.30	Cobertura RCFV - Reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição log-normal (histograma).	114
D.1	O logaritmo da série e a série suavizada - coberturas Casco e RCFV.	139
D.2	O logaritmo da série IBNR, o nível e a periodicidade suavizadas - coberturas Casco e RCFV.	140
D.3	Funções de autocorrelação das inovações padronizadas e das inovações padronizadas ao quadrado - Casco.	142
D.4	Funções de autocorrelação das inovações padronizadas e das inovações padronizadas ao quadrado - RCFV.	142
D.5	Coberturas Casco e RCFV - Q-Q plotes das inovações padronizadas.	143
G.1	GAS (1,1) gama - IC simulado para t=2 (2 ^o trim.09), Casco.	149
G.2	GAS (1,1) gama - IC simulado para t=3 (3 ^o trim.09), Casco.	149
G.3	GAS (1,1) gama - IC simulado para t=4 (4 ^o trim.09), Casco.	149
G.4	GAS (1,1) gama - IC simulado para t=5 (1 ^o trim.10), Casco.	149
G.5	GAS (1,1) gama - IC simulado para t=6 (2 ^o trim.10), Casco.	150
G.6	GAS (1,1) gama - IC simulado para t=7 (3 ^o trim.10), Casco.	150
G.7	GAS (1,1) gama - IC simulado para t=8 (4 ^o trim.10), Casco.	150
G.8	GAS (1,1) gama - IC simulado para t=9 (1 ^o trim.11), Casco.	150
G.9	GAS (1,1) gama - IC simulado para t=10 (2 ^o trim.11), Casco.	150
G.10	GAS (1,1) gama - IC simulado para t=11 (3 ^o trim.11), Casco.	150
G.11	GAS (1,1) gama - IC simulado para t=12 (4 ^o trim.11), Casco.	151
G.12	GAS (1,1) gama - IC simulado para t=13 (1 ^o trim.12), Casco.	151
G.13	GAS (1,1) gama - IC simulado para t=14 (2 ^o trim.12), Casco.	151
G.14	GAS (1,1) gama - IC simulado para t=15 (3 ^o trim.12), Casco.	151
G.15	GAS (1,1) gama - IC simulado para t=16 (4 ^o trim.12), Casco.	151
G.16	GAS (1,1) gama - IC simulado para t=17 (1 ^o trim.13), Casco.	151
G.17	GAS (1,1) gama - IC simulado para t=18 (2 ^o trim.13), Casco.	152
H.1	GAS (1,1) Log-normal - IC simulado para t=2 (2 ^o trim.09), Casco.	153

H.2	GAS (1,1) Log-normal - IC simulado para t=3 (3 ^o trim.09), Casco.	153
H.3	GAS (1,1) Log-normal - IC simulado para t=4 (4 ^o trim.09), Casco.	153
H.4	GAS (1,1) Log-normal - IC simulado para t=5 (1 ^o trim.10), Casco.	153
H.5	GAS (1,1) Log-normal - IC simulado para t=6 (2 ^o trim.10), Casco.	154
H.6	GAS (1,1) Log-normal - IC simulado para t=7 (3 ^o trim.10), Casco.	154
H.7	GAS (1,1) Log-normal - IC simulado para t=8 (4 ^o trim.10), Casco.	154
H.8	GAS (1,1) Log-normal - IC simulado para t=9 (1 ^o trim.11), Casco.	154
H.9	GAS (1,1) Log-normal - IC simulado para t=10 (2 ^o trim.11), Casco.	154
H.10	GAS (1,1) Log-normal - IC simulado para t=11 (3 ^o trim.11), Casco.	154
H.11	GAS (1,1) Log-normal - IC simulado para t=12 (4 ^o trim.11), Casco.	155
H.12	GAS (1,1) Log-normal - IC simulado para t=13 (1 ^o trim.12), Casco.	155
H.13	GAS (1,1) Log-normal - IC simulado para t=14 (2 ^o trim.12), Casco.	155
H.14	GAS (1,1) Log-normal - IC simulado para t=15 (3 ^o trim.12), Casco.	155
H.15	GAS (1,1) Log-normal - IC simulado para t=16 (4 ^o trim.12), Casco.	155
H.16	GAS (1,1) Log-normal - IC simulado para t=17 (1 ^o trim.13), Casco.	155
H.17	GAS (1,1) Log-normal - IC simulado para t=18 (2 ^o trim.13), Casco.	156
I.1	GAS (1,1) gama - IC simulado para t=2 (2 ^o trim.09), RCFV.	157
I.2	GAS (1,1) gama - IC simulado para t=3 (3 ^o trim.09), RCFV.	157
I.3	GAS (1,1) gama - IC simulado para t=4 (4 ^o trim.09), RCFV.	157
I.4	GAS (1,1) gama - IC simulado para t=5 (1 ^o trim.10), RCFV.	157
I.5	GAS (1,1) gama - IC simulado para t=6 (2 ^o trim.10), RCFV.	158
I.6	GAS (1,1) gama - IC simulado para t=7 (3 ^o trim.10), RCFV.	158
I.7	GAS (1,1) gama - IC simulado para t=8 (4 ^o trim.10), RCFV.	158
I.8	GAS (1,1) gama - IC simulado para t=9 (1 ^o trim.11), RCFV.	158
I.9	GAS (1,1) gama - IC simulado para t=10 (2 ^o trim.11), RCFV.	158
I.10	GAS (1,1) gama - IC simulado para t=11 (3 ^o trim.11), RCFV.	158
I.11	GAS (1,1) gama - IC simulado para t=12 (4 ^o trim.11), RCFV.	159
I.12	GAS (1,1) gama - IC simulado para t=13 (1 ^o trim.12), RCFV.	159
I.13	GAS (1,1) gama - IC simulado para t=14 (2 ^o trim.12), RCFV.	159
I.14	GAS (1,1) gama - IC simulado para t=15 (3 ^o trim.12), RCFV.	159
I.15	GAS (1,1) gama - IC simulado para t=16 (4 ^o trim.12), RCFV.	159
I.16	GAS (1,1) gama - IC simulado para t=17 (1 ^o trim.13), RCFV.	159
I.17	GAS (1,1) gama - IC simulado para t=18 (2 ^o trim.13), RCFV.	160
J.1	GAS (1,1) log-normal - IC simulado para t=2 (2 ^o trim.09), RCFV.	161
J.2	GAS (1,1) log-normal - IC simulado para t=3 (3 ^o trim.09), RCFV.	161
J.3	GAS (1,1) log-normal - IC simulado para t=4 (4 ^o trim.09), RCFV.	161
J.4	GAS (1,1) log-normal - IC simulado para t=5 (1 ^o trim.10), RCFV.	161
J.5	GAS (1,1) log-normal - IC simulado para t=6 (2 ^o trim.10), RCFV.	162
J.6	GAS (1,1) log-normal - IC simulado para t=7 (3 ^o trim.10), RCFV.	162
J.7	GAS (1,1) log-normal - IC simulado para t=8 (4 ^o trim.10), RCFV.	162
J.8	GAS (1,1) log-normal - IC simulado para t=9 (1 ^o trim.11), RCFV.	162
J.9	GAS (1,1) log-normal - IC simulado para t=10 (2 ^o trim.11), RCFV.	162
J.10	GAS (1,1) log-normal - IC simulado para t=11 (3 ^o trim.11), RCFV.	162
J.11	GAS (1,1) log-normal - IC simulado para t=12 (4 ^o trim.11), RCFV.	163
J.12	GAS (1,1) log-normal - IC simulado para t=13 (1 ^o trim.12), RCFV.	163
J.13	GAS (1,1) log-normal - IC simulado para t=14 (2 ^o trim.12), RCFV.	163
J.14	GAS (1,1) log-normal - IC simulado para t=15 (3 ^o trim.12), RCFV.	163

- J.15 GAS (1,1) log-normal - IC simulado para $t=16$ (*4^otrim.12*), RCFV.163
- J.16 GAS (1,1) log-normal - IC simulado para $t=17$ (*1^otrim.13*), RCFV.163
- J.17 GAS (1,1) log-normal - IC simulado para $t=18$ (*2^otrim.13*), RCFV.164

Lista de tabelas

2.1	Triângulo de <i>run-off</i> - representação incremental.	25
2.2	Triângulo de <i>run-off</i> - representação acumulada.	25
2.3	Reparametrização “por linhas” do triângulo de <i>run-off</i> .	26
2.4	Reparametrização “pela diagonal” do triângulo de <i>run-off</i> .	28
2.5	Triângulo de <i>run-off</i> - representação incremental com efeito linha.	29
5.1	Triângulo contendo os dados trimestrais dos sinistros ocorridos e avisados da cobertura Casco, em R\$ mil.	74
5.2	Triângulo contendo os dados trimestrais dos sinistros ocorridos e avisados da cobertura RCFV, em R\$ mil.	74
5.3	Estatísticas descritivas dos montantes ocorridos e avisados - coberturas Casco e RCFV e seus logaritmos.	77
5.4	Parâmetros estimados - modelo Atherino bivariado (formulação SUTSE).	79
5.5	Testes e diagnósticos aplicados às inovações padronizadas para o modelo Atherino bivariado.	83
5.6	GAS (1,1) com distribuição gama - estimativas dos parâmetros variantes no tempo (cobertura Casco).	85
5.7	Cobertura Casco - estimativas dos parâmetros estáticos (vetor ψ). Modelo GAS (1,1) com distribuição gama.	86
5.8	Cobertura Casco - resíduos quantílicos, modelo GAS (1,1) com distribuição gama.	86
5.9	Cobertura Casco - resíduos de Pearson, modelo GAS (1,1) com distribuição gama.	87
5.10	Cobertura Casco - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição gama.	90
5.11	Cobertura Casco - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição gama.	91
5.12	GAS (1,1) com distribuição log-normal - estimativas dos parâmetros variantes no tempo (cobertura Casco).	93
5.13	Cobertura Casco - estimativas dos parâmetros estáticos (vetor ψ). Modelo GAS (1,1) com distribuição log-normal.	93
5.14	Cobertura Casco - resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.	94
5.15	Cobertura Casco - resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.	94
5.16	Cobertura Casco - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição log-normal.	97
5.17	Cobertura Casco - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição log-normal.	98
5.18	Cobertura Casco - previsão fora da amostra (GAS (1,1) com distribuição gama).	100
5.19	Cobertura Casco - previsão fora da amostra (GAS (1,1) com distribuição log-normal).	100

5.20	GAS (1,1) com distribuição gama - estimativas dos parâmetros variantes no tempo (cobertura RCFV).	101
5.21	Cobertura RCFV - estimativa dos parâmetros estáticos (vetor ψ). Modelo GAS (1,1) com distribuição gama.	102
5.22	Cobertura RCFV - resíduos quantílicos, modelo GAS (1,1) com distribuição gama.	102
5.23	Cobertura RCFV - resíduos de Pearson, modelo GAS (1,1) com distribuição gama.	103
5.24	Cobertura RCFV - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição gama.	105
5.25	Cobertura RCFV - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição gama.	106
5.26	GAS (1,1) com distribuição log-normal - estimativas dos parâmetros variantes no tempo (cobertura RCFV).	108
5.27	Cobertura RCFV - estimativas dos parâmetros estáticos (vetor ψ) pelo modelo GAS (1,1) com distribuição log-normal.	108
5.28	Cobertura RCFV - resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.	109
5.29	Cobertura RCFV - resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.	109
5.30	Cobertura RCFV - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição log-normal.	112
5.31	Cobertura RCFV - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição log-normal.	113
5.32	Cobertura RCFV - previsão fora da amostra (GAS (1,1) com distribuição gama).	115
5.33	Cobertura RCFV - previsão fora da amostra (GAS (1,1) com distribuição log-normal).	116
5.34	Cobertura Casco - medidas de aderência (dentro da amostra).	117
5.35	Cobertura Casco - medidas de aderência (fora da amostra).	117
5.36	Cobertura RCFV - medidas de aderência (dentro da amostra).	117
5.37	Cobertura RCFV - medidas de aderência (fora da amostra).	118
5.38	Cobertura Casco - reservas IBNR estimadas.	118
5.39	Cobertura Casco - coeficientes de variação estimados.	119
5.40	Cobertura RCFV - reservas IBNR estimados.	120
5.41	Cobertura RCFV - coeficientes de variação estimados.	120
5.42	Cobertura Casco - teste Diebold e Mariano (p -valores).	121
5.43	Cobertura RCFV - teste Diebold e Mariano (p -valores).	122
6.1	Cobertura Casco - avaliação das métricas aplicadas.	124
6.2	Cobertura RCFV - avaliação das métricas aplicadas.	125
D.1	Parâmetros estimados e relações sinal/ruído para as séries Casco e RCFV - modelo Atherino univariado.	139
D.2	Cobertura Casco - inovações padronizadas e resíduos auxiliares (Atherino univariado).	141
D.3	Cobertura RCFV - inovações padronizadas e resíduos auxiliares (Atherino univariado).	141

D.4	Testes e diagnósticos aplicados à série dos resíduos (Modelo de Atherino univariado) - cobertura Casco.	143
D.5	Testes e diagnósticos aplicados à série dos resíduos (Modelo de Atherino univariado) - cobertura RCFV.	144
E.1	Cobertura Casco - parâmetros estimados (modelo simples com distribuição gama).	145
E.2	Cobertura Casco - medidas de aderência com distribuição gama.	145
E.3	Cobertura Casco - reservas IBNR e coeficiente de variação (Modelo simples com distribuição gama).	146
F.1	Cobertura RCFV - parâmetros estimados (Modelo simples com distribuição gama).	147
F.2	Cobertura RCFV - medidas de aderência com distribuição gama.	147
F.3	Cobertura RCFV - reservas IBNR e coeficiente de variação (Modelo simples com distribuição gama).	148

Lista de Abreviaturas

APP – Acidentes Pessoais a Passageiros
ARIMA – Autor-Regressivos Integrados de Médias Móveis
BFGS – Broyden Fletcher Goldfarb Shanno
CNSP – Conselho Nacional de Seguros Privados
DM - Diebold e Mariano
DPVAT – Danos Pessoais Causados por Veículos Automotores de Vias Terrestres
DSC – *Dynamic Conditional Score*
EM – *Expectation maximization*
EQM – Erro Quadrático Médio
FAC – Função de auto Correlação
FACP – Função de Auto Correlação Parcial
FIPE – Fundação Instituto de Pesquisas Econômicas
GAS – *Generalized Autoregressive Score*
IAIS – *International Association of Insurance Supervisors*
IBNR – *Incurred But Not Reported*
IC – intervalo de confiança
MAPE – *Mean Absolute Percentage Error*
MEE – Modelo em Espaço de Estado
MEM – Modelos de Erro Multiplicativo
MLG – Modelos Lineares Generalizados
MSE – *Mean squared error*
RCFV – Responsabilidade Civil Facultativa de Veículos
SUSEP – Superintendência de Seguros Privados
SUTSE – *Seemingly Unrelated Time Series Equations*
VAR – Vetores Auto-Regressivos
SARIMA – Sazonal Autor-Regressivos Integrados de Médias Móveis
VGBL – Vida Gerador de Benefício Livre
VMR – Valor de Mercado Referenciado

1 Introdução

Provisões IBNR - *Incurring But Not Reported* - são reservas financeiras que devem ser estabelecidas para a maioria dos segmentos em que as companhias de seguro atuam. Esta reserva, em determinada ocasião, deve ser equivalente ao montante necessário para cobrir as indenizações de todos os sinistros ocorridos, porém ainda não informados, pelos contratantes das apólices, à seguradora. Desta forma, o valor estimado da reserva IBNR direciona a seguradora quanto à magnitude do montante que a mesma deverá provisionar em um determinado período.

O cálculo de reservas IBNR desperta o interesse de pesquisadores do mundo inteiro há mais de quarenta anos, acarretando um volume de mais de 800 trabalhos publicados com abordagens paramétricas e não paramétricas para a obtenção dessas reservas (SCHMIDT, 2013).

Uma seguradora, ao realizar a previsão de reservas IBNR - *Incurring But Not Reported* - otimiza o retorno sobre o patrimônio líquido (Return on Equity - ROE), sem afetar sua solvência. A obrigatoriedade legal da provisão IBNR agrega, indubitavelmente, maior importância ao tema. No Brasil, a provisão para sinistros IBNR passou a ser obrigatória a partir do ano de 1998, através da Resolução CNSP nº 18, de 25 de agosto de 1998.

Esta tese tem como objetivo principal a formulação e a estimação de modelos para prever reservas IBNR, tanto na forma univariada quanto na forma bivariada, utilizando duas classes de modelos com parâmetros variantes no tempo. A primeira classe apresenta a extensão do modelo proposto por Atherino et al (2010). Nessa extensão levou-se em consideração a dependência entre as reservas IBNR apresentadas em distintos triângulos de *run-off* e se propôs uma abordagem bivariada, utilizando os dados no formato denominado “empilhamento” destes triângulos (Subseção 2.3.4).

Segundo De Jong (2011), a margem de risco geral, envolvendo mais de uma linha de negócio, precisa refletir a associação entre os triângulos de *run-off* que contêm dados de ramos de seguros distintos. De acordo com o autor, os pagamentos em diferentes triângulos “caminham” juntos. Contudo, se esses movimentos forem independentes, então a variância dos pagamentos totais é a soma das variâncias individuais para cada linha de negócio. No caso de

IBNR - positivamente e perfeitamente correlacionados (seguros perfeitamente complementares) - a margem de risco aumenta de acordo com o aumento do número de ramos de negócios.

Na segunda parte da tese, propomos um arcabouço geral para estimação de reservas IBNR, utilizando os modelos GAS (CREAL et. al, 2013). A partir desse arcabouço é possível especificar e estimar qualquer distribuição condicional que se julgue adequada para caracterizar as entradas do triângulo de *run-off*. Embora em nossa tese tenha sido proposta a dinâmica dos parâmetros da distribuição ao longo dos anos de origem, é possível também especificar modelos em que essa dinâmica ocorra ao longo dos anos de desenvolvimento.

Cabe ressaltar que os modelos GAS (*Generalized Autoregressive Score*) vem sendo estudado desde 2008 e até então não foram aplicados a dados de sinistros IBNR. Os modelos em questão foram utilizados na grande maioria das vezes em séries financeiras, como pode ser visto em Creal et al. (2010), Creal et al. (2011), Xin et al. (2011), Caivano e Harvey (2013) Calvori et al. (2013) e Bazzi et al. (2014), por exemplo.

Outras contribuições mais específicas dentro do arcabouço dos modelos GAS se delineiam na construção de intervalos de confiança para as reservas estimadas, tanto para os períodos de acidente quanto para as reservas totais, como também na “extrapolação” do triângulo de *run-off*, chamado de efeito linha (ou reserva de prêmio), no qual se projetam os valores dos sinistros que ainda não aconteceram e serão, quando suceder, avisados com algum tempo de atraso.

No capítulo 2, é apresentada, de forma sucinta, uma revisão bibliográfica sobre sinistros ocorridos e não avisados - IBNR, com distintas abordagens acerca do tema. Entre os artigos destaca-se Mack (1993), onde é apresentado o método *benchmark* para estimação de IBNR, o *Chain Ladder*. É notório que esse método determinístico é utilizado pelas empresas seguradoras até os dias de hoje, devido a sua simplicidade.

No capítulo 3, são apresentados os modelos estruturais para séries temporais - especificamente a formulação SUTSE (*seemingly unrelated time series equations*) introduzida por Harvey (1989) e os modelos em espaço de estado multivariados, necessários para a estimação das componentes do modelo SUTSE proposto. Adicionalmente, será apresentado o modelo Atherino (2008), que instituiu a previsão de reservas do tipo IBNR, utilizando uma ordenação do triângulo de *run-off* distinta da apresentada por Mack (1993). Nesse arranjo, Atherino apresentou o triângulo de *run-off*, na forma incremental, empilhando por linhas, originando, assim, uma série temporal univariada.

No capítulo 4, são apresentados os modelos GAS, também conhecidos

como *Dynamic Conditional Score* (DSC) e sua adaptação para dados em painel, que pode ser considerado o caso das séries contidas nos triângulos de *run-off*.

No capítulo 5, estão descritos todos os resultados obtidos nesta tese, bem como os testes aplicados aos modelos aqui propostos. O capítulo apresenta os resultados do modelo Atherino bivariado (formulação SUTSE) e os resultados dos modelos GAS (1,1) com distribuições gama e log-normal.

Por fim, no capítulo 6, são apresentadas as principais conclusões e os trabalhos a serem desenvolvidos no futuro.

2

Estimação de reservas IBNR via triângulo de run-off

As estimativas das reservas IBNR necessitam ser precisas para que os balanços e os demonstrativos de resultados das seguradoras reflitam a sua eficiência operacional na emissão dos prêmios, posto que os agentes econômicos utilizam essas informações para avaliar o valor da companhia e tomar suas decisões de investimento. Complementarmente, reservas precisas revelam a verdadeira situação financeira da seguradora, minimizando o seu risco de insolvência e contribuindo para o alinhamento dos interesses entre a companhia e o agente regulador.

Com o objetivo de examinar algumas das metodologias utilizadas para o cálculo das reservas IBNR publicadas ao longo dos anos, esta seção apresenta um breve resumo dos principais resultados da literatura considerados relevantes para a contextualização dos modelos propostos nesta tese.

2.1

Sinistros no ramo de seguros

Segundo Pereira (2013), no mercado de seguros, o sinistro ocorre quando existe dano à pessoa ou ao bem segurado, não necessariamente ocorrendo dolo por parte do cliente. O segurado, ao envolver-se em um sinistro, por razões diversas, nem sempre notifica o mesmo no ato da ocorrência. Como exemplos podemos citar os sinistros do ramo vida, no qual os beneficiários da apólice muitas vezes desconhecem a mesma, ou ainda, quando os beneficiários adiam sua notificação à seguradora devido ao seu período de luto, ou mesmo pela demora no processo de inventário do segurado.

No segmento que utilizaremos nesta tese, automóveis, o atraso na notificação do sinistro não é tão comum como no ramo vida. Contudo, a proporção do valor pago pelos reportes atrasados é significativa. Como exemplo podemos citar que, para uma dada seguradora, do total de sinistros do segmento de automóveis (Casco, RCFV e APP), ocorridos no primeiro trimestre de 2009, aproximadamente dois terços foram reportados a esta seguradora no mesmo trimestre e o restante foram reportados ao longo dos trimestres seguintes. Sendo assim, o empenho para prever o valor das reservas IBNR do segmento de automóveis é relevante.

2.2

SUSEP - Órgão regulador de seguros no Brasil

A SUSEP - Superintendência de Seguros Privados - dentre outras atividades, determina critérios para o cálculo de reservas IBNR e acompanha as reservas estabelecidas pelas seguradoras que atuam no Brasil. Como signatária da IAIS - *International Association of Insurance Supervisors* - a SUSEP, na regulação das sociedades seguradoras, segue os princípios estabelecidos por ela, aplicando as abordagens de avaliação de solvência, identificação de riscos e requerimento de capital que norteiam o Solvência II¹.

2.3

Formato dos dados de sinistros ocorridos e não avisados (IBNR)

A maior parte das metodologias de cálculo de reserva IBNR proposta até hoje faz uso de um formato de dados particular chamado triângulo de *run-off*, no qual os dados são agrupados conforme datas de ocorrência e pagamento ou aviso do sinistro, entre outras. Como exemplo de trabalhos publicados que utilizaram essa forma de agrupamento, têm-se: Bornhuetter e Ferguson (1972), Skurnick (1973), Berquist e Sherman (1977), Verrall (1989), Mack (1993), De Jong (2004), Verrall, Nielsen e Jessen (2010) e Adams (2013).

Os dados de sinistros IBNR organizados nesta forma também podem ser expostos em outras duas, a saber: empilhamento e organização pela diagonal. As três formas de agrupamento dos dados serão explicitadas nas próximas subseções.

2.3.1

Triângulo de run-off

A função do triângulo de *run-off* é apresentar, de forma concisa, os sinistros ocorridos e não avisados, em sua totalidade, no mesmo período de ocorrência. Para tanto, os montantes de valores pagos relativos a sinistros ocorridos são apresentados em uma estrutura matricial na qual os sinistros aparecem classificados por período de ocorrência, representados pelas linhas, e por prazos de desenvolvimentos dos pagamentos, representados pelas colunas.

Os triângulos de *run-off* são usados para projetar o desenvolvimento futuro de um *portfólio*, com base em um certo padrão histórico. A crença de que existem determinados padrões no comportamento inerente aos avisos de

1

Projeto de consolidação de normas relativas à solvência de seguradoras iniciado na Europa em 2009. Para maiores informações pesquisar o relatório “Capital adicional relativo ao Risco operacional” publicado pela SUSEP em março de 2012.

ocorrência dos sinistros é o ponto chave a todos os métodos de previsão de reservas IBNR, sobretudo aos aplicados em triângulos de *run-off*.

A previsão, neste caso, é obtida, “preenchendo” a parte inferior do triângulo de *run-off*. Tais valores, a serem estimados, são propriamente designados de reservas IBNR.

Com os dados observados de pagamentos estruturados na forma do triângulo de *run-off*, o objetivo passa a ser inferir sobre os valores de pagamentos futuros referentes aos sinistros já ocorridos, ou seja, sobre os valores abaixo do triângulo, denominado de triângulo inferior (ALTIERI, 2013, p. 42).

Como pode ser visto em Taylor e Ashe (1983), Verrall (1989), Mack (1993), De Jong e Heller (2008) e Costa (2010), o triângulo de *run-off* foi apresentado como forma de organização dos dados de sinistros do tipo IBNR de forma similar. A seguir são descritas as notações necessárias para a apresentação dos dados no formato mencionado:

t : Período de origem: data de ocorrência do sinistro;

i : Período de desenvolvimento: número de período referentes ao atraso no pagamento;

$C_{t,i}$: Valor pago pela seguradora referente aos sinistros originados no período t , mas que foram pagos com i instantes de tempo de atraso e

$D_{t,i}$: Valor pago pela seguradora referente aos sinistros originados no período t , mas que foram pagos até i instantes de tempo de atraso.

O triângulo de *run-off* possui duas representações básicas, a saber: a incremental e a acumulada, que serão descritas a seguir.

2.3.2

Representação incremental do triângulo de *run-off*

Assim como exposto por Atherino (2008), na forma incremental, cada célula do triângulo é representada por $C_{t,i}$, no qual, $1 < t < T$ e $1 < i < N$, como mostra a Tabela 2.1

Tabela 2.1: Triângulo de *run-off* - representação incremental.

Período de origem t	Período de desenvolvimento i					
	1	2	3	...	N - 1	N
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$...	$C_{1,N-1}$	$C_{1,N}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$...	$C_{2,N-1}$	
⋮	⋮	⋮				
T - 1	$C_{T-1,1}$	$C_{T-1,2}$				
T	$C_{T,1}$					

Esta representação dos dados de sinistros IBNR pode ser vista como dados em painel. A dimensão de origem é representada pelas linhas do triângulo, $t = 1, 2, \dots, T$ e a dimensão de desenvolvimento é representada pela coluna, $i = 1, 2, \dots, N$.

2.3.3

Representação acumulada do triângulo de run-off

Na forma acumulada do triângulo de *run-off*, as células são calculadas como demonstrado na Equação 2-1:

$$D_{t,i} = \sum_{i=1}^N C_{t,i} \tag{2-1}$$

O triângulo de *run-off* contendo os dados acumulados dos sinistros ocorridos e não avisados está apresentado na Tabela 2.2.

Tabela 2.2: Triângulo de *run-off* - representação acumulada.

Período de origem t	Período de desenvolvimento i					
	1	2	3	...	N - 1	N
1	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$...	$D_{1,N-1}$	$D_{1,N}$
2	$D_{2,1}$	$D_{2,2}$	$D_{2,3}$...	$D_{2,N-1}$	
⋮	⋮	⋮				
T - 1	$D_{T-1,1}$	$D_{T-1,2}$				
T	$D_{T,1}$					

2.3.4

Empilhamento dos dados contidos no triângulo de run-off

Segundo a proposta de Atherino et al. (2010), para a aplicação do modelo dos blocos, aqui chamado modelo Atherino, será necessário uma reorganização dos dados dos sinistros contidos no triângulo de *run-off* em

uma forma chamada “empilhamento”. Essa forma requer que se estructurem os dados utilizando a forma incremental do triângulo (Subseção 2.3.2). Para um triângulo de *run-off* com $T = N$ tem-se:

Tabela 2.3: Reparametrização “por linhas” do triângulo de *run-off*.

Período de origem t	Período de desenvolvimento i				
	1	2	3	...	N
1	y_1	y_2	y_3	...	y_N
2	y_{N+1}	y_{N+2}	...	y_{2N-1}	y_{2N}
3	y_{2N+1}	y_{2N+2}	...	y_{3N-1}	y_{3N}
⋮	⋮	⋮	⋮	⋮	⋮
T	$y_{(N-1)N+1}$	$y_{(N-1)N+2}$...	y_{N^2-1}	y_{N^2}

Como pode ser visto na Tabela 2.3, os índices t e i cedem lugar a um único índice t . Nesta representação apresentada por Atherino et al. (2010) y se refere a cada célula do triângulo de *run-off* incremental, que na Subseção 2.3.2 foi denominada de C . Nos capítulos seguintes utilizaremos a notação apresentada na Tabela 2.3 e nas sequências 2-2 e 2-3.

Empilhando os dados apresentados no triângulo de *run-off* 2.3, tem-se a série:

$$y_1, \dots, y_N, y_{N+1}, \dots, y_{2N}, y_{2N+1}, \dots, y_{3N}, \dots, y_{(N-1)N+1}, \dots, y_{N^2} \quad (2-2)$$

Contudo, os triângulos de *run-off*, no qual os dados empilhados serão retirados para o ajuste do modelo proposto, apresentam dados ausentes. Empilhando os dados com elementos ausentes, assim como feito na Equação 2-2, e completando os “espaços vazios” com “NA”, tem-se a série 2-3.

$$y_1, y_2, \dots, y_N, y_{N+1}, y_{N+2}, \dots, y_{2N-1}, NA, y_{2N+1}, y_{2N+2}, \dots, y_{3N-2}, NA, NA, y_{3N+1}, \dots, y_{4N-3}, NA, NA, NA, y_{4N+1}, \dots \quad (2-3)$$

Com esta ordenação e instituindo o uso da formulação em espaço de estado (Capítulo 3), as estruturas de dependência entre os sinistros, apresentadas no triângulo de *run-off*, podem ser modeladas como uma série temporal.

Cabe aqui ressaltar que o índice t adotado não representa a ordem cronológica dos pagamentos, nem da ocorrência do sinistro, mas sim, uma

nova forma de organizar os dados. Assim, possibilita a análise do triângulo através de modelos que admitam componentes como: nível, tendência, ciclo, sazonalidade e periodicidade (Subseção 3.2). É importante salientar também que toda a teoria apresentada nesta tese pode ser aplicada a triângulos de *run-off* com quantidades de linhas e colunas diferentes, ou seja: $T \neq N$.

A reserva IBNR propriamente dita consiste na soma dos valores faltantes (NAs) apresentados na sequência 2-3, ou seja:

$$IBNR = \sum_{t: y_t \text{ é ausente}} y_t \quad (2-4)$$

Nesta tese os dados dos avisos de ocorrência serão agrupados ora na forma triângulo de *run-off*, ora na forma chamada empilhamento, como apresentado por Atherino (2008).

2.3.5

Organização pela diagonal do triângulo de run-off

De Jong e Zehnwirth (1983) e Verrall (1994) reorganizaram os dados de sinistros IBNR pelas diagonais dos triângulos de *run-off* incrementais com a finalidade de aplicar modelos dinâmicos de estimação utilizando o filtro de Kalman.

[...] Ordenar o Triângulo de *run-off* pela sua diagonal nada mais é que reorganizar os índices de linhas e colunas de forma que a última diagonal preenchida do triângulo corresponda a um dado vetor. Este contendo todos os lançamentos daquele período calendário independente de qual o seu ano de origem ou de desenvolvimento (PEREIRA, 2013, p. 24).

Nesta organização, a cada novo período, o vetor y_T aumenta sua dimensão, ou seja, no período $t - 1$ ele terá $t - 1$ elementos e no período t terá t elementos e assim por diante. Nesta representação para um dado $y_{t,i}$ haverá um $y_j(t)$ cujo j denota o período de desenvolvimento, $j = 0, 1, \dots, t-2, t-1$. Cabe ressaltar que t é a representação do tempo de calendário e não de ocorrência ou desenvolvimento do sinistro, como no caso das ordenações apresentadas nos Subitens 2.3.2 e 2.3.3.

Tabela 2.4: Reparametrizaco “pela diagonal” do tringulo de *run-off*.

Perodo de origem t	Perodo de desenvolvimento i				
	1	2	3	...	N
1	$y_0(1)$	$y_1(2)$...	$y_{N-1}(T-1)$	$y_N(T)$
2	$y_0(2)$	$y_1(3)$...	$y_{N-1}(T)$	
...		
...	$y_2(T)$		
T-1	$y_0(T-1)$	$y_1(T)$			
T	$y_0(T)$				

Logo, $y_j(t)$  o valor do tringulo de *run-off* observado no desenvolvimento j e perodo de calendrio t . A diagonal destacada na Tabela 2.4 ser reescrita na forma vetorial, como mostra a equaco a seguir.

$$y_{(T)} = \begin{bmatrix} y_0(T) \\ y_1(T) \\ \dots \\ y_{N-1}(T) \\ y_N(T) \end{bmatrix} \tag{2-5}$$

O vetor $y_j(T)$ possui dimenso $T \times 1$ no instante T . Como pode ser visto, o primeiro valor do vetor  o valor de menor ndice.

Isso posto, a quantidade de vetores ser proporcional  quantidade de diagonais existentes em cada tringulo de *run-off*. Verrall (1989, 1994) utilizou um modelo em espaço de estado para cada vetor contendo os dados das diagonais dos tringulos de *run-off* selecionados.

2.4

Tipos de reservas calculadas

Nos tpicos a seguir sero descritos trs tipos de reservas calculadas a partir do tringulo de *run-off*: reserva por linhas, reserva total e por perodo de calendrio. Essas so as denominaçes de reservas mais utilizadas nos artigos de aturia.

Reserva por linhas: Apresenta a soma dos valores desconhecidos cuja ocorrncia se d no perodo t . A reserva por linhas foi utilizada em: Bornhuetter e Ferguson (1972), Verrall (1989), Mack (1993), De Jong (2004), Mack e Re(2006), Atherino et al.(2010), Verrall, Nielsen e Jessen (2010), Costa(2010), entre outros.

Reserva total: Representa a soma de todas as reservas estimadas para os períodos t . A reserva total foi considerada em: De Jong(2004), Klaus et al.(2008), Atherino et al. (2010), Costa(2010), entre outros.

Reserva por período de calendário: É dada pela soma de todas as reservas a serem formadas no mesmo período calendário. A reserva em questão foi utilizada por Venter (2007), entre outros.

Ressalta-se que, embora, existam três tipos distintos de reservas IBNR, nesta tese, serão calculadas apenas as reservas por linhas e reservas totais.

2.5

Adição do efeito linha ao triângulo de run-off

O triângulo de *run-off* é uma representação estática, em uma determinada data, dos períodos de origem e períodos de desenvolvimento. A medida que se avança no tempo, as linhas e as colunas da matriz aumentam conforme o acréscimo de novas informações à base das seguradoras.

A Tabela 2.5 apresenta o triângulo de *run-off* incremental com efeito linha - também conhecido como reserva de prêmio.

Tabela 2.5: Triângulo de *run-off* - representação incremental com efeito linha.

Período de origem t	Período de desenvolvimento i					
	1	2	3	...	$N - 1$	N
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$...	$C_{1,N-1}$	$C_{1,N}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$...	$C_{2,N-1}$	
⋮	⋮	⋮				
$T - 1$	$C_{T-1,1}$	$C_{T-1,2}$				
T	$C_{T,1}$					
$T + 1$	$C_{(T+1),1}$					
⋮	⋮	...				
τ	$C_{\tau,1}$...				

A parte cinza da Tabela 2.5 representa os sinistros que irão ocorrer e, por algum motivo, não serão avisados no mesmo tempo de ocorrência. As previsões destes valores são fundamentais para o funcionamento futuro da seguradora, visto que, a perenidade de suas atividades, assim como, das demais operadoras de seguro, é um preceito base para a robustez do mercado de seguros. Maiores detalhes serão apresentados na Subseção 4.2.4.

A previsão assertiva destes valores, de certa forma, está associada à antecipação do montante que sairá do caixa da empresa e consecutividade mitiga o risco da seguradora.

2.6

Metodologia Univariada para o Cálculo das Reservas IBNR proposta por Mack (1993) - Chain Ladder

De acordo com Mack (1993 *apud* Costa, 2010, p.16), o *Chain Ladder* é provavelmente o método mais popular para estimação de reservas IBNR, devido, principalmente, à simplicidade de sua utilização. Como referências importantes sobre o método, citam-se Kremer (1982), Mack (1993), Verrall (1994), England e Verrall (1998, 2002), Renshaw e Verrall (1998) e Taylor (2000).

2.6.1

Chain Ladder - Notações e resultados básicos

Assume-se que existem $l - 1$ fatores (*developments factors*) e que $C_{t,i}$ é uma variável aleatória que possui observações se $t + i \leq 1 + l$. O objetivo do método é estimar a quantidade de sinistros finais, $C_{t,l}$, e as reservas de sinistros excedentes. Seja:

$$E(C_{t,i+1}|C_{t,1}, \dots, C_{t,i}) = C_{t,i}f_i \quad (2-6)$$

$$\hat{f}_i = \frac{\sum_{j=1}^{l-t} C_{j,t+1}}{\sum_{j=1}^{l-t} C_{j,t}} \quad (2-7)$$

$$E(\hat{f}_{l+1-t} \cdot \dots \cdot \hat{f}_{l-1}) = f_{l+1-t} \cdot \dots \cdot f_{l-1} \quad (2-8)$$

$$\hat{C}_{t,l} = C_{t,l+1-t} \cdot \hat{f}_{l+1-t} \cdot \dots \cdot \hat{f}_{l-1} \quad (2-9)$$

Tem-se que: $f_1, \dots, f_{l-1} > 0$ e $1 \leq l \leq 1 \leq i \leq l - 1$.

Segundo Mack (1993), \hat{f}_i é a razão dada pelo somatório dos sinistros avisados até $i + 1$, ocorridos em $t = 1, \dots, l - 1$, com o somatório dos sinistros avisados até i .

Esta igualdade é provada no artigo por Mack (1993) usando a lei das expectativas iteradas, a própria definição de esperança e a regra de probabilidade total.

Utiliza-se \hat{f}_i na Equação 2-6 ao invés de f_i . A prova de que tal substituição é factível também foi desenvolvida em Mack (1993).

Para o cálculo de $\hat{C}_{t,l}$, em cada período t , leva-se em conta a quantidade total de sinistros acumulados e o produtório dos fatores referentes ao período de desenvolvimento $i = l + 1 - t, \dots, l - 1$.

A mesma reserva pode ser calculada através de duas equações distintas:

$$\hat{R}_t = C_{t,l+1-t}(\hat{f}_{l+1-t} \cdot \dots \cdot \hat{f}_{l+1} - 1)\hat{R}_t = C_{t,l} - C_{t,l+1-t}$$

A reserva, \hat{R}_t , é a diferença entre o total acumulado de sinistros ocorridos e o provisionado até o período de desenvolvimento l .

É fato estabelecido que o método *Chain Ladder* é o *Benchmark* no mercado de seguros, pois não utiliza informações externas ao triângulo de *run-off*, pela simplicidade de sua implementação e pela produção de estimativas IBNR adequadas. Nesta tese o resultado das estimativas produzidas pelo modelo *Chain Ladder* será utilizado na comparação com os resultados dos modelos propostos (Seção 5.4).

Ressalta-se que Verrall (1989), England e Verrall (2002), Merz e Wutrich (2008), Atherino et al. (2010), Costa (2010), Pereira (2013), entre outros optaram, também, por comparar os modelos propostos em seus trabalhos com o método *Chain Ladder*.

2.6.2

Erro médio quadrático obtido por Mack

O $MSE(\hat{C}_{tl})$ do \hat{C}_{tl} de C_{tl} pode ser definido pelas seguintes equações:

$$MSE(\hat{C}_{tl}) = E((\hat{C}_{tl} - C_{tl})^2 | D) \quad (2-10)$$

No qual $D = \{C_{ti} | t + i \leq l + 1\}$.

$$MSE(\hat{C}_{tl}) = \text{Var}(C_{tl} | D) + (E(C_{tl} | D) - \hat{C}_{tl})^2$$

A forma de calcular cada um desses termos é dada a seguir.

$$\text{Var}(C_{tl} | D) = \hat{C}_{tl}^2 \sum_{i=l+1-t}^{l-1} \frac{\sigma_i^2 | f_i^2}{\hat{C}_{ti}} \quad (2-11)$$

$$(E(C_{tl} | D) - \hat{C}_{tl})^2 = C_{t,l+1-t}^2 \hat{f}_{l+1-t}^2 \cdot \dots \cdot \hat{f}_{l+1}^2 \sum_{i=l+1-t}^{l-1} \frac{\hat{\sigma}_i^2 | \hat{f}_i^2}{\sum_{j=1}^{l-1} C_{ji}} \quad (2-12)$$

Em que: $1 \leq t \leq l$, $1 \leq i \leq l - 1$.

2.7

Metodologias multivariadas para o cálculo das reservas IBNR

Outros avanços na teoria abordada por Mack (1993) foram publicados em Braun e Munich (2004). Os autores ressaltaram que aplicar o método *Chain Ladder*, apresentado na Subseção 2.6 para o triângulo de *run-off* agregado, com a finalidade de obter uma reserva total e seu respectivo erro padrão, não é uma boa solução, porque as estimativas das reservas obtidas dessa forma não serão idênticas as estimativas obtidas pela soma das estimativas desagregadas. Além disso, a agregação de triângulos de *run-off*, com diferentes padrões de desenvolvimento, é como “misturar maçãs e laranjas” e normalmente conduz a resultados inválidos.

Em seu artigo Braun e Munich (2004) assumiram ser possível encontrar relação entre dois triângulos de *run-off*. Nesta abordagem há um coeficiente de correlação individual para cada período de desenvolvimento, diferentemente de outras abordagens que calculam apenas um coeficiente de correlação para os dois triângulos de *run-off*. Não obstante, é factível realizar algumas modificações para obter apenas um coeficiente e viabilizar a comparação entre a abordagem de Braun e Munich e outras distintas.

Schmidt e Zocher (2006) expuseram uma versão multivariada do modelo aditivo de perdas de reservas. O modelo apresentado consiste em um modelo linear que faz uso de matrizes e estruturas de variâncias. Os autores apresentaram a forma matemática da adequação do modelo univariado para o multivariado e ressaltaram que esse modelo é apropriado para certas carteiras compostas de várias subcarteiras correlatas.

[...] Quarg e Mack (2004) instauraram uma extensão bivariada especial do método *Chain Ladder*. No mesmo ano, Braun e Munich propuseram um outro modelo bivariado, a fim de construir estimadores dependentes para os erros das previsões provenientes do modelo *Chain Ladder*. O modelo de Braun e Munich (2004) foi, então, estendido por Prohl e Schmidt (2005) que utilizaram um critério de otimização para desenvolver uma versão multivariada do método *Chain Ladder*. Um outro artigo associado foi o publicado por Kremer (2005), no qual propuseram uma outra extensão multivariada do método *Chain Ladder*, no entanto, segundo o autor, este método não é adequado para aplicações que envolvam inversões de matrizes muito grandes (SCHMIDT; ZOCHER, 2006, p.1).

Merz e Wuthrich (2008) também aplicaram o método *Chain Ladder* multivariado para uma carteira de n triângulos de *run-off* correlacionados. Conforme descrito pelos autores, o estudo simultâneo de várias subcarteiras correlacionadas foi motivado por abordar o problema da dependência entre diferentes

linhas de negócios. Uma das vantagens é que, observando um triângulo de *run-off* de uma subcarteira, se pode aprender sobre o comportamento dos outros triângulos das subcarteiras. Na prática, é natural subdividir um *portfolio* em várias subcarteiras correlacionadas, de modo que cada uma satisfaça certas condições de homogeneidade.

Algumas das premissas aceitas por Merz e Wuthrich (2008), no artigo em questão, são: as reivindicações acumuladas C_{ik} de diferentes anos do acidente i são independentes; o modelo é discreto no tempo; a dependência entre diferentes triângulos de *run-off* é definida através de uma matriz de correlação e no caso de apenas um triângulo de *run-off* as estimativas dos fatores \hat{f}_i coincidem com as estimativas dos fatores clássicos.

Merz e Wuthrich (2008) expuseram uma forma recursiva e matricial do processo de variância condicional para os N triângulos de *run-off* correlacionados e apresentaram, adicionalmente, a variância condicional por anos de acidentes individuais. No exemplo aplicado por Merz e Wuthrich (2008), o uso do método *Chain Ladder* multivariado apresentou apenas um pequeno efeito na diversificação e um resultado ligeiramente melhor para o erro de estimação condicional em relação à abordagem univariada. Apesar disso, o impacto da utilização da análise multivariada do método *Chain Ladder*, para estimar os fatores de desenvolvimento \hat{f}_i , foi a maior diferença com relação à forma tradicional.

2.8

Modelos em espaço de estado e reservas IBNR

O uso dos modelos em espaço de estado, com o propósito de estimar as reservas IBNR, tem propiciado inúmeras contribuições da literatura atuarial, cabem citar: De Jong e Zehnwirth (1983) e Verrall (1989).

Na abordagem De Jong e Zehnwirth modelou-se as provisões de sinistros IBNR, utilizando os sinistros já ocorridos e organizados pela diagonal do triângulo de *run-off* incremental. Para isso, estimou-se os parâmetros estocásticos do modelo e sequencialmente calculou a reserva IBNR, fazendo uso de modelo de espaço de estados e do filtro de Kalman.

De Jong e Zehnwirth (1983) ressaltaram que a abordagem de espaço de estado, para a modelagem de reservas, tem vantagens importantes por ser flexível e unificada, sendo adequada a uma ampla variedade de circunstâncias, evitando a necessidade de cálculos complicados ou modos de análises muitas vezes confusas. Os autores também defendem que a abordagem centra-se na natureza de previsão e está alinhada à teoria de séries temporais.

Alpuim e Ribeiro (2003) sugeriram um modelo de espaço de estado

distinto, livre de distribuição para estimação da parte inferior do triângulo de *run-off*, usando também o filtro de Kalman. Os autores relacionaram o triângulo com uma coluna de valores observados escolhidos de forma a descrever melhor o valor em risco de cada ano de acidente.

O modelo proposto por Alpuim e Ribeiro é aplicável aos dados acumulados e assume que a coluna subsequente é proporcional à coluna anterior. No entanto, o modelo não faz uso dos mesmos parâmetros em todas as linhas da mesma coluna. No modelo apresentado, os autores assumem que a primeira coluna do triângulo de *run-off* contém valores determinísticos e as seguintes colunas contêm observações ou previsões de variáveis aleatórias, dependentes dos valores da primeira coluna.

Costa (2010) utilizou três metodologias para modelagem de sinistros IBNR, e dedicou-se a verificar, por meio de exercícios empíricos com dados reais, se existe uma abordagem estatisticamente mais eficaz.

[...] A primeira metodologia se baseou no método clássico *Chain Ladder*, a segunda em modelos de regressão linear com heterocedasticidade, sob o arranjo usual do triângulo de *run-off* via duplo-índice e a terceira metodologia inseriu-se no arcabouço dos modelos em espaço de estado lineares e do filtro de Kalman. O autor constatou a superioridade dos dois modelos propostos comparado ao método clássico *Chain Ladder* nas séries de dados reais utilizadas em seu estudo (COSTA, 2010, p. 18).

Atherino et al. (2010) propuseram a previsão de reservas do tipo IBNR levando em conta uma ordenação distinta do triângulo (Subseção 2.3.4). Os autores apresentaram o triângulo de *run-off* na forma incremental empilhado por linhas, originando uma série temporal univariada repleta de valores ausentes, em que a soma desses valores constitui a reserva IBNR a ser estimada.

Vale ressaltar que Atherino et al. (2010) desenvolveram duas abordagens de estimação para os valores das reservas IBNR, inteiramente baseadas na teoria dos modelos em espaço de estado e filtro de Kalman. Uma das contribuições desta tese é a extensão do modelo dos blocos apresentado por Atherino para a estimação das reservas IBNR (Subseção 3.3.1). A extensão se dá no desenvolvimento do modelo Atherino na forma multivariada, o que proporciona o estudo da dependência dos triângulos de *run-off*, através das correlações entre os resíduos e/ou através de componentes comuns.

Ao longo desta tese a extensão do modelo dos blocos será mencionada como modelo Atherino bivariado.

3

Modelo Atherino bivariado

Neste capítulo será apresentada a primeira contribuição desta tese, a extensão bivariada do modelo Atherino. O desenvolvimento aqui apresentado esta direcionado à modelagem e previsão de reservas IBNR pertencentes a dois triângulos de *run-off* incrementais de coberturas relacionadas entre si. Cabe ressaltar que o modelo Atherino bivariado pode ser estendido para um número arbitrário de carteiras (coberturas). O desenvolvimento para n carteiras será facilmente deduzido a partir do que será apresentado a seguir.

O modelo Atherino bivariado desenvolvido nesta tese tem como objetivo estimar reservas IBNR (Subseção 2.4) captando, caso exista, a dependência entre as coberturas e a dependência entre as células de cada triângulo de *run-off* (Subseção 2.3.1). A parte do modelo que captura a dependência entre as duas coberturas está relacionada à estrutura das matrizes de covariâncias presentes na formulação SUTSE e a parte do modelo que capta a dependência entre as células de cada triângulo foi desenvolvida por Atherino et al. (2010).

O modelo Atherino pode ser usado para captar a dependência entre as células de cada triângulo de *run-off* somente se os dados das coberturas estiverem no formato “empilhado”, conforme apresentado na Subseção 2.3.4. Essa forma de organizar os dados possibilita a análise dos dados contidos nos triângulos através de modelos que admitam componentes como: tendência, ciclo e periodicidade (Subseção 3.2).

Para melhor a organização e a exposição, este capítulo se divide em quatro seções. A primeira seção denota a suposição de log-normalidade para os dados contidos no triângulo de *run-off*. A segunda seção apresenta o modelo Atherino univariado para triângulos de *run-off* (aplicação dos modelos estruturais), os modelos em espaço e de estado (que são a chave para lidar com o modelo Atherino), o cálculo da matriz de covariância condicional das parcelas do IBNR e o cálculo dessas reservas propriamente ditas. Na terceira seção são apresentados o modelo Atherino bivariado e a formulação SUTSE - *seemingly unrelated time series equations*). Por fim, a última seção deste capítulo apresenta as medidas de aderências que serão utilizadas nesta tese.

3.1

Uso da distribuição log-normal para a série composta pelos dados contidos no triângulo de *run-off*

O uso da distribuição log-normal para a modelagem dos dados contidos no triângulo de *run-off* teve como inspiração o trabalho do Mack publicado em 1994. No artigo em questão, ainda que o autor tenha desenvolvido seu modelo sem precisar admitir qualquer distribuição, apresentou um exemplo considerando log-normalidade para a construção gráfica das reservas IBNR.

Segundo Atherino (2008), a distribuição log-normal foi amplamente utilizada na literatura atuarial (cf. Taylor, 2000, Capítulo 9), tanto na área de Modelos Lineares Generalizados (MLG) quanto na área de série temporais. Em MLG, pode-se citar os trabalhos de Kremer (1982), Renshaw (1989), Christofides (1990), Verrall (1991) e Doray (1996); na área de séries temporais, têm-se de Jong & Zehnwirth (1983), Verrall (1989), de Jong (2004) e de Jong (2006).

A distribuição log-normal, para a série de reservas IBNR na forma “empilhada” induz, por primeiros princípios, a modelagem do logaritmo das entradas dos triângulos de *run-off* incrementais. Sendo assim, $x_t = \log(y_t)$, y_t apresentado na Tabela 2.3. Desta forma, após a aplicação do modelo Atherino bivariado ao logaritmo das séries “emplilhadas”, será necessária uma transformação da escala log-normal para a normal (Subseção 3.2.4). Após a transformação têm-se as reservas IBNR e seus Coeficientes de Variação.

3.2

Modelo Atherino univariado

Como será visto, o modelo Atherino é uma aplicação dos modelos estruturais quando os dados estão expressos na forma “empilhada” apresentada na Subseção 2.3.4.

Na abordagem estrutural, segundo Harvey (1989), uma série temporal é decomposta em componentes que possuem interpretação, a saber: a tendência, que representa os movimentos de longo prazo da série; a sazonal, que captura os movimentos periódicos até um ano; o cíclico, que acompanha o curso do ciclo de negócios ou ciclos econômicos com duração maior do que um ano; e, finalmente, a componente de erro ou componente irregular, que caracteriza os efeitos aleatórios.

Estimar as reservas IBNR usando este tipo de modelo é intuitivo desde que se considere a reordenação dos dados contidos no triângulo de *run-off* como apresentado na Subseção 2.3.4.

O conjunto de equações que será apresentado a seguir refere-se a um modelo estrutural com nível e componente periódica estocásticas para apenas uma série “empilhada”.

Equação das observações:

$$x_t = \mu_t + \gamma_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (3-1)$$

Equação de estado:

$$\mu_{t+1} = \mu_t + \eta_t \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (3-2)$$

$$\gamma_{t+1} = -\sum_{j=1}^{N-1} \gamma_{t+1-j} + \omega_t \quad \omega_t \sim N(0, \sigma_\omega^2) \quad (3-3)$$

$$E[\varepsilon_t \eta_t] = E[\varepsilon_t \omega_t] = 0 \quad \forall t \text{ e } s$$

$$E[\varepsilon_t \alpha_1] = E[\eta_t \alpha_1] = E[\omega_t \alpha_1] = 0 \quad \forall t$$

Temos que N , apresentado na Equação 3-3, é o número de colunas que o triângulo de *run-off* apresenta (vide Tabela 2.3), ou seja, N está relacionado à periodicidade apresentada pela série “empilhada” a qual pretende-se ajustar o modelo. Nesta representação μ_t representa o nível do volume de sinistros ocorridos e não avisados em cada período de acidente, γ_t é a componente periódica que captura as flutuações no volume de sinistros IBNR entre períodos de origens sucessivos e, por fim, ε_t é o termo aleatório.

Cabe destacar que nesta tese o modelo Atherino bivariado será apresentado e aplicado sem componente cíclica. Os triângulos de *run-off* que serão utilizados na aplicação (Tabelas 5.1 e 5.1) contêm dados do primeiro trimestre de 2009 até o segundo trimestre de 2013, período este com um intervalo de tempo muito curto para a identificação de ciclos de negócios ou econômicos.

A chave para lidar com o modelo Atherino é, sem dúvidas, a sua colocação na forma em espaço de estado. Nesse modelo as componentes são estimadas recursivamente através do filtro de Kalman (FK) ou de forma mais completa, utilizando o algoritmo de suavização (*smoothing*).

Para maiores detalhes sobre modelos estruturais, forma em espaço de estado, filtro de Kalman e algoritmos de suavização ver: Harvey (1989), Harvey (1991), Durbin e Koopman (2001) e Commandeur e Koopman (2007).

O modelo estrutural univariado proposto nas Equações 3-1, 3-2 e 3-3 pode ser reescrito na formulação espaço de estado, que será apresentada na próxima subseção.

3.2.1

Forma em espaço de estado linear e Gaussiana

Considere a forma em espaço de estado linear e Gaussiana como apresentada em Durbin e Koopman (2001, p. 38).

Equação das observações:

$$\begin{matrix} x_t & = & Z_t & \alpha_t & + & \varepsilon_t & & t = 1, 2, \dots, T \\ p \times 1 & & p \times m & m \times 1 & & p \times 1 & & \end{matrix} \quad (3-4)$$

Equação do estado:

$$\begin{matrix} \alpha_{t+1} & = & T_t & \alpha_t & + & R_t & \eta_t & t = 1, 2, \dots, T \\ m \times 1 & & m \times m & m \times 1 & & m \times r & r \times 1 & \end{matrix} \quad (3-5)$$

Sendo:

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \text{NID} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right) \quad (3-6)$$

$$E[\varepsilon'_t \alpha_1] = E[\eta'_t, \alpha_1] = 0, \forall t$$

$$\alpha_1 \sim N(a_1, p_1) \quad (3-7)$$

As matrizes do sistema, Z_t , T_t , e R_t , são supostas determinísticas. O vetor x_t contém os valores das p séries temporais no período t .

Dada uma série temporal de tamanho t do processo X_t , definam-se $X_j \equiv (x_1, \dots, x_j)$, $a_{t|j} \equiv E[\alpha_t | X_j]$ e $P_{t|j} \equiv \text{Var}[\alpha_t | X_j]$. As equações de predição e de suavização fornecem fórmulas recursivas para o cálculo dos momentos condicionais de α_t quando $j = t - 1$ e para $j = n$, respectivamente. Suas expressões analíticas se encontram nas Equações 3-8, 3-9 e 3-10. Suas derivações, sob os pressupostos da forma em espaço de estado adotada nesta tese, podem ser obtidas em Durbin e Koopman (2001).

$$\begin{aligned} v_t &= x_t - Z_t \alpha_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z'_t + H_t, \\ K_t &= T_t P_{t|t-1} Z'_t F^{-1}, & L_t &= T_t - K_t Z_t, \quad t = 1, \dots, T, \end{aligned} \quad (3-8)$$

Equações de predição:

$$\begin{aligned} a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} L'_t + R_t Q_t R'_t, \\ r_{t-1} &= Z'_t F_t^{-1} v_t + L'_t r_t, & N_{t-1} &= Z'_t F_t^{-1} Z_t + L'_t N_t L_t. \end{aligned} \quad (3-9)$$

Equações de suavização:

$$\begin{aligned} a_{t|n} &= a_{t|t-1} + P_{t|t-1}r_{t-1}, & P_{t|n} &= P_{t|t-1} = P_{t|t-1}N_{t-1}P_{t|t-1}, \\ r_n &= 0, & N_n &= 0 & t = 1, \dots, T. \end{aligned} \quad (3-10)$$

Diversos modelos estatísticos e econométricos podem ser representados em uma forma de espaço de estado. Citam-se como exemplos: modelos estruturais para séries temporais, modelos ARIMA e SARIMA, modelos autorregressivos vetoriais (VAR), modelos de regressão com coeficientes fixos ou variantes de forma estocástica no tempo, entre outros.

O modelo estrutural, para uma única série, proposto nas Equações 3-1, 3-2 e 3-3 tem a seguinte forma em espaço de estado:

$$x_t = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{pmatrix} \mu_t \\ \gamma_t \\ \gamma_{t-1} \\ \vdots \\ \gamma_{t-N+1} \end{pmatrix} + \varepsilon_t \quad (3-11)$$

$$\begin{pmatrix} \mu_{t+1} \\ \gamma_{t+1} \\ \gamma_t \\ \vdots \\ \gamma_{t-N} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{pmatrix} \mu_t \\ \gamma_t \\ \gamma_{t-1} \\ \vdots \\ \gamma_{t-N+1} \end{pmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \eta_t \\ \omega_t \end{pmatrix} \quad (3-12)$$

Colocando o modelo Atherino na forma em espaço de estado, fazendo uso do filtro de Kalman e algoritmo de suavização (ver Atherino et al., 2010), as componentes podem ser estimadas facilmente.

O próximo passo é calcular a matriz de covariância condicional das parcelas do IBNR (vide Equação 2-4). A subseção seguinte apresenta, de forma sucinta, os resultados desenvolvidos em Atherino et al. (2010).

3.2.2 Matriz de covariância condicional

Considerando a forma em espaço de estado (Subseção 3.2.1) e toda a notação correspondente a esta (Anexo A), adicionalmente define-se:

$$- \mathcal{J} \equiv \{t: x_t \text{ é não ausente}\};$$

- $\tilde{X} \equiv \{x_{i_j} : i_j \in \mathfrak{J}, \forall_j\}^1$;
- $X \equiv \{x_t : t = 1, \dots, n\}$;
- $L_t^* \equiv \begin{cases} L_t, & \text{se } t \in \mathfrak{J} \\ T_t & \text{caso contrário} \end{cases}$
- $N_t^* \equiv \sum_{k=t+1}^n L_{t+1}^{*'} \dots L_{k-1}^{*'} Z_k^{*'} F_k^{-1} Z_k^* L_{k-1}^* \dots L_{t+1}^*$

Observa-se que \tilde{X} , no nosso contexto, consiste no logaritmo da informação proveniente do triângulo de *run-off* (vide Tabela 2.3) e \mathfrak{J} se trata do vetor contendo as posições das observações não ausentes do triângulo de *run-off*.

O Teorema 3.1 a seguir, cujo detalhamento da dedução encontra-se no trabalho de Atherino et al. (2010), é necessário para o cálculo da matriz de covariância condicional das parcelas do IBNR.

Teorema 3.1 *Para t, j arbitrários, tem-se que*

$$Cov(x_t, x_j | \tilde{X}) = \begin{cases} 0 & \text{se } t \in \mathfrak{J} \text{ ou } j \in \mathfrak{J} \\ Z_t(P_{t|t-1} - P_{t|t-1} N_{t-1}^* P_{t|t-1}) Z_t' + H_t & \text{se } t = j \text{ e } t, j \notin \mathfrak{J} \\ Z_t P_{t|t-1} L_t^{*'} L_{t+1}^{*'} \dots L_{j-1}^{*'} (I - N_{j-1}^* P_{j|j-1}) Z_j' & \text{se } t < j \text{ e } t, j \notin \mathfrak{J} \end{cases} \quad (3-13)$$

Segundo Atherino (2008), a implementação computacional envolve o armazenamento de algumas matrizes provenientes das recursões do filtro de Kalman: $P_{t|t-1}$ e N_{t-1}^* para todo $t \notin \mathfrak{J}$, e ainda as matrizes $L_\rho^*, L_{\rho+1}^*, \dots, L_{\rho'-1}^*$, $1 \leq \rho < \rho' \leq n$, nas quais ρ e ρ' são o primeiro e o último instantes, respectivamente, em que existam valores faltantes.

3.2.3

Cálculo da reserva IBNR quando os dados contidos no triângulo de run-off estão expostos na forma empilhada

A reserva IBNR consiste na soma dos valores faltantes contidos na série empilhada proveniente do triângulo de *run-off*, ou seja, os valores da série não observados. No contexto de modelos em espaço de estado a estimativa pode ser obtida pela suavização e assim segue que:

$$IBNR = \sum_{t: y_t \text{ é ausente}} E[y_t | \mathbf{y}_n] \quad (3-14)$$

onde $\mathbf{y}_n = y_1, y_2, \dots, y_n$

Um dos diferenciais do modelo de Atherino et al. (2010) é justamente no tratamento dos valores ausentes, de forma que a estimação se torna natural

¹Observe que \tilde{X} está definido como um conjunto de vetores aleatórios e não como um vetor aleatório empilhado. Mas, para o que segue, o mesmo pode ser visto como tal.

mediante a utilização do filtro de Kalman e das equações de suavização (Anexo A).

3.2.4

Reserva IBNR estimada e o Erro Médio Quadrático: Casos total e por ano

Nesta etapa da modelagem se tem a matriz de sinistros IBNR estimados através do modelo Atherino, $E[X|\tilde{X}]$, e a matriz de covariância condicional obtida através do Teorema 3.1. Para se obter o resultado desejado é necessário selecionar dessas matrizes as posições de dados ausentes ($X \ni \tilde{X}$). Assim, como em Atherino (2010), utilizamos um vetor de seleção.

Para um vetor $a = (a_1, a_2, \dots, a_n)'$ qualquer, tem-se que $E(a'X|\tilde{X}) = a'E(X|\tilde{X})$ e $Cov(a'X|\tilde{X}) = a'Cov(X|\tilde{X})a$. A reserva IBNR é calculada escolhendo-se um vetor $a^{(T)}$, de forma que $a_i^{(T)} = 1$ se $i \notin \mathbb{I}$ e $a_i^{(T)} = 0$ caso contrário. Para as parcelas do IBNR relativas a um específico ano de acidente (linhas), utilize o mesmo vetor a com coordenadas nulas nos lugares apropriados - os que não correspondem ao ano. Esta análise por ano de acidente é interessante porque permite a identificação de fontes de incertezas nas parcelas do IBNR (ATHERINO, 2008, p.24).

$$\widehat{\text{IBNR}} \equiv \hat{R} = a'E(X|\tilde{X}) \quad (3-15)$$

$$\text{ep}(\widehat{\text{IBNR}}) \equiv \widehat{\text{ep}}(R) = \sqrt{a'Cov(X|\tilde{X})a} \quad (3-16)$$

É necessário ressaltar que toda teoria aqui apresentada foi proposta para o logaritmo dos dados contidos no triângulo de *run-off*. Desta forma, será necessária uma transformação da escala log-normal para a normal. Sendo assim: após utilizar o método Atherino multivariado para obter $\hat{x}_t \equiv E[x_t|\tilde{X}]$, $\sigma_{\hat{x}_t}^2 \equiv \text{Var}[x_t|\tilde{X}]$ e $\sigma_{\hat{x}_t, \hat{x}_j} \equiv \text{Cov}(x_t, x_j|\tilde{X})$. Calcule:

$$\hat{y}_t = \exp \left[\hat{x}_t + \frac{\sigma_{\hat{x}_t}^2}{2} \right] \quad (3-17)$$

$$\sigma_{\hat{y}_t}^2 = \exp \left[2\hat{x}_t + \sigma_{\hat{x}_t}^2 \right] \left(e^{\sigma_{\hat{x}_t}^2} - 1 \right) \quad (3-18)$$

$$\sigma_{\hat{y}_t, \hat{y}_i}^2 = \exp \left[\hat{x}_t + \hat{x}_i + \frac{\sigma_{\hat{x}_t}^2}{2} + \frac{\sigma_{\hat{x}_i}^2}{2} \right] \left(e^{\sigma_{\hat{x}_t, \hat{x}_i}} - 1 \right) \quad (3-19)$$

3.3

Modelo Atherino bivariado

Uma das contribuições desta tese é formular a extensão do modelo apresentado na seção anterior, que aqui chamaremos de modelo Atherino bivariado.

Como já mencionado, o modelo bivariado proposto tem como objetivo captar as dependências entre os IBNR de uma mesma série “empilhada” (matriz de covariância condicional, Subseção 3.2.2) e as dependências entre os IBNR de duas séries “empilhadas” modeladas de forma conjunta. Para tal, será necessária a introdução da formulação SUTSE.

3.3.1

Formulação SUTSE

A classe de modelos chamada de SUTSE (*seemingly unrelated time series equations*) foi introduzida por Harvey (1989) e posteriormente utilizada por Fernandez e Harvey (1990), Moauro e Savio (2005), Jalles (2009), entre outros.

Quando se modelam séries de sinistros IBNR, individualmente, descartam-se possíveis associações entre diversas coberturas. Em algumas séries empilhadas (Subseção 2.3.4), provenientes do triângulo de *run-off*, pode-se observar o coeficiente de correlação significativo indicando possíveis relações multivariadas entre as séries. Modelos SUTSE em que há componentes comuns ou resíduos correlacionados podem ser aplicados para modelar, quando existirem, as relações entre os sinistros IBNR de coberturas específicas.

O modelo Atherino com nível e componente periódica estocásticas, para duas séries de sinistros IBNR “empilhadas”, na formulação SUTSE é descrito nas equações a seguir:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} + \begin{bmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad (3-20)$$

$$\begin{bmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} \quad (3-21)$$

$$\begin{bmatrix} \gamma_{1,t+1} \\ \gamma_{2,t+1} \end{bmatrix} = \begin{bmatrix} -\sum_{j=1}^{N-1} \gamma_{1,t+1-j} \\ -\sum_{j=1}^{N-1} \gamma_{2,t+1-j} \end{bmatrix} + \begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \quad (3-22)$$

No qual,

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_1\varepsilon_2} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_2}^2 \end{bmatrix} \right] \quad (3-23)$$

$$\begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1\eta_2} \\ \sigma_{\eta_1\eta_2} & \sigma_{\eta_2}^2 \end{bmatrix} \right] \quad (3-24)$$

$$\begin{bmatrix} \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\omega_1}^2 & \sigma_{\omega_1\omega_2} \\ \sigma_{\omega_1\omega_2} & \sigma_{\omega_2}^2 \end{bmatrix} \right] \quad (3-25)$$

Na Equação 3-22, assim como na Equação 3-3, N refere-se ao número de colunas do triângulo de *run-off*, ou seja, o período máximo de desenvolvimento i (vide Tabela 2.1, por exemplo).

Nos modelos SUTSE a dependência entre cada triângulo de *run-off* é capturada através das correlações intra perturbações do modelo: $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, $\eta_t = (\eta_{1,t}, \eta_{2,t})'$ e $\omega_t = (\omega_{1,t}, \omega_{2,t})'$.

De acordo com a Subseção 3.2.1, o modelo Atherino multivariado (Equações 3-20, 3-21 e 3-22) pode ser representado da seguinte forma:

Equação das observações:

$$x_t = Z_t \alpha_t + \varepsilon_t \quad t = 1, 2, \dots, T$$

No qual, para um modelo bivariado, as matrizes x_t e ε_t têm dimensão 2×1 , a matriz Z_t tem dimensão $2 \times m$, sendo m o total de variáveis de estado e α_t possui dimensão $m \times 1$.

$$x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} \quad (3-26)$$

$$Z_t = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (3-27)$$

$$\alpha_t = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \gamma_{1,t} \\ \gamma_{1,t-1} \\ \gamma_{1,t-2} \\ \vdots \\ \gamma_{1,t-N+1} \\ \gamma_{2,t} \\ \gamma_{2,t-1} \\ \gamma_{2,t-2} \\ \vdots \\ \gamma_{2,t-N+1} \end{bmatrix} \quad (3-28)$$

No qual N representa a dimensão da componente periódica - depende do número de colunas do triângulo de *run-off* que será modelado - (vide Tabela 2.1, por exemplo) .

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad (3-29)$$

Equação do estado:

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad t = 1, 2, \dots, T$$

No qual, para um modelo bivariado, a matriz T_t tem dimensão $m \times m$, a matriz R_t tem dimensão $m \times 4$ e η_t possui dimensão 4×1 .

$$\alpha_{t+1} = \begin{bmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \\ \gamma_{1,t+1} \\ \gamma_{1,t} \\ \gamma_{1,t-1} \\ \gamma_{1,t-2} \\ \vdots \\ \gamma_{1,t-N} \\ \gamma_{2,t+1} \\ \gamma_{2,t} \\ \gamma_{2,t-1} \\ \gamma_{2,t-2} \\ \vdots \\ \gamma_{2,t-N} \end{bmatrix} \quad (3-30)$$

$$T_t = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & -1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & & & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \mathbf{I}_{N \times N} & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & & & & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & -1 & \cdots & -1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & & & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \mathbf{I}_{N \times N} & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & & & & 0 \end{bmatrix} \quad (3-31)$$

$$R_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (3-32)$$

$$\eta_t = \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \quad (3-33)$$

No qual $\eta_t \sim N(0, Q)$.

$$Q_t = \begin{bmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1\eta_2} & 0 & 0 \\ \sigma_{\eta_2\eta_1} & \sigma_{\eta_2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\omega_1}^2 & \sigma_{\omega_1\omega_2} \\ 0 & 0 & \sigma_{\omega_2\omega_1} & \sigma_{\omega_2}^2 \end{bmatrix} \quad (3-34)$$

A possibilidade de existência de componentes comuns entre as séries pode ser investigada observando-se, em um modelo SUTSE, o posto da matriz variância-covariância associada a cada um dos vetores de choques das componentes, $\eta_t = (\eta_{t1}, \eta_{t2})$ para a componente nível, e $\omega_t = (\omega_{t1}, \omega_{t2})$ para a componente periódica. Se essas matrizes, Σ_η e Σ_ω , respectivamente, possuem posto maior ou igual a 2 (para o caso $x_t \sim 2 \times 1$; se $x_t \sim p \times 1$ a condição é posto maior ou igual a p) então, isso indica que o melhor modelo a ser aplicado para aquele conjunto de dados é o modelo SUTSE irrestrito, ou seja, cada série tem suas componentes individuais. Quando o posto é menor do que 2 (para o caso $x_t \sim 2 \times 1$) teríamos um modelos SUTSE restrito, com menos componentes, onde a componente de uma série seria uma função afim da outra componente da outra série, daí a denominação de modelos de “componentes comuns”.

Na próxima subseção será apresentada a formulação SUTSE com nível comum (Harvey, 1989). A representação quando há outras componentes comuns

- sazonalidade, ciclo, tendência e periodicidade - segue o mesmo raciocínio.

3.3.2

Formulação SUTSE - nível comum

O reconhecimento de fatores comuns permite o uso de modelos que não somente têm interpretações interessantes, mas também fornecem previsões e inferências mais eficientes, comparadas às fornecidas pelo modelo SUTSE geral.

Um modelo SUTSE bivariado de nível local pode ser descrito a partir das Equações 3-35 e 3-36.

$$\begin{aligned}x_{1,t} &= \mu_{1,t} + \epsilon_{1,t} \\x_{2,t} &= \mu_{2,t} + \epsilon_{2,t}\end{aligned}\quad (3-35)$$

$$\begin{aligned}\mu_{1,t+1} &= \mu_{1,t} + \eta_{1,t} \\ \mu_{2,t+1} &= \mu_{2,t} + \eta_{2,t}\end{aligned}\quad (3-36)$$

Neste caso a matriz de variância e covariância do vetor $\eta_t = (\eta_{1,t}, \eta_{2,t})'$ é dada pela Equação 3-37.

$$\Sigma_\eta = \begin{pmatrix} \sigma_{\eta_1}^2 & \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} \\ \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} & \sigma_{\eta_2}^2 \end{pmatrix}\quad (3-37)$$

A correlação é dada por $\rho_\eta = \frac{\sigma_{\eta_1 \eta_2}}{\sigma_{\eta_1} \sigma_{\eta_2}}$.

Pode-se mostrar que o modelo apresentado nas Equações 3-35, 3-36 e 3-37 pode ser reescrito como:

$$\begin{aligned}x_{1,t} &= \mu_{1,t} + \epsilon_{1,t} \\x_{2,t} &= \pi_\eta \mu_{1,t} + \bar{\mu}_t + \epsilon_{2,t} \\ \mu_{1,t+1} &= \mu_{1,t} + \eta_{1,t} \\ \bar{\mu}_{t+1} &= \bar{\mu}_t + \bar{\eta}_t\end{aligned}$$

no qual a matriz de variâncias e covariâncias é dada por:

$$\text{Cov} \begin{pmatrix} \eta_{1t} \\ \bar{\eta}_t \end{pmatrix} = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & \sigma_{\eta_2}^2 - \rho_\eta^2 \sigma_{\eta_1}^2 \end{pmatrix}\quad (3-38)$$

Prova.

Tem-se que: $\eta_{t,2} = \pi_\eta \eta_{t,1} + \bar{\eta}_t$, no qual $\pi_\eta = \rho_\eta \frac{\sigma_{\eta_2}}{\sigma_{\eta_1}}$, logo:

$$\begin{aligned}
 E[\bar{\eta}_t] &= 0 \\
 \text{Cov}[\eta_{1,t}, \bar{\eta}_t] &= E[\eta_{1,t} \bar{\eta}_t] \\
 &= E[\eta_{1,t}(\eta_{2,t} - \pi_\eta \eta_{1,t})] \\
 &= E[\eta_{1,t} \eta_{2,t}] - \pi_\eta E[\eta_{1,t} \eta_{1,t}] \\
 &= \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} - \rho_\eta \frac{\sigma_{\eta_2}}{\sigma_{\eta_1}} \sigma_{\eta_1}^2 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}[\bar{\eta}_t] &= \text{Var}[\eta_{2,t} - \pi_\eta \eta_{1,t}] \\
 &= \text{Var}[\eta_{2,t}] + \pi_\eta^2 \text{Var}[\eta_{1,t}] - 2\pi_\eta \text{Cov}[\eta_{1,t}, \eta_{2,t}] \\
 &= \sigma_{\eta_2}^2 + \pi_\eta^2 \sigma_{\eta_1}^2 - 2\pi_\eta \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} \\
 &= \sigma_\eta^2 - \rho_\eta^2 \sigma_{\eta_2}^2
 \end{aligned}$$

Por conseguinte:

$$\text{Cov} \begin{pmatrix} \eta_{1,t} \\ \bar{\eta}_t \end{pmatrix} = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & \sigma_{\eta_2}^2 - \rho_\eta^2 \sigma_{\eta_2}^2 \end{pmatrix}$$

Adicionalmente, tem-se:

$$\begin{aligned}
 \mu_{2,t+1} &= \mu_{2,t} + \eta_{2,t} \\
 &= \mu_{2,t} + (\pi_\eta \eta_{1,t} + \bar{\eta}_t) \\
 &= \mu_{2,t} + \pi_\eta (\mu_{1,t+1} - \mu_{1,t}) + \bar{\eta}_t
 \end{aligned}$$

De modo que se pode escrever 3-39:

$$\mu_{2,t+1} - \pi_\eta \mu_{1,t+1} = \mu_{2,t} - \pi_\eta \mu_{1,t} + \bar{\eta}_t$$

Fazendo $\bar{\mu}_t = \mu_{2,t} - \pi_\eta \mu_{1,t}$, tem-se:

$\bar{\mu}_{t+1} = \bar{\mu}_t + \bar{\eta}_t$ e $\mu_{2,t} = \pi_\eta \mu_{1,t} + \bar{\mu}_t$, de modo que:

Então,

$$\begin{aligned}
 x_{2,t} &= \mu_{2,t} + \epsilon_{2,t} \\
 &= \pi_\eta \mu_{1,t} + \bar{\mu}_t + \epsilon_{2,t}
 \end{aligned}$$

Desta forma, o modelo bivariado 3-35 e 3-36 pode ser reescrito como

$$\begin{aligned} x_{1,t} &= \mu_{1,t} + \epsilon_{1,t} \\ x_{2,t} &= \pi_\eta \mu_{1,t} + \bar{\mu}_t + \epsilon_{2,t} \\ \mu_{1,t+1} &= \mu_{1,t} + \eta_{1,t} \\ \bar{\mu}_{t+1} &= \bar{\mu}_t + \bar{\eta}_t \end{aligned}$$

no qual a matriz de variâncias e covariâncias é dada por:

$$Cov \begin{pmatrix} \eta_{1t} \\ \bar{\eta}_t \end{pmatrix} = \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & \sigma_{\eta_2}^2 - \rho_\eta^2 \sigma_{\eta_1}^2 \end{pmatrix}$$

No caso em que $\rho_\eta = \pm 1$ $\eta_{1,t}$ e $\eta_{2,t}$ são perfeitamente correlacionados, o que equivale a dizer que a matriz de variância e covariância de η_t , Σ_η , não tem posto cheio. $\bar{\mu}_t$ é constante, e há só uma tendência comum, que será denotada por μ_t^+ .



O Modelo 3-35 tem a forma:

$$\begin{aligned} x_{1,t} &= \mu_t^+ + \epsilon_{1,t} \\ x_{2,t} &= \pi_\eta \mu_t^+ + \bar{\mu} + \epsilon_{2,t} \\ \mu_{t+1}^+ &= \mu_t^+ + \bar{\eta}_t \end{aligned}$$

De outro modo na forma matricial:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \pi_\eta & 1 \end{bmatrix} \begin{bmatrix} \mu_t^+ \\ \bar{\mu} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \quad (3-39)$$

Se $\pi_\eta = 1$, a tendência da segunda série fica sempre a uma distância constante $\bar{\mu}$ da tendência na primeira séries, ou seja, $\mu_{2,t} = \bar{\mu} + \mu_{1,t}$.

Pré multiplicando o vetor das observações pelo vetor $(-\pi_\eta, 1)$, tem-se:

$$x_{2,t} = \pi_\eta x_{1,t}^+ + \bar{\mu} + \epsilon_t \quad (3-40)$$

no qual $\epsilon_t = \epsilon_{2,t} - \pi_\eta \epsilon_{1,t}$ e $\pi_\eta = \frac{|\rho| \sigma_{\eta_2}}{\rho \sigma_{\eta_1}}$. Quando a condição linear $x_{2,t} - \pi_\eta x_{1,t}$ for estacionária, tem-se que as duas séries são co-integradas.

Na prática, aplica-se a formulação SUTSE apresentada na Equação 3-20 e, posteriormente, calculam-se as correlações entre os resíduos e verifica-se se há componentes em comum. Adicionalmente, no caso de nível comum, utilizam-se

$\sigma_{\eta_1}^2$ e $\sigma_{\eta_2}^2$ (vide Equação 3-34) no cálculo do π e a partir deste ponto modelam-se novamente os dados utilizando a formulação SUTSE específica para o caso com nível comum.

Neste ponto já se tem toda a teoria necessária para se formular o modelo Atherino bivariado. Como já mencionado, este modelo tem a finalidade de estimar de forma conjunta as reservas IBNR (vide Equação 4-33). O modelo será aplicado a duas séries “empilhadas” (vide Subseção 2.3.4) provenientes da reorganização dos dados contidos em triângulos de *run-off*.

Neste modelo bivariado o volume de sinistros ocorridos de uma cobertura é escrito como uma combinação linear do volume de sinistros ocorridos de uma outra cobertura. Cada cobertura apresenta uma distinta componente periódica e essa está diretamente ligada ao número de colunas N que o triângulo de *run-off* apresenta (vide Tabela 2.2).

Cabe ressaltar que o modelo proposto pode ser adaptado facilmente a um número arbitrário de séries (coberturas) que apresentem relações entre si.

De acordo com a Subseção 3.2.1, o modelo Atherino bivariado (formulação SUTSE com componentes comuns) pode ser representado da seguinte forma:

Equação das observações:

$$x_t = Z_t \alpha_t + \varepsilon_t$$

No qual, para um modelo bivariado, as matrizes x_t e ε_t têm dimensão 2×1 , a matriz Z_t tem dimensão $2 \times m$, sendo m o total de variáveis de estado e α_t possui dimensão $m \times 1$.

$$x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} \quad (3-41)$$

$$Z_t = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \pi_\mu & 1 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (3-42)$$

$$\alpha_t = \begin{bmatrix} \mu_{1,t} \\ \bar{\mu} \\ \gamma_{1,t} \\ \gamma_{1,t-1} \\ \gamma_{1,t-2} \\ \vdots \\ \gamma_{1,t-N+1} \\ \gamma_{2,t,2} \\ \gamma_{2,t-1} \\ \gamma_{2,t-2} \\ \vdots \\ \gamma_{2,t-N+1} \end{bmatrix} \quad (3-43)$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad (3-44)$$

N representa a dimensão da componente periódica - depende do número de colunas do triângulo de *run-off* que será modelado (vide Tabela 2.2, por exemplo).

Equação do estado:

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad t = 1, 2, \dots, T$$

No qual, para um modelo bivariado, a matriz T_t tem dimensão $m \times m$, a matriz R_t tem dimensão $m \times 3$ e η_t possui dimensão 3×1 .

em que:

$$\alpha_{t+1} = \begin{bmatrix} \mu_{1,t+1} \\ \bar{\mu} \\ \gamma_{1,t+1} \\ \gamma_{1,t} \\ \gamma_{1,t-1} \\ \gamma_{1,t-2} \\ \vdots \\ \gamma_{1,t-N} \\ \gamma_{2,t+1} \\ \gamma_{2,t} \\ \gamma_{2,t-1} \\ \gamma_{2,t-2} \\ \vdots \\ \gamma_{2,t-N} \end{bmatrix} \quad (3-45)$$

$$T_t = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & -1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & & & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \mathbf{I}_{N \times N} & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & & & & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & -1 & \cdots & -1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & & & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \mathbf{I}_{N \times N} & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & & & & 0 \end{bmatrix} \quad (3-46)$$

$$R_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} \quad (3-47)$$

$$\eta_t = \begin{bmatrix} \eta_t^+ \\ \omega_{1,t} \\ \omega_{2,t} \end{bmatrix} \quad (3-48)$$

sendo $\eta_t \sim N(0, Q)$.

$$Q_t = \begin{bmatrix} \sigma_{\eta_t^+}^2 & 0 & 0 \\ 0 & \sigma_{\omega_1}^2 & \sigma_{\omega_1\omega_2} \\ 0 & \sigma_{\omega_2\omega_1} & \sigma_{\omega_2}^2 \end{bmatrix} \quad (3-49)$$

No modelo $x_{1,t}$ e $x_{2,t}$ representam os logaritmos das duas coberturas², π_μ relaciona os níveis das duas coberturas através das variâncias e covariâncias dos resíduos presentes nas equações dos níveis (quando essas não são comuns), sendo $\pi_\mu = \rho_{\eta} \frac{\sigma_{\eta_2}}{\sigma_{\eta_1}}$. Adicionalmente, $\gamma_{1,t}$ e $\gamma_{2,t}$ denotam as periodicidades das coberturas relacionadas, assim como na Equação 3-3. η_t descreve as componentes irregulares do nível e das periodicidades e Q_t apresenta as variâncias e covariâncias das equações que descrevem a evolução do nível e da periodicidade.

Em função da dificuldade na programação do modelo Atherino bivariado, assim como outras formulações em espaço de estado multivariadas (Lima, 2010), optou-se por utilizar uma abordagem univariada para a problemática em questão. As adaptações da teoria univariada, primeiramente, apresentada no artigo de Koopman e Durbin (2000) serão expostas no Anexo A.

O cálculo da matriz de covariância condicional das parcelas do IBNR³

²Vide capítulo de aplicação (5).

³Vide Equação 4-33.

modeladas de forma bivariada (Equações 3-41 e 3-45) é feito simultaneamente (Subseção 3.2.2). O cálculo das reservas IBNR e do erro médio quadrático é semelhante a forma apresentada na Subseção 3.2.3, utilizando as Equações 3-15 e 3-16.

3.4

Análise do poder preditivo - medidas de aderência

O poder preditivo de um modelo ajustado a uma base de dados pode ser checado através da comparação de valores reais dos dados com os valores previstos dos mesmos a partir de um modelo adequado.

Como medidas de aderência para avaliar a qualidade de ajuste do modelo “*in sample*” e “*out of sample*” utilizou-se o MAPE, o MSE, o Pseudo e o R^2 apresentados a seguir.

$$\text{MAPE} = \frac{1}{n} \sum_{t=d+1}^n \left(\frac{|y_t - \hat{y}_{t|t-1}|}{|y_t|} \right) * 100, \quad |y_t| \neq 0 \quad (3-50)$$

$$\text{Pseudo}R^2 = [\text{Corr}(y_t, \hat{y}_{t|t-1})]^2 \quad (3-51)$$

$$\text{EQM} = \frac{1}{n-d} \sum_{t=d+1}^n (y_t - \hat{y}_{t|t-1})^2 \quad (3-52)$$

onde $d = \dim(\delta)$ é a dimensão não estacionária do vetor de estado; y_t representa cada observação de uma dada cobertura e $\hat{y}_{t|t-1}$ representa a previsão um passo a frente. Vale ressaltar que nos interessa avaliar o poder preditivo do modelo referente a variável original, y_t e não $x_t = \log(y_t)$, logo utilizaremos y_t nas medidas de aderência. Diferentemente da análise residual que será feita, para o modelo Atherino, utilizando a variável $x_t = \log(y_t)$.

4

Modelos GAS para triângulo de run-off

Modelos GAS (*Generalized Autoregressive Score*), também conhecidos como *Dynamic Conditional Score* (DCS), desenvolvidos por Creal et al. (2013), fornecem um arcabouço geral para especificar e estimar modelos de séries temporais e dados em painel com parâmetros variantes no tempo e consecutivamente calcular as reservas IBNR. A facilidade na estimação e a inferência são características principais desses modelos. A partir desse arcabouço é possível especificar e estimar qualquer distribuição condicional que se julgue adequada para caracterizar as entradas do triângulo de *run-off*.

Nos modelos GAS, a verossimilhança está disponível na forma fechada, o que não é o caso para modelos em espaço de estado com distribuições não Gaussianas. Além disso, a estrutura completa da distribuição preditiva é transportada para o mecanismo de atualização dos parâmetros variantes no tempo, via função *score*.

Alguns modelos bastante conhecidos e utilizados na literatura podem ser vistos como casos particulares dos modelos GAS. Como exemplos cabem citar: os modelos autoregressivos condicionalmente heterocedásticos (GARCH) de Engle e Bollerslev (1986), os modelos autoregressivos de intensidade e duração condicional, de Engle e Russell (1998) e os modelos de erro multiplicativo (MEM) de Engle (2002).

A segunda classe de modelos propostos para o cálculo das reservas IBNR é o GAS. Na próxima seção, será apresentado o modelo GAS e, na Seção 4.2 a sua adaptação para dados em painel, que será utilizada na modelagem univariada do triângulo de *run-off* para estimação da reserva IBNR.

4.1

Modelos GAS - especificações e propriedades

Seja y_t a variável dependente de interesse, f_t o vetor com parâmetros variantes no tempo e ψ o vetor com os parâmetros fixos e desconhecidos do modelo. Adicionalmente, define-se os conjuntos de informação $Y^t = (y_1, y_2, \dots, y_t)$ e $F^t = (f_0, f_1, \dots, f_t)$. Assim teremos que a informação disponível até o instante $t - 1$ será dada por:

$$F_{t-1} = \{Y^{t-1}, F^{t-1}\}, \quad \text{para } t = 1, 2, \dots, T. \quad (4-1)$$

Supondo y_t um processo estocástico com espaço de estado contínuo, sua densidade condicional é representada por:

$$y_t \sim p(y_t | f_t, F_{t-1}; \psi) \quad (4-2)$$

O mecanismo de atualização do vetor de parâmetros variantes no tempo, f_t , é dado por uma equação semelhante à adotada nos modelos ARMA, resultando em um modelo GAS (p,q) definido a seguir.

$$f_{t+1} = \omega + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1} \quad (4-3)$$

Onde ω é um vetor de constantes, as matrizes A_i e B_j possuem dimensões adequadas, determinadas pela dimensão de f_t , enquanto s_t é uma função dos dados passados, $s_t = s_t(y_t, f_t; \psi)$, a ser detalhada.

A estimativa dos fatores no tempo, f_t , dado as observações passadas, $Y_{t-1} = y_1, \dots, y_{t-1}$, para um valor de ψ , é semelhante ao processo de filtragem (Seção A.2). Quando a observação y_t torna-se disponível computa-se s_t como definido na Equação 4-4. Na prática, o processo de filtragem é iniciado no instante $t = 1$ com f_1 sendo um valor fixo ou até mesmo sendo estimado, fazendo parte do vetor ψ . Esse assunto será abordado com mais detalhes na Subseção 4.1.1.

Assim como em Creal et al (2008), define-se s_t , presente na Equação 4-3 como sendo:

$$s_t = S_t(F_{t-1}) \cdot \nabla_t = S_t \nabla_t \quad (4-4)$$

O vetor *score*, ∇_t , contido na Equação 4-4 é dado por:

$$\nabla_t = \frac{\partial \log p(y_t | f_t, F_{t-1}; \psi)}{\partial f_t} \quad (4-5)$$

S_t é uma matriz de ponderação obtida a partir da matriz de informação de Fisher, sendo dada por:

$$S_t = I_{t|t-1}^{-d} \quad d = 0, \frac{1}{2}, 1. \quad (4-6)$$

No qual,

$$\begin{aligned} I_{t|t-1} &= E_{t-1}(\nabla_t \nabla_t') \\ &= -E_{t-1} \left[\left(\frac{\partial^2 \log p(y_t | f_t, F_{t-1}; \psi)}{\partial f_t \partial f_t'} \right) \right] \end{aligned} \quad (4-7)$$

Sendo $E_{t-1}(\cdot) = E_{t-1}(\cdot | f_t, F_{t-1}; \psi)$. O mecanismo de ponderação do vetor *score* presente na Equação 4-4 é uma relevante característica dos modelos GAS. Escolhendo-se $d = 0$ obtém-se a matriz identidade ($S_t = I$) com a qual o mecanismo de atualização passa a utilizar somente o gradiente para a atualização dos parâmetros variantes no tempo. Se $d = 1/2$, então pode-se mostrar que $E_{t-1}(s_t) = 0$ e $V_{t-1}(s_t) = 1$, ou seja, s_t é um processo diferença martingale com variância unitária.

4.1.1

Inicialização do vetor de parâmetros variantes no tempo

Para iniciar a otimização da função de verossimilhança com relação aos parâmetros fixos do modelos GAS, é necessário que se escolha os valores iniciais de cada parâmetro e um valor inicial para f_1 , que constitui o início de todo o processo de atualização, como pode ser visto na Equação 4-3. Para o caso do vetor de parâmetros variantes no tempo, f_t , supondo um modelo GAS (1,1), segue que:

$$E[f_{t+1}] = \omega + A_1 E[s_t] + B_1 E[f_t] \quad (4-8)$$

Dado que $E[s_t | F_{t-1}] = 0$, segue que $E[s_t] = 0$, e assim temos que:

$$E[f_{t+1}] = \omega + B_1 E[f_t] \quad (4-9)$$

Supondo estacionariedade para o processo f_t , segue que $E[f_{t+1}] = E[f_t]$, o que implica que $E[f_t] = \omega / (1 - B_1)$, $B_1 \neq 1$. Assim, toma-se $f_t = E[f_t]$, e portanto:

$$f_1 = \frac{\omega}{1 - B_1}, \quad B_1 \neq 1. \quad (4-10)$$

Na estimação do vetor de parâmetros fixos dos modelos GAS, a inicialização do algoritmo de otimização será realizada a partir de um método híbrido, utilizando algoritmos genéticos. Esse será brevemente descrito na próxima subseção.

4.1.2

Algoritmos genéticos para geração de valores iniciais

Algoritmos genéticos (AGs) são métodos de otimização inspirado nos conceitos da teoria de seleção natural (princípio Darwiniano), originando-se de conceitos baseados nos processos genéticos para procurar soluções ótimas ou sub-ótimas para um dado problema. Essa classe de algoritmos evolutivos se tornou conhecida com o trabalho de John Holland na década de 70. A partir de então, muitos trabalhos foram publicados acerca dos fundamentos teóricos dos algoritmos genéticos, como por exemplo: Goldberg (1989), Rawlins (1991), Whitley (2000), Whitley e Vose (1995).

Nos AGs cada “indivíduo” representa uma possível solução para um dado problema. Cada possível solução de um problema é codificada em uma estrutura chamada de “cromossomo”, composta por uma cadeia de *bits* ou símbolos. Os “cromossomos” representam indivíduos, que são evoluídos ao longo de várias gerações, de acordo com os princípios de seleção natural e sobrevivência. Os indivíduos são, então, submetidos ao processo evolucionário que envolve avaliação, seleção, recombinação ou *crossover* e mutação.

Nesta tese utilizou-se algoritmos genéticos com a finalidade de determinar o melhor vetor de parâmetros iniciais na estimação do modelo GAS. Para tal, o pacote GA do R-project foi empregado. A inicialização adotada foi a aleatória uniforme, na qual cada gene do indivíduo recebe um valor sorteado de forma aleatória a partir de uma distribuição uniforme.

Considere ψ uma matriz de dimensão $m \times p$, na qual m é o número de elementos na população de valores iniciais, p é o número de parâmetros do modelo GAS, k o número de iterações do AG, a e b parâmetros da distribuição uniforme, y a série (Casco ou RCVF no caso desta tese) e lv a log-verossimilhança. A execução do algoritmo envolve inicialmente a definição dos parâmetros descritos acima.

O pseudo algoritmo para a geração de valores iniciais do vetor ψ utilizado nesta tese será apresentado a seguir:

Figura 4.1: Algoritmo para a geração de valores iniciais.

Algoritmo 1 Geração de Valores Iniciais por meio de Algoritmo Genéticos

```

1: procedimento ALGORITMO GENETICO( $lv(\cdot), m, p, K, a, b, y$ )
2:   Gerar uma população de candidatos a solução
3:   para  $i \leftarrow 1$  até  $m$  faça
4:     para  $j \leftarrow 1$  até  $p$  faça
5:        $\psi[i, j] \leftarrow U(a[j], b[j])$            ▷ Sorteio Aleatório da Uniforme
6:     fim para
7:     Avaliar a função de verossimilhança  $lv(\cdot|y)$  no vetor  $\psi$  gerado
8:   fim para
9:   Selecionar as 5% melhores soluções
10:  para  $k \leftarrow 1$  até  $K$  faça
11:    realizar mutação
12:    realizar crossover
13:    gerar uma nova população de soluções
14:    avaliar a log-verossimilhança nas soluções geradas
15:    retorna Selecionar as 5% melhores soluções
16:  fim para
retorna  $\psi_{otimo}$ 

```

Os valores dos parâmetros estáticos (vetor ψ) obtidos pelo AG serão utilizados para inicializar o método clássico de otimização (BFGS).

4.1.3 Diagnósticos

Um passo crucial na modelagem estatística de dados é a investigação da probabilidade do modelo utilizado ser uma aproximação razoável para o verdadeiro mecanismo probabilístico gerador dos dados. Nessa tese os diagnósticos para a avaliação dos modelos são baseados em testes realizados nos resíduos quantílicos e/ou nos resíduos de Pearson.

De acordo com Dunn e Smyth (1996) o resíduo quantílico teórico para um modelo estatístico é definido como:

$$R_{t,\psi} = \Phi^{-1}\{F(y_t|f_t, F_{t-1}; \psi)\} \quad t = 1, 2, \dots, T \quad (4-11)$$

A sua versão observada quando substituimos ψ por sua estimativa de máxima verossimilhança $\hat{\psi}$ é representado por:

$$r_{t,\hat{\psi}} = \Phi^{-1}\{F(y_t|f_t, F_{t-1}; \hat{\psi})\} \quad (4-12)$$

Onde Φ^{-1} é a função de distribuição acumulada de uma normal padrão, $F(y_t|F_{t-1}, \hat{\psi})$ é a função de distribuição condicional do processo y_t apresentado na Equação 4-14 e $\hat{\psi}$ é a estimativa do vetor ψ obtido por máxima verossimilhança.

O resíduo de Pearson é definido como:

$$r_{p,t} = \frac{y_t - E[y_t|f_t, F_{t-1}; \hat{\psi}]}{\sqrt{V[y_t|f_t, F_{t-1}; \hat{\psi}]}} \quad (4-13)$$

A ausência de valores em muitas posições do triângulo de *run-off* dificulta a análise residual. Para realizar os testes de autocorrelação serial, heterocedasticidade condicional e normalidade construiu-se uma série empilhando por linhas¹ os resíduos dos triângulos de *run-off* e não considerou-se as posições na qual existem valores ausentes.

Adicionalmente, para avaliar se nos resíduos do modelo GAS encontram-se posições mal ajustadas ou *outliers*, utiliza-se, assim como em De Jong (2004), uma análise simplificada dos resíduos em painel. Observa-se a proporção, em relação ao total, dos resíduos que apresenta valores superiores a 2,0.

4.2

Modelos GAS - adaptação para série em painel

A utilização dos modelos GAS para a modelagem dos dados em forma de triângulo de *run-off* será efetuada a partir da sua versão para dados em painel, conforme apresentado em Creal et al. (2014). Como veremos a seguir, os sinistros IBNR expostos no triângulo de *run-off* podem ser considerados neste formato, uma vez que é de se esperar que haja dependência nos dados ao longo dos períodos de origem (t) e ou ao longo dos períodos de desenvolvimento (i).

4.2.1

Especificações dos modelos GAS (1,1) para dados em painel

De acordo com Creal et al. (2014), supõe-se uma série temporal em painel, $y_{t,i}$, com $t = 1, \dots, T$ e $i = 1, \dots, N$. Por hipótese cada célula do triângulo de *run-off* (t, i) é modelada por uma mesma densidade condicional conforme a seguir:

¹

O empilhamento por linhas foi apresentado na Subseção 2.3.4. No caso dos modelos Atherino univariado e bivariado empilhou-se os dados de sinistros IBNR e logo após aplicou-se o modelo. No caso dos modelos GAS, os dados foram modelados como dados em painel e empilhou-se posteriormente, os resíduos para análise residual.

$$y_{t,i} \sim p(y_{t,i}|f_t, F_{t-1}, \psi) \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad (4-14)$$

No qual, semelhante à formulação geral dos modelos GAS, tem-se que:

$y_{t,i}$: montante de sinistros ocorrido em t e avisados em i , $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$;

f_t : vetor de parâmetros variantes no tempo;

F_{t-1} : toda a informação disponível em t ;

ψ : vetor de parâmetros fixos.

Observe que, por construção, a variação no tempo dos parâmetros do modelo ocorre apenas ao longo do período de origem t , $t = 1, 2, \dots, T$.

Para completar a especificação adequada aos modelos GAS, para os dados do triângulo de *run-off*, vamos assumir que os sinistros IBNR, os $y_{t,i}$'s, em cada tempo t , são *cross-section* independentes, condicional à f_t e ao conjunto de informações F_{t-1} . Os detalhes dessa formulação serão apresentados posteriormente, para as densidades gama e log-normal.

4.2.2

Estimação por Máxima Verossimilhança - adaptada ao triângulo de run-off

Uma propriedade conveniente dos modelos GAS é a facilidade de avaliação da função de verossimilhança. Adota-se a decomposição do erro padrão de predição. Assim como descrito em Creal et al. (2011), para uma série temporal y_1, y_2, \dots, y_T , o estimador de máxima verossimilhança, para os parâmetros fixos do modelo, ψ , pode ser expresso como:

$$\hat{\psi} = \arg \max_{\psi} \sum_{t=1}^T l_t \quad (4-15)$$

No qual $l_t = \log p(y_t|F_{t-1}; \psi)$ para uma dada realização y_t . Para a avaliação da função de verossimilhança, é suficiente implementar o mecanismo de atualização dado pela Equação 4-3 e avaliar o logaritmo da verossimilhança, l_t , em um valor particular ψ^* de ψ .

Será necessário adaptar a verossimilhança tradicional aos dados do triângulo de *run-off*. Sob a hipótese de independência *cross-section*, dado f_t e F_{t-1} , a densidade em cada tempo t será dada por:

$$p(y_t|\psi, F_{t-1}) = \prod_{i=1}^N p(y_{t,i}|f_t, F_{t-1}; \psi) \quad (4-16)$$

Uma vez que a função de verossimilhança é dada por:

$$L(\psi) = \prod_{t=1}^T p(y_t | f_t, F_{t-1}; \psi) \quad (4-17)$$

em vista de (4-16), segue que:

$$\begin{aligned} l(\psi) &= \log(L(\psi)) \\ &= \sum_{t=1}^T \log p(y_t | f_t, F_{t-1}; \psi) \\ &= \sum_{t=1}^T \sum_{i=1}^N \log p(y_{t,i} | f_t, F_{t-1}; \psi) \end{aligned} \quad (4-18)$$

Uma variável *dummy* para identificar a existência ou a ausência de valores em cada posição do triângulo de *run-off* é incluída na verossimilhança, resultando em:

$$l(\psi) = \sum_{t=1}^T \sum_{i=1}^N \delta_{t,i} \log p(y_{t,i} | f_t, F_{t-1}; \psi) \quad (4-19)$$

onde:

$$\delta_{t,i} = \begin{cases} 1, & \text{se } y_{t,i} \text{ é observável;} \\ 0, & \text{caso contrário.} \end{cases} \quad (4-20)$$

Sob certas condições de regularidade, o estimador de máxima verossimilhança $\hat{\psi}$ de ψ é consistente e possui distribuição assimétrica normal.

$$\sqrt{T}(\hat{\psi} - \psi) \longrightarrow^d N(0, W^{-1}) \quad (4-21)$$

No qual $W = -E[\frac{\partial^2 l}{\partial \psi \partial \psi'}]$ e $l = \sum_{t=1}^T l_t$.

Maiores detalhes sobre a normalidade assintótica dos estimadores de máxima verossimilhança pesquisar em Blasques et al. (2014).

4.2.3

Intervalo de confiança in sample para os parâmetros estáticos do modelo GAS(1,1)

O método para o cálculo dos intervalos de confiança (IC) suficientemente preciso e rápido em modelos GAS e que incorpore a incerteza ao vetor de parâmetros estáticos do modelo foi desenvolvido por Blasques et al. (2015) e chamado de “*simulation-based bands*”.

Os autores do artigo propõem intervalos de confiança para \hat{f}_t , de forma a refletir a incerteza dos parâmetros estáticos contidos no vetor $\hat{\psi}$. No instante que se obtém o intervalo de confiança para os \hat{f}_t s e para os $\hat{\lambda}_i$ s torna-se possível o cálculo dos intervalos de confiança para as reservas IBNR (Equações 4-33 e 4-34).

4.2.3.1

Simulation-based bands

Desde que os intervalos de confiança reflitam apenas a incerteza contida no vetor que engloba os estimadores de máxima verossimilhança, $\hat{\psi}_T$, pode-se obter intervalos exatos gerando valores $\hat{\psi}^j$ da distribuição de $\hat{\psi}_T$ e em seguida calcular o filtro de $t = 1$ a $t = T$ para cada valor simulado $\hat{\psi}^j$. Contudo, uma vez que a distribuição amostral finita de $\hat{\psi}_T$ é desconhecida, a solução é simular valores da distribuição assintótica.

$$\hat{\psi}_T^j \sim N(\hat{\psi}_T, T^{-1}\hat{W}), \quad j = 1, \dots, M \quad (4-22)$$

Onde M é o número de simulações e W é a matriz de variâncias e covariâncias de $\hat{\psi}$. Para cada $\hat{\psi}^j$, $j = 1, \dots, M$, calcula-se a sequência $\hat{f}_1^j, \hat{f}_2^j, \hat{f}_3^j, \dots, \hat{f}_T^j$ (vide Equação 4-3).

A incerteza de $\hat{\psi}_T$ é caracterizada por sua distribuição assintótica, conforme Equação 4-21. Para obter o intervalo de confiança basta calcular os percentis desejados para cada t , a partir dos M cenários gerados pela simulação de Monte Carlo para os \hat{f}_t^j .

4.2.4

Previsão fora da amostra - efeito linha no triângulo de run-off

Como já mencionado, o triângulo de *run-off* é uma representação estática, em uma determinada data, dos sinistros IBNR associados a períodos de origem e a períodos de desenvolvimento (Subseção 2.5). À medida que se avança no tempo, as linhas e as colunas da matriz aumentam conforme o acréscimo de novos dados de sinistros à base das seguradoras. A previsão destes valores são de interesse da seguradora, uma vez que eles constituem parte da reserva total que ela deverá provisionar no próximo instante de tempo.

Os valores presentes na extensão do efeito linha - também conhecido como reserva de prêmio (quando $t = T, T + 1, \dots, \tau$) apresentados no Capítulo 2 representam os sinistros “IBNR futuros”, ou seja, sinistros que ocorrerão e que, por algum motivo, não serão imediatamente informados à seguradora.

Quando t representar uma linha do triângulo superior a T será chamado de t^* , sendo $t^* = T + 1, T + 2, \dots, \tau$. O cálculo destes “IBNR futuros” é dado pela equação:

$$\text{IBNR}_{t^*,i} = E[y_{t^*,i}|f_t, F_{t-1}; \psi], \quad t^* = T + 1, T + 2, \dots, \tau. \quad (4-23)$$

Como nas linhas t^* todas as posições são ausentes de informações, $I_{t|t-1}$

e ∇_t contidos na Equação 4-41 serão iguais a zero. O instante de tempo $T + 1$ é a única exceção, neste período, temos as informações $I_{T|T-1}$ e ∇_T .

As projeções não serão feitas para colunas adicionais ao triângulo de *run-off* ($i > N$). Avaliando os triângulos disponíveis observou-se que no período de desenvolvimento N os sinistros reportados têm proporções muito pequenas do total pago no primeiro período de desenvolvimento, menos de 0,5%, sendo assim, acredita-se que os avisos em períodos superiores a N podem ser desprezados sem maiores perdas.

[...] Não é prudente que se faça uma extrapolação do efeito linha por horizontes muito longos, pois o poder preditivo dos modelos pode ser fortemente comprometido, dado que as variáveis explicativas, geralmente utilizadas para modelagem de sinistros IBNR, não contêm informações quanto a mudanças de política da empresa e nem da economia (COSTA, 2010, p. 23).

É importante ressaltar que um arcabouço geral para estimação de reservas IBNR é a utilização dos modelos GAS (CREAL et al. (2013)). A partir desse arcabouço é possível especificar e estimar qualquer distribuição condicional que se julgue adequada para caracterizar as entradas do triângulo de *run-off*. Outro ponto importante é que, embora em nossa tese tenha sido proposta a dinâmica dos parâmetros da distribuição f_t ao longo dos anos de origem t , é possível também assumir que essa dinâmica ocorra ao longo dos anos de desenvolvimento i e até de forma conjugada. Será necessário averiguar se o modelo proposto captou a dependência de forma correta através da análise residual, por exemplo.

4.3

Modelos GAS e densidades condicionais

Em se tratando da modelagem de triângulos de *run-off* via modelos GAS, a escolha da distribuição preditiva e dos parâmetros variantes no tempo devem refletir as distribuições contínuas definidas no domínio dos reais positivos. As distribuições gama e log-normal são algumas das distribuições que possuem suportes adequados às séries dispostas nos triângulos de *run-off* utilizadas nesta tese (Tabelas 5.1 e 5.2).

Em seus artigos Verrall (1989), Taylor (2003), De Jong (2004) e Atherino (2010) adotaram a suposição de log-normalidade dos dados. Já Verrall (2000) e Huang (2012) assumiram que os dados contidos nos triângulos apresentavam distribuição gama.

Por essas distribuições já terem sido adotadas na literatura atuarial e por apresentarem suportes adequados para os dados IBNR, optou-se por empregá-las nesta tese.

4.3.1

Densidade gama adaptada ao triângulo de run-off

Por construção o parâmetro β varia de acordo com cada coluna i , portanto, $\beta \rightarrow \beta_i$, $i = 1, 2, 3, \dots, N$. Por sua vez, o parâmetro α varia de acordo com cada linha t , logo, $\alpha \rightarrow \alpha_t$, $t = 1, 2, 3, \dots, T$.

A densidade para cada $y_{t,i}$ será dada pela Equação 4-24.

$$p(y_{t,i}|\alpha_t, \beta_i, F_{t-1}) = \frac{y_{t,i}^{\alpha_t-1} e^{-\frac{y_{t,i}}{\beta_i}}}{\beta_i^{\alpha_t} \Gamma(\alpha_t)}, \quad y_{t,i} > 0 \quad (4-24)$$

A esperança e a variância de uma distribuição gama são dadas a seguir:

$$E[y_{t,i}|\alpha_t, \beta_i, F_{t-1}] = \alpha_t \beta_i = e^{f_t + \lambda_i} \quad (4-25)$$

$$V[y_{t,i}|\alpha_t, \beta_i, F_{t-1}] = \alpha_t \beta_i^2 \quad (4-26)$$

O logaritmo da densidade (Equação 4-24) é dado pela equação:

$$\log p(y_{t,i}|\alpha_t, \beta_i, F_{t-1}) = (\alpha_t - 1) \log(y_{t,i}) - \frac{y_{t,i}}{\beta_i} - \alpha_t \log(\beta_i) - \log \Gamma(\alpha_t) \quad (4-27)$$

Adicionalmente, tem-se que:

$$\alpha_t \beta_i = e^{f_t + \lambda_i} \quad (4-28)$$

No qual $\lambda_i \in (-\infty, +\infty)$ e $\beta_i = e^{\lambda_i}$. A Equação 4-28 será utilizada nos cálculos das reservas IBNR. Assim como no modelo Atherino, para cada célula do triângulo inferior, as reservas IBNR serão dadas pelas esperanças condicionais e a reserva total será dada pelo somatório dessas esperanças.

4.3.1.1

Dinâmica de atualização - distribuição gama

Como visto na Seção 4.1, para o desenvolvimento do modelo GAS (1,1) será necessário calcular ∇_t e $I_{t|t-1}$. Assim como em Creal et al. (2013), segue que:

$$\nabla_t = \sum_{i=1}^N \delta_{t,i} \nabla_{t,i} = \sum_{i=1}^N \delta_{t,i} \frac{\partial \log p(y_{t,i} | \alpha_t, \beta_i, F_{t-1})}{\partial f_t} \quad (4-29)$$

$$I_{t|t-1} = E_{t|t-1} [\nabla_t \nabla_t'] = \sum_{i=1}^N \delta_{t,i} E_{t|t-1,i} [\nabla_{t,i} \nabla_{t,i}'] \quad (4-30)$$

Onde $\delta_{t,i}$ é definido de acordo com (4-20).

O modelo GAS(1,1) com distribuição gama para modelar a estrutura de dependência entre os dados contidos no triângulo de *run-off* é apresentado a seguir.

$$f_{t+1} = \omega + A_1 s_t + B_1 f_t \quad (4-31)$$

Onde:

$$s_t = I_{t|t-1}^{-d} \nabla_t, \quad d = 0, \frac{1}{2} \text{ ou } 1.$$

Quando $d = \frac{1}{2}$, tem-se:

$$\begin{aligned} s_t &= I_{t|t-1}^{-\frac{1}{2}} \nabla_t \\ &= \frac{1}{\sqrt{\sum_{i=1}^N \delta_{t,i} (\alpha_t)^2 \psi_1(\alpha_t)}} \sum_{i=1}^N \delta_{t,i} \alpha_t [\log(y_{t,i}) - \log(\beta_i) - \psi(\alpha_t)] \end{aligned}$$

Onde $\psi(\alpha_t) = \partial \log \Gamma(\alpha_t) / \partial \alpha_t$, ou seja, a função digama é a primeira derivada logarítmica da função gama e a função trigama é a segunda derivada logarítmica da função gama, $\psi_1(\alpha_t) = \partial^2 \log \Gamma(\alpha_t) / \partial \alpha_t^2$.

Adicionalmente, se $y \sim \text{gama}(\alpha, \beta)$, então: $E[\log y] = \psi(\alpha) + \log(\beta)$ e $V[\log y] = \psi_1(\alpha)$. Esse resultado será útil para provar que $E[s_t | F_{t-1}] = 0$ e $E[s_t] = 0$ (vide Subseção 4.1.1).

4.3.1.2

Cálculo do IBNR assumindo distribuição gama

Semelhante ao modelo Atherino (Subseção 3.2.3) os sinistros estimados para cada posição do triângulo de *run-off* é dado pela esperança condicional.

$$\text{IBNR}_{t,i} = E[y_{t,i} | f_t, F_{t-1}; \psi] = e^{f_t + \lambda_i} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (4-32)$$

A reserva IBNR por linhas consiste na soma dos valores faltantes de cada linha t do triângulo de *run-off*, ou seja, para cada período de origem t , tem-se

a estimativa do IBNR a ser pago pela seguradora até o período N (vide Tabela 2.2 , por exemplo).

$$\text{IBNR}_t = \sum_{i:\text{é ausente}} e^{f_t + \lambda_i} \quad (4-33)$$

A reserva IBNR total será dada pelo somatório das reservas IBNR por linhas, como pode ser visto na equação a seguir.

$$\text{IBNR}_{Total} = \sum_{t=1}^T \sum_{i:\text{é ausente}} e^{f_t + \lambda_i} \quad (4-34)$$

De acordo com a Equação 4-32, para a estimação da reserva IBNR será necessário o cálculo dos λ'_t s e dos f'_t s. Os λ'_t s são obtidos através da relação $\beta_i = e^{\lambda_i}$ e os f'_t s (Equação 4-3) dependem, dentre outros parâmetros, dos parâmetros estáticos ω , A_1 e B_1 que precisam ser estimados. Para estimar os parâmetros desconhecidos e estáticos do modelo, $\psi = (\omega, A_1, B_1, \lambda_1, \dots, \lambda_N)$, deve-se maximizar o logaritmo da função verossimilhança conforme a Equação 4-20.

Além da estimação das reservas IBNR via modelos GAS, estimou-se o coeficiente de variação e o intervalo de confiança para as reservas IBNR. O algoritmo para estimação dos coeficientes de variação e dos intervalos de confiança serão descritos nas subseções a seguir.

4.3.1.3

Cálculo do coeficiente de variação para a distribuição gama

Conforme visto nas subseções anteriores, o modelo GAS (1,1) um passo à frente está bem definido e assim a verossimilhança é conhecida. Contudo, para calcular o coeficiente de variação das reservas estimadas precisa-se obter as distribuições dos valores ausentes do triângulo de *run-off* e para tal será necessário projetar o modelo por mais de um passo à frente, e nesse caso a densidade não é mais conhecida, tendo que ser obtida através de simulação.

As variâncias das reservas IBNR estimadas foram obtidas conforme descrito abaixo:

Passo 1: Cada célula vazia do triângulo de *run-off* é modelado por uma variável aleatória gama (α_t, β_i) , no qual $\alpha_t = e^{f_t}$ e $\beta_i = e^{\lambda_i}$;

Passo 2: Usando os parâmetros α_t e β_i estimados para cada célula vazia, simula-se uma observação da distribuição gama para cada posição do triângulo;

- Passo 3:** Ao final reserva-se as somas dessas observações simuladas para cada célula vazia. Neste ponto tem-se uma realização da soma;
- Passo 4:** Repete-se os passos 2 e 3 dez mil vezes. Ao final se terá, para cada célula, dez mil valores - a média aritmética desses valores, para cada célula, converge em probabilidade para a média da gama para cada célula do triângulo (Equação 4-32);
- Passo 5:** Soma-se sequencialmente esses valores sobre as células vazias, e assim, ao final, têm-se dez mil valores da variável soma, que serão as realizações do IBNR;
- Passo 6:** A média amostral das somas obtidas no passo 5 e a variância amostral (a fórmula usual dos livros textos) serão utilizadas no cálculo do coeficiente de variação.

4.3.1.4

Cálculo do intervalo de confiança para a distribuição gama

Conforme Blasques et al. (2015), os intervalos de confiança para as estimativas do IBNR para cada t e para o IBNR total (Subseção 4.2.3) foram calculados utilizando o seguinte algoritmo:

Passo 1: Após estimar o vetor de parâmetros $\hat{\psi} = (\hat{\omega}, \hat{A}_1, \hat{B}_1, \hat{\lambda}_1, \dots, \hat{\lambda}_N)$ e a matriz Hessiana $W(\hat{\psi})$ via máxima verossimilhança adaptada ao triângulo de *run-off*, assim como apresentado da Subseção 4.2.2, gera-se, através de uma distribuição normal multivariada, M vetores de parâmetros estáticos, conforme a Equação 4-22;

Passo 2: A partir dos valores simulados de $\hat{\psi}$ calcula-se $\hat{f}_t^{(j)}$ e reserve:

$$\hat{f}_{t+1}^{(j)} = \hat{\omega} + \hat{A}_1^{(j)} s_t + \hat{B}_1^{(j)} f_t \quad (4-35)$$

(j) representa o número de simulações, $j = 1, 2, \dots, M$;

Passo 3: Neste ponto, já é possível substituir na Equação 4-32 os valores simulados de $\hat{f}_t^{(i)}$ e $\hat{\lambda}_i^{(j)}$ e gerar M reservas IBNR;

Passo 4: Após a geração de M reservas ($M=10.000$) ordena-se esses valores do menor para o maior e obtêm-se os quantis $\frac{\alpha}{2}$ e $1 - \frac{\alpha}{2}$, para o cálculo de intervalo de $(1 - \alpha)\%$. Nesta tese utilizou-se $\alpha = 0,05$, e portanto, os intervalos serão de 95%.

Assim, em cada linha t do triângulo de *run-off*, tem-se um intervalo de confiança de 95% para a reserva IBNR estimada. Esses intervalos de confiança darão maior segurança e certamente auxiliarão nas tomadas de decisões da seguradora pautadas em menor risco.

4.3.2

Densidade log-normal adaptada ao triângulo de run-off

Nesta subseção assume-se que as entradas dos triângulos de *run-off* são modeladas pela distribuição log-normal, ou seja:

$$p(y_{t,i}|\mu_{t,i}, \sigma, F_{t-1}) = \frac{1}{y_{t,i}\sigma\sqrt{2\pi}} \exp. \left[\frac{(\log y_{t,i} - \mu_{t,i})^2}{2\sigma^2} \right], y_{t,i} \in (0, +\infty) \quad (4-36)$$

Onde $\sigma > 0$, $\mu_{t,i} \in \mathfrak{R}$ e varia de acordo com a linha t e com a coluna i do triângulo, $t = 1, 2, 3, \dots T$ e $i = 1, 2, 3, \dots N$.

O logaritmo da densidade (Equação 4-36) é dada pela equação:

$$\log p(y_{t,i}|\mu_{t,i}, \sigma, F_{t-1}) = -\log(y_{t,i}) - \log(\sigma) - \frac{1}{2}\log(2\pi) - \frac{\log(y_{t,i} - \mu_{t,i})^2}{2\sigma^2} \quad (4-37)$$

A esperança e a variância da log-normal são dadas pelas equações a seguir:

$$\begin{aligned} E[y_{t,i}|\mu_{t,i}, \sigma, F_{t-1}] &= e^{\mu_{t,i} + \frac{\sigma^2}{2}} \\ V[y_{t,i}|\mu_{t,i}, \sigma, F_{t-1}] &= (e^{\sigma^2} - 1)e^{2\mu_{t,i} + \sigma^2} \end{aligned} \quad (4-38)$$

Adicionalmente, tem-se que:

$$E[y_{t,i}|F_{t-1}] = e^{\mu_{t,i} + \frac{\sigma^2}{2}} = e^{f_t + \lambda_i + \frac{\sigma^2}{2}} \quad (4-39)$$

Uma vez que $\mu_{t,i} = f_t + \lambda_i$ e $\mu_{t,i} \in \mathfrak{R}$.

A Equação 4-39 será utilizada no cálculo das reservas IBNR. Para cada célula do triângulo inferior, as reservas IBNR serão dadas pelas esperanças condicionais e a reserva total será dada pelo somatório das esperanças estimadas para as células que apresentarem valores ausentes nos triângulos de *run-off*.

4.3.2.1

Dinâmica de atualização - distribuição log-normal

O vetor *score* é dado pela equação:

$$\nabla_t = \sum_{i=1}^N \delta_{i,t} \nabla_{t,i}(f_t) = \sum_{i=1}^N \delta_{t,i} \frac{\partial \log p(y_{t,i}|\mu_{t,i}, \sigma, F_{t-1})}{\partial f_t} \quad (4-40)$$

$I_{t|t-1}$ foi descrito na Equação 4-30 e $\delta_{t,i}$ na Equação C-4.

O modelo GAS(1,1) com distribuição log-normal é apresentado a seguir.

$$f_{t+1} = \omega + A_1 s_t + B_1 f_t$$

s_t sendo:

$$s_t = I_{t|t-1}^{-d} \nabla_t \quad d = 0, \frac{1}{2} \text{ ou } 1$$

Quando $d = \frac{1}{2}$, segue que:

$$\begin{aligned} s_t &= I_{t|t-1}^{-\frac{1}{2}} \nabla_t \\ &= \frac{1}{\sqrt{\sum_{i=1}^N \delta_{t,i} \frac{1}{\sigma^2}}} \sum_{i=1}^N \delta_{t,i} \left[\frac{\log(y_{t,i}) - \mu_{t,i}}{\sigma^2} \right] \end{aligned} \quad (4-41)$$

4.3.2.2

Cálculo do IBNR assumindo distribuição log-normal

O IBNR para cada posição do triângulo de *run-off* é dada pela equação:

$$\begin{aligned} \text{IBNR}_{t,i} &= E[y_{t,i} | f_t, F_{t-1}; \psi] \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T \\ &= e^{f_t + \lambda_i + \frac{\sigma^2}{2}} \\ &= e^{\mu_{t,i} + \frac{\sigma^2}{2}} \end{aligned} \quad (4-42)$$

As reservas IBNR por linhas e total são calculadas conforme descrito na Subseção 4.3.1.2, utilizando o somatório das esperanças condicionais para todas as posições *missing*, para cada t e posteriormente soma-se todos os somatórios em t .

4.3.2.3

Cálculo do coeficiente de variação para a distribuição log-normal

Assim como na Subseção 4.3.1, os coeficientes de variação das reservas IBNR foram obtidos através simulação, visto que a densidade das posições *missing* não é conhecida. A simulação foi feita utilizando o seguinte algoritmo:

Passo 1: Cada célula vazia do triângulo de *run-off* tem uma variável aleatória log-normal $(\mu_{t,i}, \sigma)$, no qual $\mu_{t,i} = f_t + \lambda_i$ e $\sigma^2/2 = e^{\lambda_N}$;

Passo 2: Usando o $\mu_{t,i}$ e o σ estimados para cada célula vazia, simula-se uma observação da distribuição log-normal para cada célula do triângulo;

Passo 3: Ao final reserva-se as somas dessas observações simuladas para cada célula vazia. Neste ponto tem-se uma realização da soma;

Passo 4: Repete-se os passos 2 e 3 dez mil vezes. Ao final se terá em cada célula dez mil valores - A média desses valores para cada célula irá coincidir, a menos de flutuação, com o valor dado pela média da log-normal para cada célula do triângulo (4-42).

Passo 5: Soma-se sequencialmente esses valores sobre as células vazias, e assim, ao final, têm-se 10.000 valores da variável soma, que serão as realizações do IBNR;

Passo 6: A média amostral das somas obtidas no passo 5 e a variância amostral (a fórmula usual dos livros textos) serão utilizadas no cálculo do coeficiente de variação.

4.3.2.4

Cálculo do intervalo de confiança para a distribuição log-normal

O algoritmo que viabiliza o cálculo dos intervalos de confiança descrito na Subseção 4.2.3, para a distribuição log-normal, é semelhante ao apresentado na Subseção 4.3.1.4. A única diferença está no cálculo do IBNR. No caso da distribuição gama será utilizada a Equação 4-32 e no caso da distribuição log-normal a Equação utilizada será a 4-42.

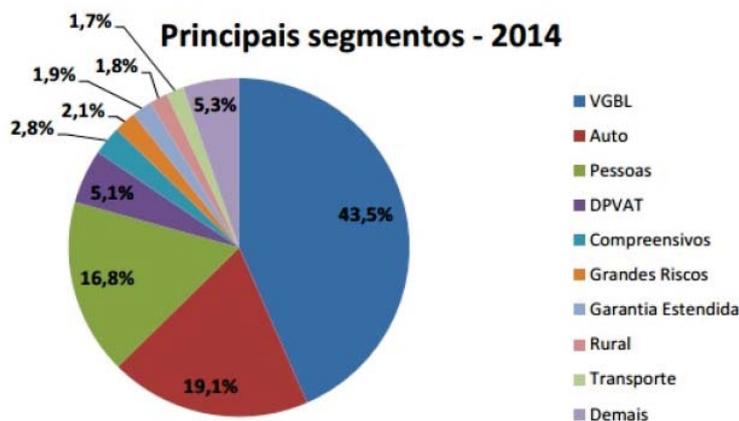
5 Aplicações

Para a aplicação dos modelos Atherino univariado e bivariado (Capítulo 3) e modelos GAS (1,1) com distribuições gama e log-normal (Capítulo 4) utilizou-se dados de triângulos de *run-off* das coberturas Casco e RCFV de uma determinada seguradora brasileira, compreendendo o período entre o primeiro trimestre de 2009 e o segundo trimestre de 2013. Antes de apresentar a base de dados faremos uma breve introdução sobre a organização dos seguros em diversos ramos, pois esse conhecimento é importante para melhor compreender a natureza da base de dados utilizada nessa tese.

Os seguros podem ser agrupados por ramos¹ que, por sua vez, constituem os segmentos de seguros. No Brasil, nove segmentos de seguros (VGBL, Automóveis, Pessoas, DPVAT, Compreensivos, Transportes, Grandes Riscos, Rural e Garantia Estendida), em 2014, representaram 94,4% do volume total de prêmios diretos pagos aos segurados.

A importância do segmento de seguros de automóveis, no Brasil, pode ser observada nos dados técnicos publicados pela SUSEP, os quais indicam que 19,1% do total de prêmios pagos em 2014 foram oriundos dessa categoria. Como pode ser observado na Figura 5.1, no Brasil, o ramo de automóveis é atualmente o segundo maior seguimento de seguros.

Figura 5.1: Participações dos segmentos de seguros no Brasil. Fonte: SUSEP.



¹Conjunto de coberturas diretamente relacionadas ao objeto ou objetivo do plano de seguro (Circular SUSEP 395/09).

A contratação de um seguro de automóvel tem como objetivo garantir ao segurado a indenização ou reembolso, em caso dos sinistros, dos prejuízos sofridos e despesas incorridas, devidamente comprovadas, decorrentes dos riscos cobertos e relativos aos veículos segurados, conforme disposto nas condições gerais de cada apólice contratada. O segmento de automóveis abrange as seguintes coberturas: Casco, RCFV e APP².

Para o ramo de automóveis, a cobertura APP representa, em média, apenas 0,14% do montante pago pela seguradora na forma de sinistros IBNR e por esse motivo a cobertura APP foi preterida nas aplicações desta tese. Abaixo uma breve descrição das coberturas a partir dos quais formam a nossa base de dados.

- **Casco de automóveis:** Cobre o Valor de Mercado Referenciado (VMR), denominado pelo valor da tabela FIPE³ acrescido de um fator de ajuste.
- **RCFV:** Responsabilidade Civil Facultativa de Veículos, também conhecido como seguro para terceiros, pois são destinados aos pagamentos de danos causados a bens, objetos e pessoas alheias ao veículo segurado. O seguro prevê, além disso, o pagamento de advogados e custos judiciais, esse pode ser desmembrado em DM (Danos Materiais) e DC (Danos Corporais). A cobertura RCFV pode ser contratada isoladamente ou em conjunto com as coberturas básicas. Caso essa cobertura seja contratada independentemente das demais, está sujeita à vistoria prévia da seguradora.

Como visto, o seguro de Responsabilidade Civil Facultativa de Veículos é de contratação facultativa e complementa o valor de eventual indenização pago pelo seguro obrigatório de danos pessoais causados por veículos automotores de via terrestres (DPVAT) por danos corporais causados a outras pessoas.

Os triângulos de *run-off* apresentados nas Tabelas 5.1 e 5.2 contêm os dados das coberturas Casco e RCFV com desenvolvimentos trimestrais e compreende o período entre o primeiro trimestre de 2009 e o segundo trimestre de 2013, perfazendo um total de 171 observações.

²O seguro APP - Acidentes Pessoais a Passageiros - tem como objetivo a indenização por danos decorrentes de acidentes pessoais ocorridos aos passageiros.

³A Tabela FIPE expressa os preços médios de veículos no mercado nacional, servindo como um parâmetro para negociações e ou avaliações dos mesmos.

Tabela 5.1: Triângulo contendo os dados trimestrais dos sinistros ocorridos e avisados da cobertura Casco, em R\$ mil.

Trimestre de origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1º trim. 09	677.907	191.328	21.480	6.808	1.934	1.670	983	589	456	363	256	221	76	178	102	104	23	77
2º trim. 09	712.818	182.781	18.911	4.979	3.356	1.427	1.161	687	539	859	319	126	131	298	146	99	33	
3º trim. 09	738.850	177.280	20.738	6.934	3.636	3.090	1.363	687	366	706	200	227	304	178	175	56		
4º trim. 09	728.489	192.991	30.002	8.901	4.271	1.996	1.298	677	458	419	254	255	492	247	57			
1º trim. 10	755.997	208.364	29.660	6.591	3.088	4.069	1.249	1.290	606	756	226	521	70	142				
2º trim. 10	774.917	211.131	24.284	7.202	4.069	1.474	1.948	816	489	331	270	329	86					
3º trim. 10	866.604	143.440	23.776	7.188	3.836	3.073	1.410	963	1.376	530	387	128						
4º trim. 10	914.646	154.792	29.836	7.769	5.148	2.493	1.432	957	1.065	615	294							
1º trim. 11	960.763	146.687	25.064	8.024	3.173	2.779	1.153	618	725	391								
2º trim. 11	949.109	150.757	21.866	6.091	2.732	2.257	1.698	758	385									
3º trim. 11	989.798	149.949	26.117	6.382	5.007	3.648	1.513	1.075										
4º trim. 11	958.699	172.402	28.814	7.821	4.574	1.918	1.125											
1º trim. 12	929.962	182.624	24.379	6.663	3.380	1.459												
2º trim. 12	968.549	156.246	24.337	6.617	4.407													
3º trim. 12	977.076	170.912	23.852	7.872														
4º trim. 12	969.794	161.899	24.743															
1º trim. 13	985.334	135.579																
2º trim. 13	1.020.016																	

Tabela 5.2: Triângulo contendo os dados trimestrais dos sinistros ocorridos e avisados da cobertura RCFV, em R\$ mil.

Trimestre de origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1º trim. 09	98.971	52.556	15.447	6.507	3.176	3.162	3.423	2.562	1.234	1.802	1.921	2.199	993	858	1.460	552	1.199	399
2º trim. 09	114.811	58.024	13.149	6.722	6.383	4.789	4.901	2.768	2.621	2.996	1.814	1.131	1.488	818	1.733	642	1.012	
3º trim. 09	123.011	56.657	14.651	8.164	4.907	5.819	2.722	3.115	2.955	2.166	1.889	1.091	1.992	1.746	1.685	1.289		
4º trim. 09	126.756	67.427	20.338	8.462	5.096	3.731	4.183	5.636	3.375	1.324	1.484	2.630	2.035	1.593	440			
1º trim. 10	119.358	66.624	21.526	9.704	4.298	4.562	5.805	4.334	2.100	1.709	2.480	2.235	1.165	1.523				
2º trim. 10	136.160	72.874	14.668	6.081	5.042	5.709	5.374	2.787	1.992	2.678	2.308	1.178	1.261					
3º trim. 10	161.138	55.712	15.664	7.683	7.123	4.163	2.957	3.636	3.232	1.935	1.368	1.488						
4º trim. 10	167.363	64.322	19.715	10.450	7.033	3.501	3.423	4.677	3.519	2.286	2.859							
1º trim. 11	160.567	57.139	17.671	8.766	4.078	4.589	3.905	3.764	3.567	2.683								
2º trim. 11	188.510	63.566	21.466	9.215	6.077	8.283	5.314	3.227	2.955									
3º trim. 11	195.104	59.368	17.115	8.758	6.104	4.445	4.416	5.027										
4º trim. 11	179.785	65.475	20.397	8.742	6.293	4.417	4.690											
1º trim. 12	182.794	61.830	17.710	6.607	4.642	4.604												
2º trim. 12	200.149	61.649	17.321	6.459	5.412													
3º trim. 12	207.339	62.207	13.470	6.813														
4º trim. 12	210.190	57.643	14.814															
1º trim. 13	198.489	49.569																
2º trim. 13	235.076																	

Cabe destacar que os triângulos superiores contidos nas Tabelas 5.1 e 5.2 comportam os dados de sinistros que ocorreram e foram avisados à seguradora, quer seja no período de ocorrência, quer seja com algum atraso (período de desenvolvimento superior a 1). As partes inferiores das Tabelas 5.1 e 5.2 serão preenchidas com as reservas IBNR estimadas. Essas reservas são destinadas aos pagamentos dos sinistros que ocorreram nos períodos de origem t apresentados nas Tabelas 5.1 e 5.2 e que por algum motivo ainda não foram reportados às seguradoras.

Os dados utilizados nesta tese são correspondentes aos lançamentos contábeis das seguradoras, exatamente como foram reportados. Não sendo aplicado nenhum

fator de correção por inflação ou outro indexador. Caso seja necessário ajustar a base de dados mencionada acima, pode-se fazer uso da correção dos valores escolhendo o último período de ocorrência como data base para o ajuste. No caso dos triângulos 5.1 e 5.2, faria-se a correção dos valores desde a efetivação do seu pagamento até o segundo trimestre de 2013. Sendo assim:

Para a diagonal principal (sinistros pagos em T, independente da data de ocorrência):

$$VP_{C_{t,i}} = C_{t,i} \quad (5-1)$$

Onde $VP_{C_{t,i}}$ é o valor presente de $C_{t,i}$ na data base para o ajuste. Observe que, nesse caso, não é necessário nenhum fator de correção, pois o pagamento ocorre na dada base do ajuste.

Para o restante do triângulo superior:

$$VP_{C_{t,i}} = C_{t,i}(1 + r) \quad (5-2)$$

Onde r é a variação acumulada de um índice de correção (por exemplo o IPCA, FIPE, IGPM, etc.) até a data base para o ajuste.

Segundo as informações disponíveis até o momento (Tabelas 5.1 e 5.2), o montante pago pelas seguradoras para seus segurados, referentes aos sinistros da cobertura Casco, é muito superior ao montante pago aos segurados referente à cobertura RCFV, nos trimestres de desenvolvimentos iniciais. Avaliando o trimestre de origem 1, observa-se que, para a cobertura Casco, 74,9% do total pago em todos os períodos de desenvolvimento ($i=1,2,3,\dots,18$) foi pago no mesmo trimestre da ocorrência do sinistro. De maneira distinta, na cobertura RCFV, menos de 50,0% foi pago no primeiro trimestre de desenvolvimento. Possivelmente, esses padrões individuais se repetem nos demais trimestres de ocorrência dos sinistros.

As Figuras 5.2 e 5.3 apresentam as séries provenientes do empilhamento dos dados (Subseção 2.3.4) contidos nos triângulos de *run-off* incrementais 5.1 e 5.2 e seus logaritmos.

Figura 5.2: Cobertura Casco - série empilhada de sinistros ocorridos e avisados e seus logaritmos.

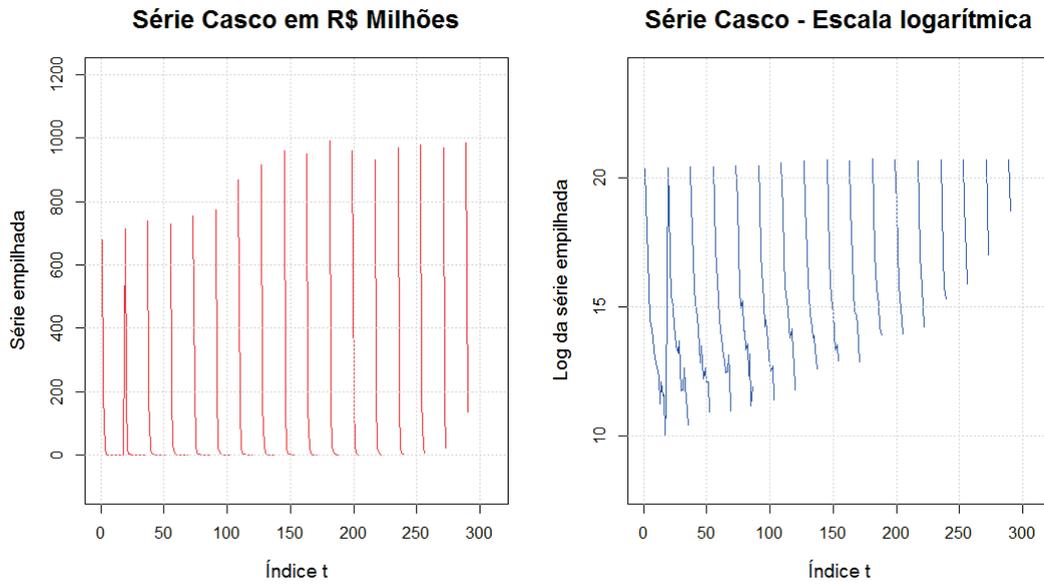
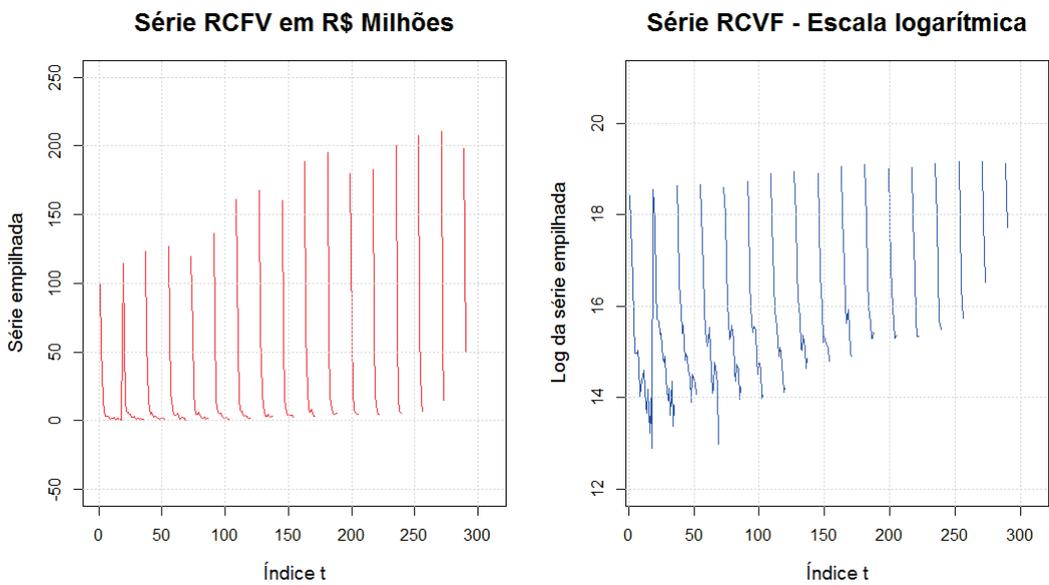


Figura 5.3: Cobertura RCFV - série empilhada de sinistros ocorridos e avisados e seus logaritmos.



A Tabela 5.3 apresenta as estatísticas descritivas das séries provenientes do empilhamento dos dados (Subseção 2.3.4) contidos nos triângulos de *run-off* 5.1 e 5.2 e seus logaritmos.

Tabela 5.3: Estatísticas descritivas dos montantes ocorridos e avisados - coberturas Casco e RCFV e seus logaritmos.

Estatísticas descritivas*	Séries			
	Casco	Casco (logaritmo)	RCFV	RCFV (logaritmo)
Min.	22.846	10,03	399.120	12,89
Máx.	1.020.015.724	20,74	235.076.407	19,27
Média	104.946.733	15,23	27.857.555	15,80
DP	258.817.611	2,80	52.338.529	1,56
Assimetria	2,62	0,54	2,37	0,71
Excesso de curtose	5,43	-0,72	4,65	-0,45
r(1)	0,06	0,55	0,20	0,54
r(2)	-0,04	0,24	0,01	0,19
r(18)	0,06	0,27	0,09	0,24
Teste de R. U. (ADF)	-7,25 (0,01)	-6,06 (0,01)	-5,87 (0,01)	-6,96 (0,01)

Nota: “**” - Deve-se exercer cautela na leitura desses resultados, devido aos valores faltantes de cada uma das séries.

Com base na Tabela 5.3, analisando estritamente as séries Casco e RCFV, podemos perceber que a série Casco apresenta dispersão maior, comparado à série RCFV. As duas séries apresentam assimetria positiva e possuem excesso de curtose.

Ainda com base na Tabela 5.3, desta vez analisando somente os logaritmos das séries Casco e RCFV, observamos que a assimetria é de 0,54 e 0,71, respectivamente e os excessos de curtoses das séries são negativos e próximos de zero.

Como se pode observar na Tabela 5.3, tanto para as séries quanto para o logaritmo destas, de acordo com o teste ADF com constante, rejeita-se a hipótese de raiz unitária. As autocorrelações das séries são inferiores as autocorrelações dos logaritmos destas em todas as ordens apresentadas.

Atherino et al. (2010) evidenciaram em seu artigo que o modelo ajustado ao logaritmo dos dados apresenta melhores resultados e melhores medidas de aderência do que o modelo ajustado à série original. Deste modo, para o desenvolvimento a seguir utilizou-se o logaritmo das séries provenientes dos triângulos de *run-off* (coberturas Casco e RCFV).

Nas próximas seções serão apresentados os resultados dos modelos Atherino bivariado (formulação SUTSE) e GAS (1,1) com distribuições gama e log-normal. O resultado da aplicação do método *Chain Ladder* e do modelo Atherino univariado tem como objetivo a comparação com os resultados dos modelos propostos nesta tese. Isto posto, os resultados da estimação do modelo Atherino univariado (parâmetros estimados, componentes suavizadas e análise residual) encontram-se no Anexo D.

5.1

Resultados do modelo Atherino bivariado (formulação SUTSE)

Nesta tese utilizamos o conceito de agrupamento dos dados do triângulo de *run-off* conforme proposto por Atherino et. al (2010) e definido na Subseção 2.3.4. Essa forma de apresentar os dados (empilhamento) permite que se observe padrões nas coberturas que podem ser representados por componentes como nível e

periodicidade. A apresentação dos dados como uma série temporal e o uso de modelos estruturais permitem a captura de dependência entre cada célula do triângulo de *run-off*, conforme apresentado na Subseção 3.2.2.

A dependência entre as coberturas selecionadas (Casco e RCFV) é capturada quando se modela, de forma conjunta, as duas séries temporais provenientes dos triângulos de *run-off*. Essas associações entre a ocorrência dos sinistros IBNR de coberturas distintas é modelada através da formulação bivariada (formulação SUTSE) apresentada no Capítulo 3.

Como podemos observar (Tabelas 5.1 e 5.2) os triângulos de *run-off* que utilizaremos nesta aplicação possuem 18 colunas, logo $N = 18$ (Equação 3-43).

Cabe ressaltar que a aplicação da formulação SUTSE acrescido do modelo Atherino (aqui denominado de Atherino bivariado) se faz justificável quando as coberturas envolvidas apresentam algum tipo de relação. Assim, o desenvolvimento de uma cobertura traz informações acerca da outra cobertura relacionada e vice-versa. Uma medida simples que pode ser calculada para averiguar se as séries são relacionadas é a correlação linear. Para o caso das séries escolhidas nesta tese (coberturas Casco e RCFV) a correlação linear foi de 98,0%, indicando que as mesmas possuem alguma dependência.

Inicialmente ajustou-se o modelo proposto nesta tese - modelo Atherino bivariado (formulação SUTSE) com componentes de nível e de periodicidade estocásticas (Equações 3-31 e 3-30) as séries das coberturas Casco e RCFV obtidas a partir dos respectivos triângulos de *run-off* - Figuras 5.2 e 5.3. Posteriormente, a partir das matrizes de covariâncias estimadas $\hat{\Sigma}_\eta$ e $\hat{\Sigma}_\omega$, calculou-se o posto destas matrizes. Por conseguinte, averiguou-se a existência de uma componente comum entre as séries modeladas. Logo, decidiu-se por aplicar uma formulação distinta, em que o nível é comum as duas séries (Subseção 3.3.2). Comparou-se também as medidas de aderência para os dois modelos, com e sem nível comum, a primeira formulação aplicada mostrou-se mais aderente aos dados.

A estimação via máxima verossimilhança (BFGS) das variâncias dos choques aleatórios (Equação 3-49) utilizou a inicialização difusa exata (Anexo A.4) e o algoritmo EM (Subseção A.6). Essas estimativas encontram-se na Tabela 5.4:

Tabela 5.4: Parâmetros estimados - modelo Atherino bivariado (formulação SUTSE).

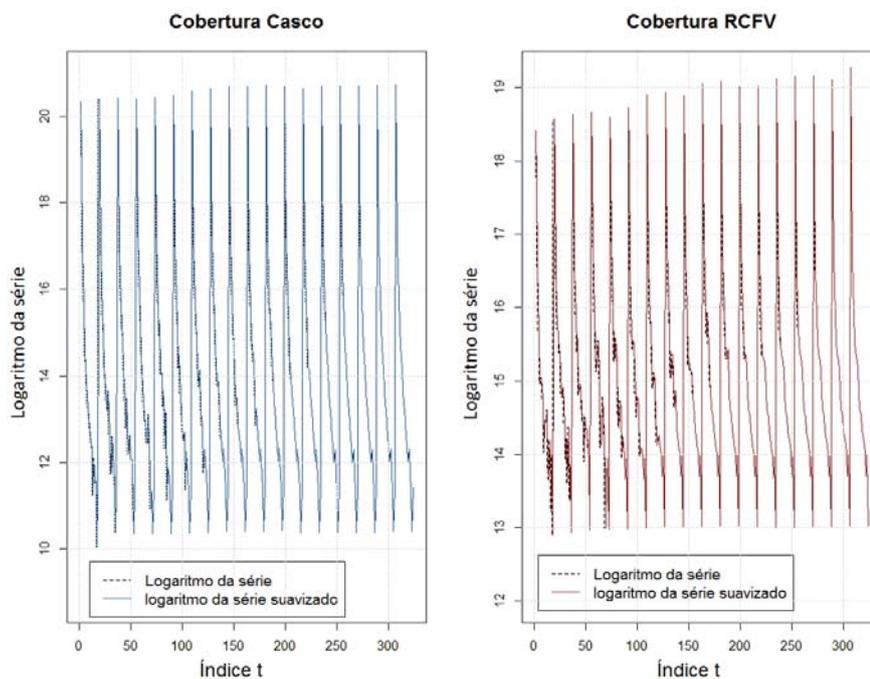
Parâmetro	estimativa
$\sigma_{\varepsilon_1}^2$	$8,48 \times 10^{-02}$
$\sigma_{\varepsilon_2}^2$	$6,60 \times 10^{-02}$
$\sigma_{\varepsilon_1\varepsilon_2}$	$2,25 \times 10^{-02}$
$\sigma_{\eta_t^+}^2$	$1,94 \times 10^{-04}$
$\sigma_{\omega_1}^2$	$1,33 \times 10^{-19}$
$\sigma_{\omega_2}^2$	$2,04 \times 10^{-19}$
$\sigma_{\omega_1\omega_2}$	$-1,04 \times 10^{-10}$

Como pode ser observado na Tabela 5.4, as estimativas dos parâmetros de variância e covariância revelaram, para as duas coberturas, periodicidade determinística e ausência de associação entre as componentes periódicas das duas séries (Casco e RCFV).

A variável π_η denotada na Equação 3-42 (ver também Equação 3-27), para a modelagem SUTSE aplicada às séries Casco e RCFV, é igual a 0,715. Como visto, $\mu_{2,t} = \pi_\eta \mu_{1,t} + \bar{\mu}$ (sendo $\bar{\mu} = 5,14$), ou seja, o nível da cobertura RCFV pode ser escrito como uma combinação linear do nível da cobertura Casco.

Na Figura 5.4, observa-se as séries obtidas através da abordagem de suavização para ambas as coberturas selecionadas.

Figura 5.4: Coberturas Casco e RCFV - logaritmo das séries IBNR e as séries suavizadas.



Analisando a Figura 5.4, pode-se constatar que as estimativas suavizadas aderem-se bem aos dados. Nas Figuras 5.5 e 5.6 apresentam-se os níveis e as periodicidades modeladas de forma bivariada.

Figura 5.5: Cobertura Casco - logaritmo da série IBNR, o nível e a periodicidade suavizadas (modelo Atherino bivariado - formulação SUTSE).

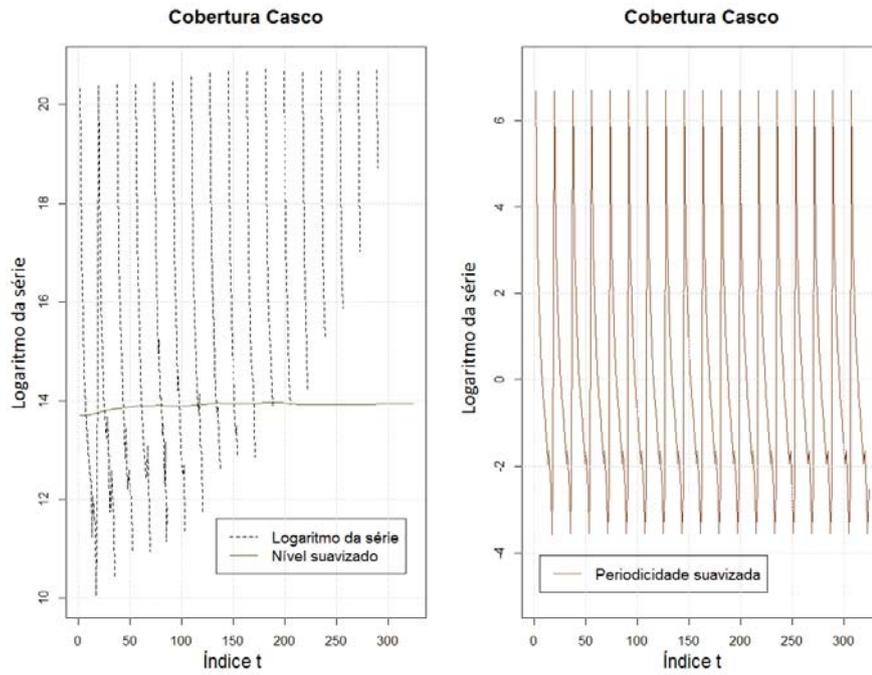
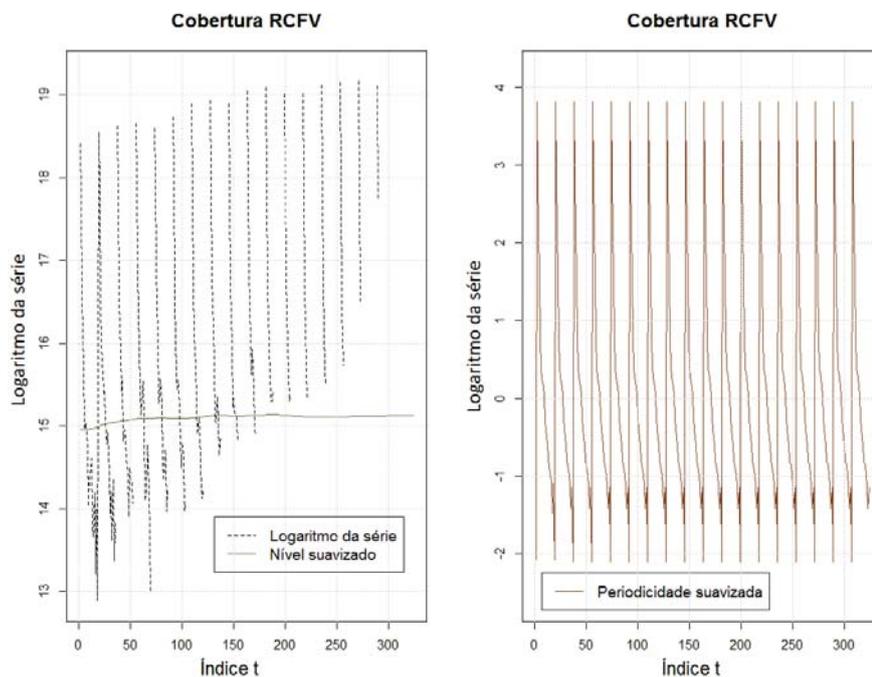


Figura 5.6: Cobertura RCFV - logaritmo da série IBNR, o nível e a periodicidade suavizadas (modelo Atherino bivariado - formulação SUTSE).



As Figuras 5.7, 5.8, 5.9 e 5.10 apresentam os gráficos contendo as inovações padronizadas, os resíduos suavizados, as funções de autocorrelação e os Q-Q plots.

Figura 5.7: Casco - inovações padronizadas e resíduos suavizados (modelo Atherino bivariado - formulação SUTSE).

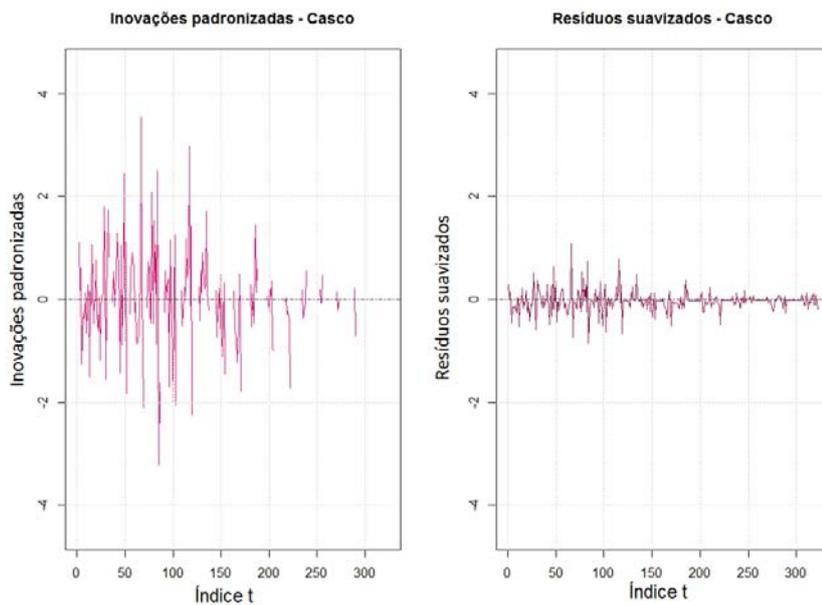


Figura 5.8: RCFV - inovações padronizadas e resíduos suavizados (modelo Atherino bivariado - formulação SUTSE).

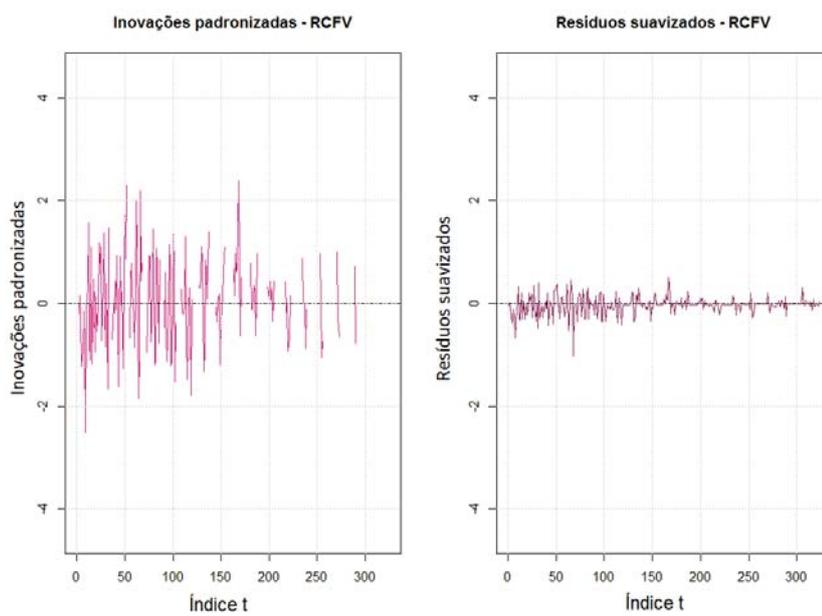
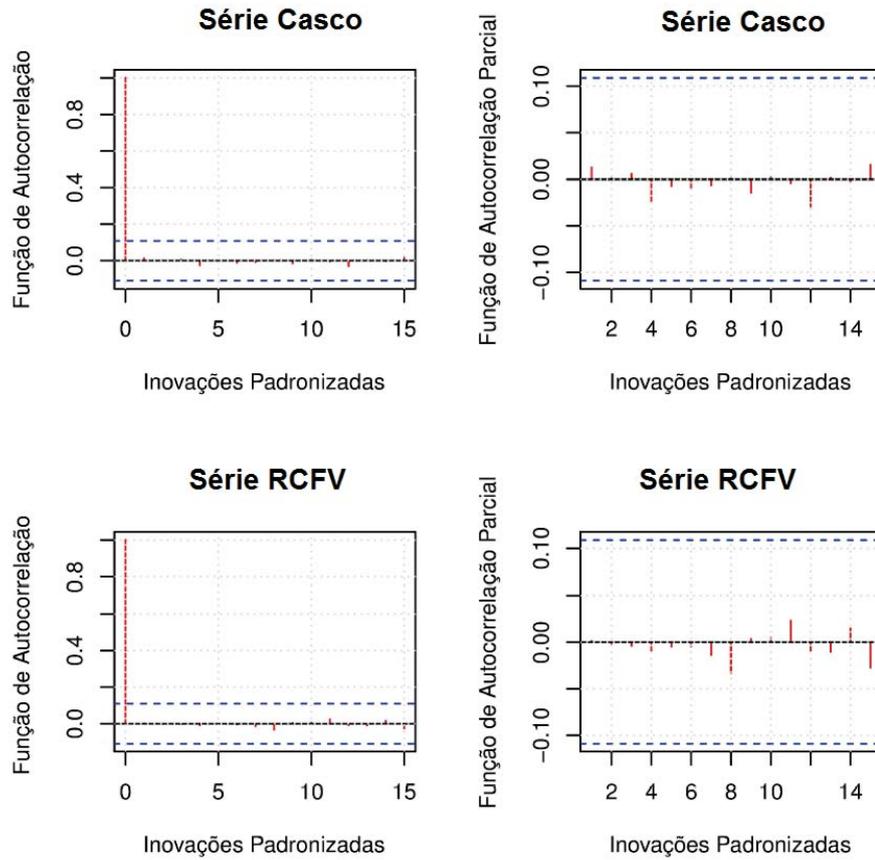
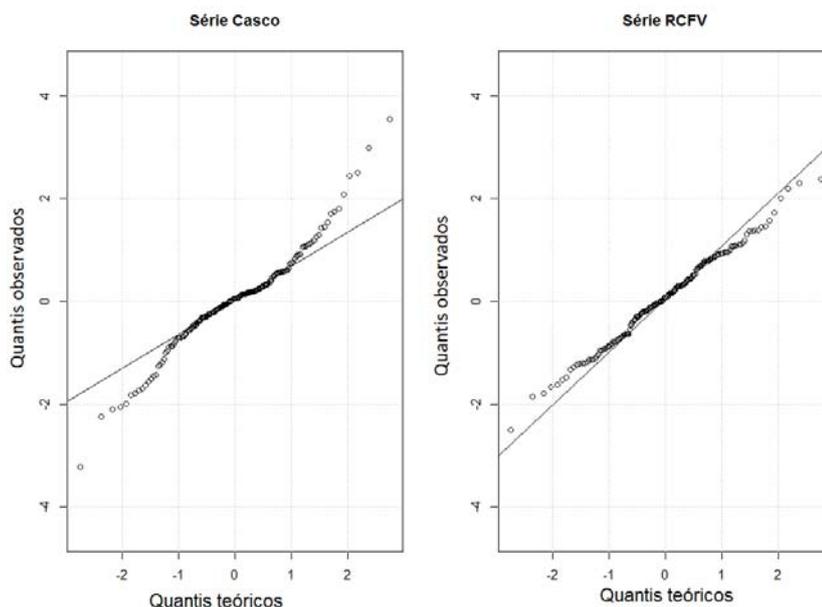


Figura 5.9: Coberturas Casco e RCFV - função de autocorrelação e função de autocorrelação parcial.



Verificou-se, tanto na FAC quanto na FACP das inovações padronizadas, que não há evidências fortes de dependências lineares.

Figura 5.10: Coberturas Casco e RCFV - Q-Q plotes das inovações padronizadas, modelo Atherino bivariado.



Como pode ser observado, os pontos se ajustam bem as retas apenas nas partes centrais dos gráficos. Há nos extremos alguns dados que não se ajustam as retas. Esses dados podem comprometer a normalidade das inovações padronizadas.

Na Tabela 5.5 são apresentados: o teste de autocorrelação, o teste de normalidade e o teste de homogeneidade da variância.

Tabela 5.5: Testes e diagnósticos aplicados às inovações padronizadas para o modelo Atherino bivariado.

Série/teste	estatísticas de teste e <i>p</i> -valores		
	Homocedasticidade (Qui-Quadrado)	Autocorrelação (Ljung-Box)	Normalidade (Jarque-Bera)
Casco	0,414 (0,996)	24,526 (0,748)	11,532 (0,003)
RCFV	0,342 (0,999)	38,040 (0,149)	22,036 (0,000)

No teste de heterocedasticidade, obteve-se *p*-valor igual a 0,996 para a série Casco e 0,999 para a série RCFV, não sugerindo a rejeição de H_0 , ou seja, não há evidências de heterocedastidade. Nos testes Box-Ljung, não há evidências para rejeitar a hipótese nula, isto é, as séries são descorrelatadas. No teste de normalidade Jarque-Bera, rejeita-se a hipótese nula, a hipótese de normalidade dos resíduos para ambas as séries.

A não normalidade dos resíduos provavelmente ocorre devido às posições mal ajustadas nos extremos (ver Figura 5.10), essas podem ser acomodadas no modelo

via variáveis *dummies*, fazendo com que não se rejeite a hipótese de normalidade para as inovações padronizadas.

As medidas de aderência, as reservas IBNR estimadas e os coeficientes de variação provenientes do modelo Atherino bivariado (formulação SUTSE) serão descritos no final deste capítulo, juntamente com todos os resultados do modelo GAS (1,1) com distribuições gama e log-normal.

5.2

Resultados do modelo GAS (1,1)

Como já descrito no Capítulo 4, aplicou-se o modelo GAS (1,1) com distribuição gama e log-normal nos dados contidos nos triângulos de *run-off* apresentados nas Tabelas 5.1 e 5.2. Para essa aplicação, diferentemente da aplicação do modelo Atherino bivariado, os dados observados nos triângulos foram modelados como dados em painel, onde t representa o período de origem e i representa o período de desenvolvimento, $t = 1, \dots, T$ e $i = 1, \dots, N$. Nesta aplicação $t = 1$ representa o primeiro trimestre de 2009 e $t = T$ representa o segundo trimestre de 2013.

Para as inicializações dos vetores de parâmetros estáticos e desconhecidos do modelo GAS (1,1) com distribuição gama: $\psi = (\omega, A_1, B_1, \lambda_1, \dots, \lambda_N)$ ou do modelo GAS (1,1) com distribuição log-normal: $\psi = (\omega, A_1, B_1, \lambda_1, \dots, \lambda_N, \sigma)$, utilizou-se uma versão do algoritmo genético (Subseção 4.1.3), de forma a minimizar o tempo computacional e evitar que o processo de maximização retorne máximos locais distantes do ótimo global.

Com relação a inicialização do vetor de parâmetros variantes no tempo, f_1 , foram feitas algumas tentativas. A primeira foi colocá-lo no vetor de parâmetros desconhecidos; a segunda, associá-lo a média da primeira coluna do triângulo de *run-off* e a terceira tentativa, assumir estacionariedade do processo e calcular f_1 conforme Equação 4-10. Analisando os gradientes, as matrizes Hessianas e o tempo de convergência, observou-se que a melhor solução foi a última testada.

Com o objetivo de evidenciar a existência de dependência temporal nos dados contidos nos triângulos de *run-off* (Tabelas 5.1 e 5.2), aplicou-se, inicialmente, o modelo com distribuição gama estático, no qual, $A = 0$, $B = 0$, $\alpha = e^\omega$ e $f_t = \omega$, ou seja f_t constante (observar Equação 4-41). As tabelas apresentadas nos Apêndices E e F apontam uma menor aderência desses modelos estáticos aos dados, além de reservas IBNR e coeficientes de variação mais elevados. Portanto, conclui-se que há ganhos na aplicação do modelo GAS com coeficientes variantes no tempo, no qual se leva em consideração o fator temporal contido nas coberturas Casco e RCFV.

As distribuições escolhidas (gama e log-normal) são contínuas e possuem suportes nos reais positivos, e assim tornam-se adequadas à modelagem das coberturas Casco e RCFV.

Nas próximas subseções serão apresentados os resultados da estimação do modelo GAS para cada uma das coberturas.

5.2.1

Cobertura Casco

5.2.1.1

Modelo GAS (1,1) com distribuição gama

Utilizando o triângulo de *run-off* com dados IBNR da cobertura Casco (Tabela 5.1) e o modelo GAS(1,1) com distribuição gama (Subseção 4.3.1), estimou-se os seguintes parâmetros variantes no tempo (Tabela 5.6):

Tabela 5.6: GAS (1,1) com distribuição gama - estimativas dos parâmetros variantes no tempo (cobertura Casco).

Parâmetros variantes no tempo			
t	f_t	∇_t	S_t
1	2,522	-43,943	0,065
2	2,464	1,470	0,069
3	2,561	-10,304	0,068
4	2,483	11,322	0,073
5	2,563	2,673	0,073
6	2,500	-5,857	0,078
7	2,527	13,326	0,080
8	2,540	21,921	0,083
9	2,547	-5,724	0,087
10	2,496	-5,373	0,094
11	2,528	16,534	0,098
12	2,550	3,071	0,104
13	2,510	-2,616	0,114
14	2,523	3,804	0,124
15	2,530	4,075	0,138
16	2,528	2,016	0,160
17	2,525	-0,360	0,196
18	2,519	2,412	0,278

Como pode ser observado na Tabela 5.6, o vetor *score* ∇_t apresenta variações mais significativas comparadas as variações dos parâmetros f_t e S_t ao longo das linhas do triângulo de *run-off*.

Na Tabela 5.7 estão descritos, para o triângulo de *run-off* contendo os dados dos sinistros da cobertura Casco (Tabela 5.1), os parâmetros denominados estáticos e os parâmetros que mudam de acordo com a coluna.

Tabela 5.7: Cobertura Casco - estimativas dos parâmetros estáticos (vetor ψ). Modelo GAS (1,1) com distribuição gama.

ω	A	B									
4,127 (0,913)	0,020 (0,001)	-0,637 (0,149)									
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9			
4,255 (0,012)	2,608 (0,013)	0,686 (0,013)	-0,574 (0,012)	-1,203 (0,014)	-1,646 (0,015)	-2,219 (0,014)	-2,714 (0,016)	-2,961 (0,013)			
λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}			
-3,120 (0,017)	-3,814 (0,018)	-3,877 (0,019)	-4,162 (0,021)	-4,089 (0,023)	-4,635 (0,027)	-4,969 (0,033)	-6,077 (0,048)	-5,099 (0,077)			

Como pode ser visto na Tabela 5.7, o parâmetro variante no tempo f_t (Equação 4-3) depende negativamente do seu passado (B=-0,637) e positivamente do vetor *score* ponderado (A=0,020). O erro padrão das estimativas ($EP(\lambda_i)$) cresce a medida que a coluna (i) avança. Este fato ocorre, pois, quando i cresce (Tabela 5.1) temos menos observações para estimar cada parâmetro.

Para a análise residual do modelo GAS (1,1) com distribuição gama foram calculados os resíduos quantílicos (Equação 4-12) e resíduos de Pearson (Equação 4-13). Esses encontram-se descritos nas Tabelas 5.8 e 5.9.

Tabela 5.8: Cobertura Casco - resíduos quantílicos, modelo GAS (1,1) com distribuição gama.

Trim. origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 ^o T09	-0,78	0,54	-0,39	-0,01	-1,91	-1,11	-0,97	-1,03	-1,06	-1,28	-0,15	-0,44	-2,46	-0,45	-0,50	0,77	-0,67	0,16
2 ^o T09	-0,40	0,57	-0,60	-0,83	-0,07	-1,37	-0,22	-0,32	-0,31	1,97	0,83	-1,93	-0,98	1,62	0,96	0,77	0,78	
3 ^o T09	-0,64	0,13	-0,65	-0,09	-0,15	0,89	-0,02	-0,68	-1,88	0,88	-1,11	-0,48	1,67	-0,59	1,34	-1,50		
4 ^o T10	-0,40	0,71	0,93	1,10	0,71	-0,39	0,08	-0,44	-0,90	-0,68	-0,04	0,19	3,08	0,83	-2,10			
1 ^o T10	-0,57	0,71	0,61	-0,27	-0,71	1,99	-0,34	1,66	-0,27	1,14	-0,73	2,77	-2,17	-1,33				
2 ^o T10	-0,25	0,98	0,11	0,27	0,47	-1,41	1,53	0,13	-0,75	-1,47	0,12	1,06	-2,03					
3 ^o T10	0,03	-0,49	-0,06	0,16	0,17	0,98	0,22	0,64	2,83	-0,06	1,35	-2,13						
4 ^o T10	0,18	-0,28	0,71	0,40	1,22	0,16	0,23	0,57	1,97	0,43	0,28							
1 ^o T11	0,33	-0,48	0,05	0,49	-0,56	0,54	-0,55	-0,97	0,43	-1,14								
2 ^o T11	0,46	-0,21	-0,24	-0,30	-0,86	-0,03	1,01	-0,11	-1,48									
3 ^o T11	0,50	-0,34	0,27	-0,26	1,15	1,66	0,47	1,05										
4 ^o T11	0,31	0,06	0,55	0,38	0,73	-0,77	-0,64											
1 ^o T12	0,34	0,41	0,09	-0,04	-0,21	-1,48												
2 ^o T12	0,44	-0,18	0,03	-0,11	0,68													
3 ^o T12	0,45	0,10	-0,06	0,48														
4 ^o T12	0,43	-0,08	0,07															
1 ^o T13	0,50	-0,66																
2 ^o T13	0,65																	

Tabela 5.9: Cobertura Casco - resíduos de Pearson, modelo GAS (1,1) com distribuição gama.

Trim. origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1ª T09	-0,80	0,47	-0,46	-0,10	-1,70	-1,07	-0,97	-1,01	-1,03	-1,20	-0,24	-0,51	-2,14	-0,52	-0,56	0,72	-0,71	0,07
2ª T09	-0,48	0,50	-0,65	-0,85	-0,17	-1,27	-0,31	-0,41	-0,39	2,24	0,79	-1,65	-0,97	1,75	0,94	0,72	0,73	
3ª T09	-0,68	0,03	-0,70	-0,18	-0,23	0,86	-0,11	-0,72	-1,63	0,85	-1,07	-0,55	1,82	-0,64	1,40	-1,37		
4ª T10	-0,47	0,65	0,90	1,10	0,65	-0,47	-0,01	-0,51	-0,91	-0,72	-0,13	0,09	2,66	0,79	-1,76			
1ª T10	-0,62	0,66	0,54	-0,36	-0,75	2,25	-0,41	1,80	-0,35	1,15	-0,76	2,98	-2,36	-1,25				
2ª T10	-0,34	0,97	0,01	0,17	0,39	-1,30	1,65	0,04	-0,78	-1,34	0,02	1,06	-1,90					
3ª T10	-0,06	-0,55	-0,15	0,07	0,07	0,97	0,13	0,57	2,97	-0,15	1,42	-1,79						
4ª T10	0,08	-0,36	0,66	0,31	1,25	0,07	0,14	0,50	2,21	0,35	0,19							
1ª T11	0,24	-0,55	-0,04	0,41	-0,62	0,46	-0,60	-0,96	0,35	-1,10								
2ª T11	0,38	-0,29	-0,32	-0,38	-0,87	-0,12	1,00	-0,20	-1,35									
3ª T11	0,43	-0,42	0,17	-0,34	1,17	1,81	0,39	1,04										
4ª T11	0,22	-0,03	0,47	0,30	0,68	-0,80	-0,69											
1ª T12	0,25	0,33	-0,01	-0,14	-0,30	-1,35												
2ª T12	0,36	-0,27	-0,06	-0,21	0,62													
3ª T12	0,37	0,01	-0,16	0,40														
4ª T12	0,35	-0,17	-0,02															
1ª T13	0,42	-0,71																
2ª T13	0,58																	

Como pode ser observado nas Tabelas 5.8 e 5.9, menos do que 5,0% das posições apresentam valores absolutos superiores a 2,0, indicando um número pequeno de posições mal ajustadas pelo modelo. A correlação entre as duas séries de resíduos - de Pearson e quantílico - foi de 98,2%, indicando uma forte relação entre eles.

A ausência de valores em muitas posições do triângulo de *run-off* dificulta a análise residual. Para realizar os testes de autocorrelação serial, heterocedasticidade condicional e normalidade construiu-se uma série empilhando por linhas os resíduos dos triângulos de *run-off* e não se considerou as posições na qual existem NAs (posições *missing*). Observa-se que pela construção do modelo (Equações 4-41 e 4-41) a estimação de uma determinada posição do triângulo de *run-off* depende de todas as informações dos períodos de ocorrência e dos períodos de desenvolvimentos anteriores a esse.

As Figuras 5.11, 5.12, 5.13 e 5.14 apresentam os resíduos quantílicos e de Pearson empilhados, as funções de autocorrelação associadas e os Q-Q plots.

Figura 5.11: Casco - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição gama.

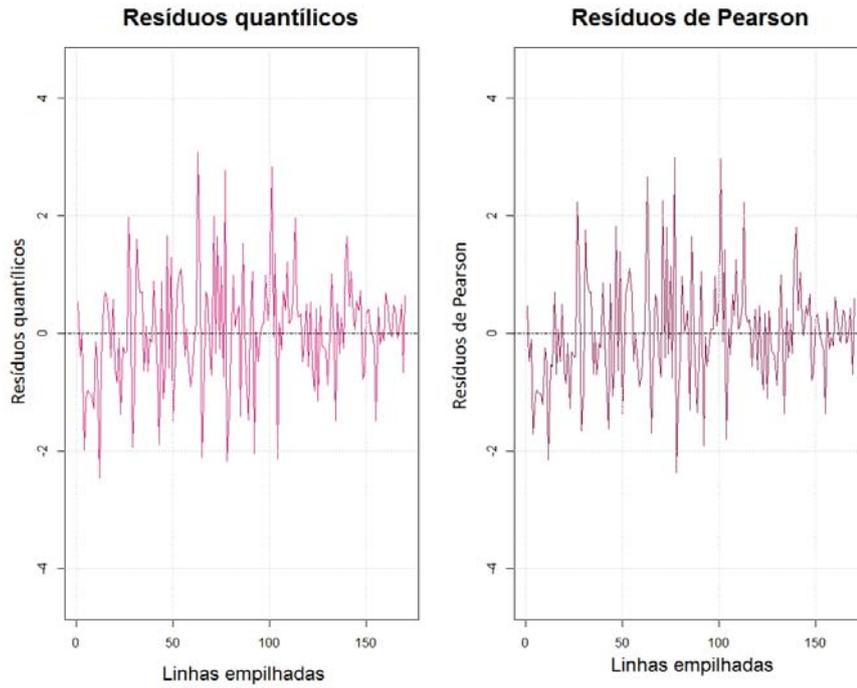


Figura 5.12: Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição gama.

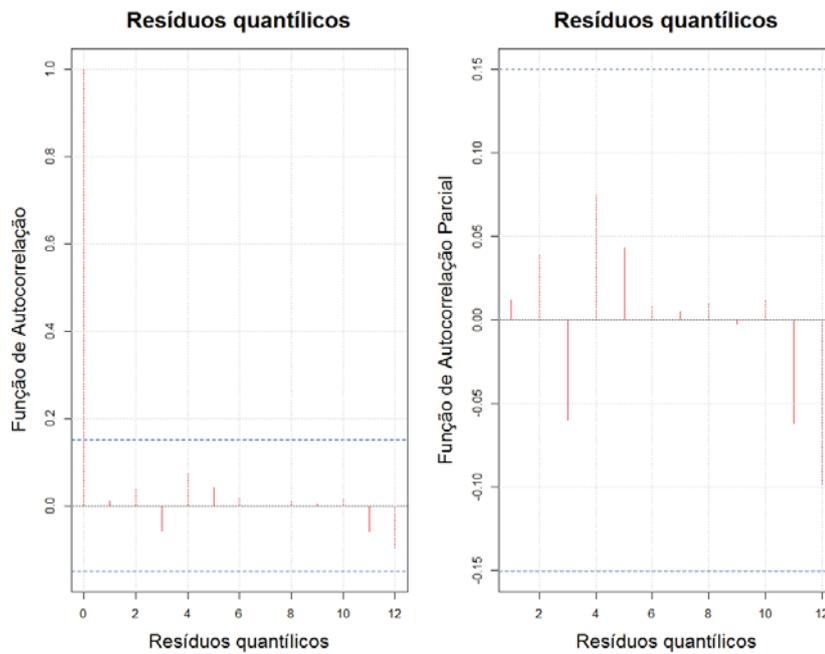
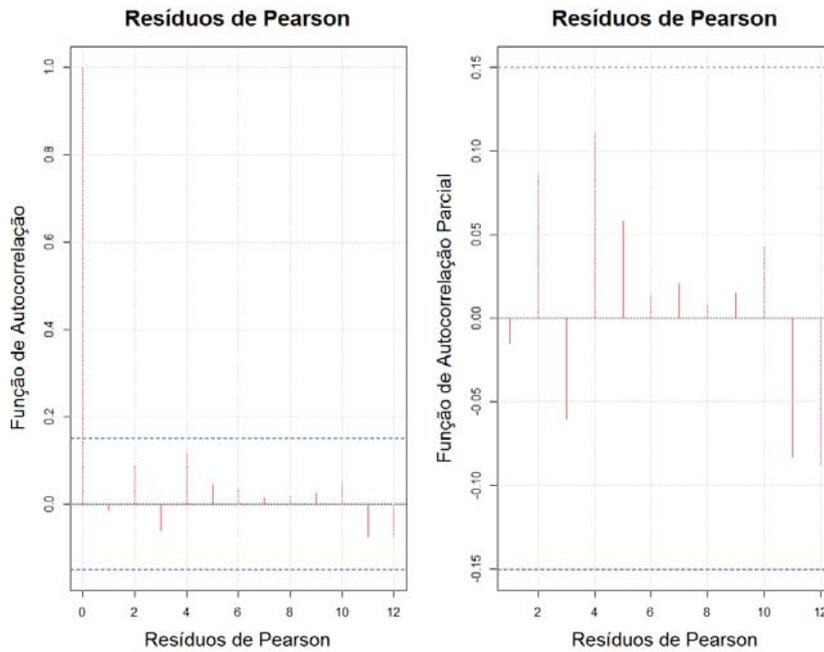
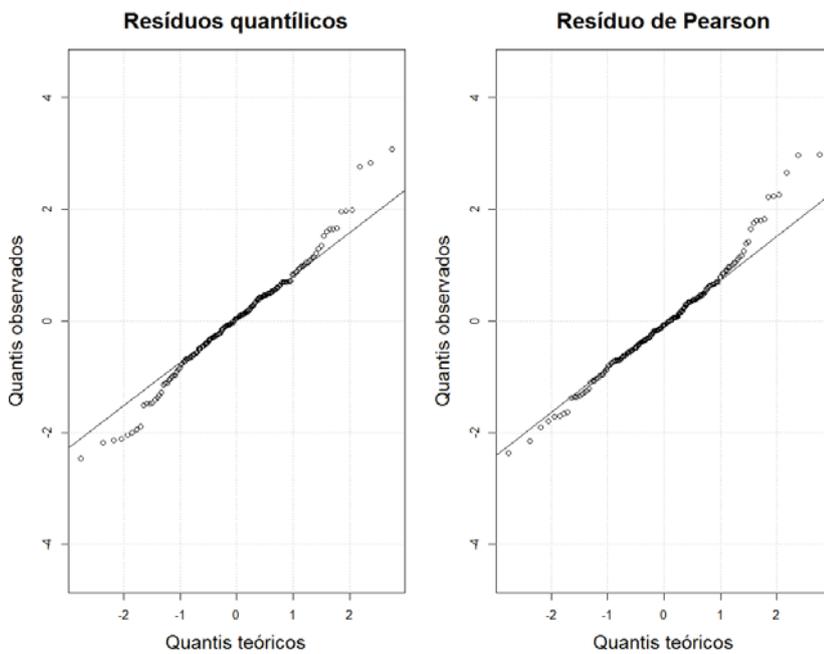


Figura 5.13: Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.



Ao analisar as funções de autocorrelação apresentadas acima, constatou-se que não há evidências de dependência linear.

Figura 5.14: Casco - Q-Q plots dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.



Como pode ser observado nos Q-Q plots, os pontos se ajustam bem as retas

apenas nas partes centrais dos gráficos. Há nos extremos alguns dados que não se ajustam as retas.

Para auxiliar na análise residual do modelo GAS (1,1) com distribuição gama foram feitos os testes de Heterocedasticidade, Box-Ljung e Jarque-Bera nos resíduos quantílicos e nos de Pearson (Tabela 5.10).

Tabela 5.10: Cobertura Casco - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição gama.

Série/teste	estatísticas de teste e <i>p-valores</i>		
	Homocedasticidade (Qui-Quadrado)	Autocorrelação (Ljung-Box)	Normalidade (Jarque-Bera)
resíduo de Pearson	0,392 (0,999)	0,0369 (0,8477)	13.394 (0,001)
resíduo quantílico	0,390 (0,998)	0,0251 (0,8741)	5,353 (0,068)

Nos resíduos de Pearson, o teste de heterocedasticidade apresentou *p-valor* igual a 0,999, não sugerindo a rejeição de H_0 , ou seja, não há evidências de heterocedasticidade. No teste Box-Ljung, não há evidências para rejeitar a hipótese nula, isto é, a série de resíduos é decorrelatada. No teste de normalidade Jarque-Bera, rejeita-se a hipótese de normalidade.

Para os resíduos quantílicos aceita-se, ao nível de 5,0%, a hipótese de normalidade, a hipótese de ausência de correlação e a hipótese de homocedasticidade.

5.2.1.2

GAS (1,1) com distribuição gama - reserva IBNR e intervalo de confiança, cobertura Casco

Como visto na Subseção 4.2.3, os intervalos de confiança para os parâmetros estáticos do modelo viabilizam o cálculo dos intervalos para as reservas IBNR, sendo dado pelas esperanças condicionais dos y_{it} , $E[y_{it}|f_t, F_{t-1}, \hat{\psi}]$, $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$.

Na Tabela 5.11, estão apresentados os intervalos de confiança e as reservas IBNR por linhas e totais simuladas a partir dos desenvolvimentos apresentados na Subseção 4.2.3.

Tabela 5.11: Cobertura Casco - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição gama.

GAS (1,1) com distribuição gama			
Intervalo de confiança de 95,0%			
Trimestre	2,5%	Reservas estimadas	97,5%
2 ^o trim.09	R\$ 41.490	R\$ 74.802	R\$ 124.614
3 ^o trim.09	R\$ 72.930	R\$ 112.638	R\$ 171.845
4 ^o trim.09	R\$ 142.784	R\$ 190.091	R\$ 252.095
1 ^o trim.10	R\$ 266.270	R\$ 332.545	R\$ 415.238
2 ^o trim.10	R\$ 422.528	R\$ 518.600	R\$ 626.123
3 ^o trim.10	R\$ 626.012	R\$ 736.816	R\$ 883.165
4 ^o trim.10	R\$ 792.949	R\$ 993.467	R\$ 1.171.797
1 ^o trim.11	R\$ 1.140.333	R\$ 1.324.481	R\$ 1.635.221
2 ^o trim.11	R\$ 1.259.568	R\$ 1.737.174	R\$ 2.031.693
3 ^o trim.11	R\$ 2.185.965	R\$ 2.535.739	R\$ 3.308.794
4 ^o trim.11	R\$ 2.152.931	R\$ 3.311.684	R\$ 3.889.790
1 ^o trim.12	R\$ 4.103.148	R\$ 4.731.851	R\$ 6.204.447
2 ^o trim.12	R\$ 4.433.968	R\$ 6.860.886	R\$ 7.877.368
3 ^o trim.12	R\$ 9.642.586	R\$ 11.219.347	R\$ 14.935.868
4 ^o trim.12	R\$ 10.697.607	R\$ 17.503.952	R\$ 20.306.064
1 ^o trim.13	R\$ 37.034.803	R\$ 43.942.873	R\$ 60.058.060
2 ^o trim.13	R\$ 87.677.434	R\$ 203.883.038	R\$ 245.953.434
Total	R\$ 241.047.942	R\$ 300.009.993	R\$ 350.278.281

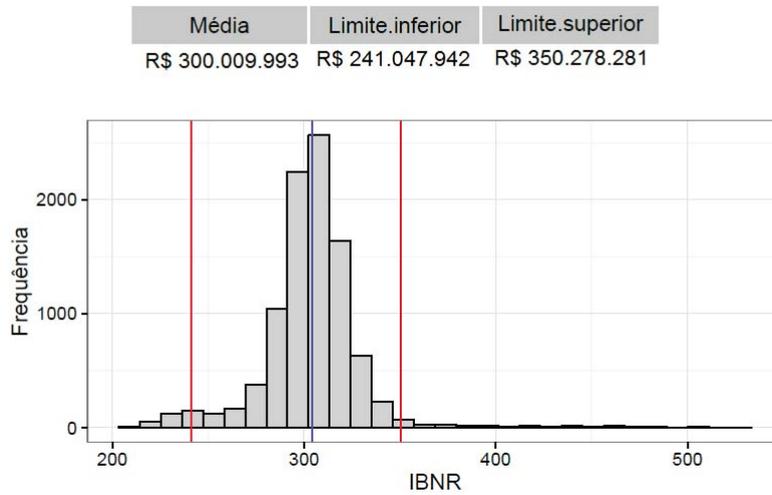
Como pode ser visto na Tabela 5.11, o intervalo de confiança de 95,0% para a reserva IBNR total é dado por:

$$R\$241.047.942,00 \leq IBNR_{Total} \leq R\$350.278.281,00 \quad (5-3)$$

Ou seja, de acordo com o modelo GAS (1,1) com distribuição gama, as reservas IBNR totais têm 95,0% de chance de estarem entre R\$241.047.942,00 e R\$350.278.281,00.

A Figura 5.15 apresenta o intervalo de confiança para as reservas IBNR totais simuladas via modelo GAS (1,1) com distribuição gama.

Figura 5.15: Cobertura Casco - reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição gama (histograma).



No Anexo G serão apresentados, para cada trimestre de origem, os gráficos contendo as 10.000 simulações que geraram os intervalos de confiança descritos na Tabela 5.11.

5.2.1.3 Modelo GAS (1,1) com distribuição log-normal

Utilizando o triângulo de *run-off* com dados IBNR da cobertura Casco e o modelo GAS(1,1) com distribuição log-normal, foram estimados os seguintes parâmetros variantes no tempo (Tabela 5.12):

Tabela 5.12: GAS (1,1) com distribuição log-normal - estimativas dos parâmetros variantes no tempo (cobertura Casco).

Parâmetros variantes no tempo			
t	f_t	∇_t	S_t
1	0,806	-3,089	0,245
2	0,792	-0,397	0,252
3	0,813	-0,021	0,260
4	0,800	0,616	0,269
5	0,812	0,851	0,278
6	0,806	-0,532	0,289
7	0,803	1,032	0,300
8	0,813	1,602	0,314
9	0,810	-0,303	0,329
10	0,801	-0,696	0,347
11	0,804	1,126	0,368
12	0,814	0,215	0,393
13	0,801	-0,359	0,425
14	0,806	0,154	0,465
15	0,807	0,203	0,520
16	0,807	0,069	0,601
17	0,806	-0,084	0,736
18	0,804	0,147	1,040

De acordo com a Tabela 5.12, assim como no modelo GAS(1,1) com distribuição gama, o vetor *score* ∇_t apresenta variações mais significativas comparadas as variações dos parâmetros f_t e S_t ao longo das linhas do triângulo de *run-off*.

Na Tabela 5.13 estão descritos, para o triângulo contendo os dados dos sinistros da cobertura Casco (Triângulo 5.1), os parâmetros denominados estáticos, os parâmetros que mudam de acordo com a coluna e o parâmetro σ da distribuição log-normal.

Tabela 5.13: Cobertura Casco - estimativas dos parâmetros estáticos (vetor ψ). Modelo GAS (1,1) com distribuição log-normal.

ω	A	B	σ						
1,342	0,019	-0,666	0,079						
(1,077)	(0,015)	(0,385)	(1,114)						
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	
5,965	4,318	2,396	1,135	0,485	0,011	-0,519	-1,027	-1,339	
(0,610)	(0,611)	(0,619)	(0,612)	(0,611)	(0,612)	(0,618)	(0,614)	(0,613)	
λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	
-1,458	-2,117	-2,272	-2,741	-2,410	-3,004	-3,292	-4,380	-3,381	
(0,616)	(0,614)	(0,616)	(0,617)	(0,619)	(0,622)	(0,631)	(0,640)	(0,669)	

Como pode ser visto na Tabela 5.13, o parâmetro variante no tempo f_t depende negativamente do seu passado (B=-0,666) e positivamente do vetor *score* ponderado (A=0,019).

Os resíduos quantílicos e de Pearson encontram-se descritos nas Tabelas 5.14 e 5.15.

Tabela 5.14: Cobertura Casco - resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.

Trim. origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 ^o T09	-0,89	0,46	-0,48	-0,08	-2,23	-1,08	-1,08	-1,09	-0,90	-1,29	-0,18	-0,16	-1,97	-0,43	-0,31	0,80	-0,73	0,06
2 ^o T09	-0,53	0,49	-0,74	-1,01	-0,10	-1,45	-0,30	-0,36	-0,12	1,96	0,78	-1,96	-0,15	1,58	1,15	0,80	0,73	
3 ^o T09	-0,72	0,06	-0,74	-0,15	-0,13	0,97	-0,05	-0,68	-1,81	0,95	-1,18	-0,19	2,50	-0,56	1,47	-1,56		
4 ^o T10	-0,52	0,61	0,83	0,99	0,69	-0,33	0,03	-0,48	-0,76	-0,65	-0,09	0,47	4,47	0,85	-2,24			
1 ^o T10	-0,65	0,62	0,52	-0,34	-0,72	1,94	-0,37	1,54	-0,03	1,18	-0,77	2,74	-2,73	-1,38				
2 ^o T10	-0,35	0,88	0,03	0,19	0,47	-1,45	1,42	0,13	-0,57	-1,54	0,08	1,33	-1,76					
3 ^o T10	-0,05	-0,58	-0,14	0,09	0,17	1,06	0,18	0,63	2,99	0,03	1,26	-2,10						
4 ^o T10	0,12	-0,34	0,64	0,34	1,19	0,29	0,21	0,58	2,06	0,54	0,27							
1 ^o T11	0,26	-0,56	-0,01	0,42	-0,21	0,64	-0,60	-1,01	0,67	-1,11								
2 ^o T11	0,40	-0,29	-0,32	-0,38	-0,92	0,08	0,96	-0,10	-1,40									
3 ^o T11	0,43	-0,42	0,20	-0,33	1,12	1,67	0,43	1,02										
4 ^o T11	0,26	0,01	0,48	0,33	0,73	-0,67	-0,68											
1 ^o T12	0,27	0,34	0,02	-0,11	-0,21	-1,52												
2 ^o T12	0,38	-0,25	-0,03	-0,18	0,69													
3 ^o T12	0,39	0,04	-0,12	0,41														
4 ^o T12	0,37	-0,14	0,02															
1 ^o T13	0,43	-0,76																
2 ^o T13	0,58																	

Tabela 5.15: Cobertura Casco - resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.

Trim. origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 ^o T09	-0,88	0,33	-0,55	-0,21	-1,70	-1,01	-1,01	-1,02	-0,88	-1,15	-0,30	-0,28	-1,72	-0,52	-0,41	0,71	-0,75	-0,08
2 ^o T09	-0,60	0,36	-0,77	-0,96	-0,23	-1,26	-0,41	-0,46	-0,24	2,32	0,69	-1,55	-0,28	1,74	1,14	0,71	0,63	
3 ^o T09	-0,75	-0,08	-0,76	-0,27	-0,26	0,91	-0,18	-0,72	-1,47	0,89	-1,08	-0,31	2,89	-0,62	1,59	-1,33		
4 ^o T10	-0,59	0,50	0,75	0,94	0,59	-0,43	-0,11	-0,56	-0,78	-0,69	-0,22	0,34	3,90	0,76	-1,70			
1 ^o T10	-0,70	0,50	0,40	-0,44	-0,75	2,28	-0,47	1,69	-0,16	1,18	-0,79	3,07	-1,93	-1,21				
2 ^o T10	-0,45	0,81	-0,11	0,05	0,34	-1,26	1,51	-0,01	-0,64	-1,31	-0,06	1,38	-1,44					
3 ^o T10	-0,18	-0,64	-0,27	-0,05	0,02	1,03	0,03	0,51	3,30	-0,10	1,29	-1,63						
4 ^o T10	-0,02	-0,44	0,53	0,20	1,19	0,15	0,07	0,46	2,50	0,42	0,12							
1 ^o T11	0,12	-0,63	-0,15	0,28	-0,63	0,53	-0,65	-0,96	0,56	-1,03								
2 ^o T11	0,26	-0,39	-0,42	-0,48	-0,90	-0,06	0,90	-0,23	-1,23									
3 ^o T11	0,30	-0,51	0,06	-0,43	1,10	1,88	0,30	0,98										
4 ^o T11	0,11	-0,12	0,35	0,19	0,63	-0,71	-0,72											
1 ^o T12	0,13	0,20	-0,12	-0,24	-0,33	-1,30												
2 ^o T12	0,24	-0,36	-0,16	-0,30	0,58													
3 ^o T12	0,25	-0,09	-0,25	0,28														
4 ^o T12	0,23	-0,26	-0,12															
1 ^o T13	0,30	-0,78																
2 ^o T13	0,45																	

PUC-Rio - Certificação Digital Nº 1121532/CA

Como pode ser observado nas Tabelas 5.8 e 5.9, menos do que 5,0% das posições apresentam valores absolutos superiores a 2,0, sob normalidade desses resíduos. Isso indica a presença de poucos *outliers*. A correlação entre as duas séries de resíduos - de Pearson e quantílico - foi de 98,2%, indicando uma forte relação entre eles.

As Figuras 5.16, 5.17, 5.18 e 5.19 apresentam os resíduos quantílicos e de Pearson empilhados, as funções de autocorrelação associadas e os Q-Q plotes.

Figura 5.16: Casco - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição log-normal.

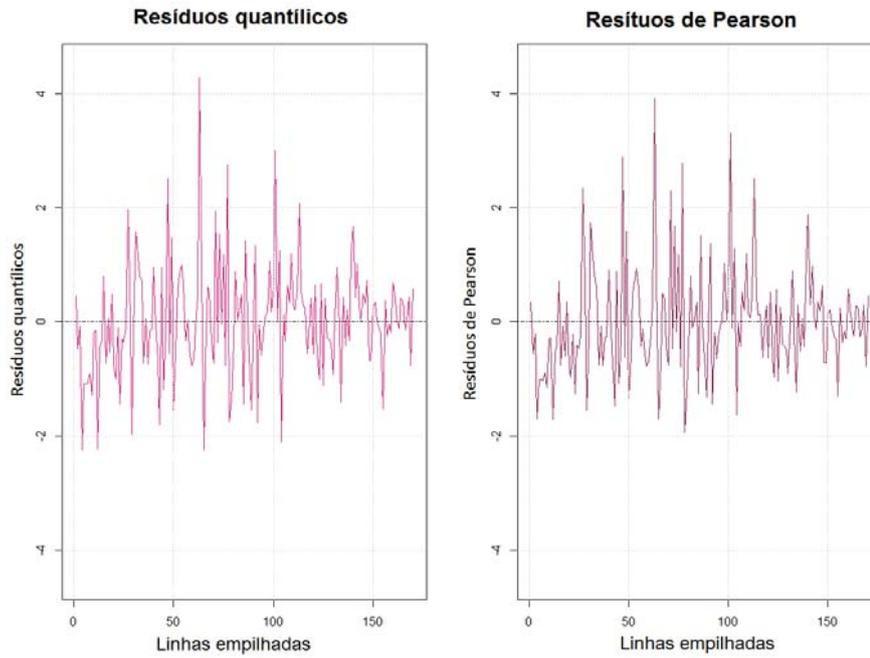


Figura 5.17: Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.

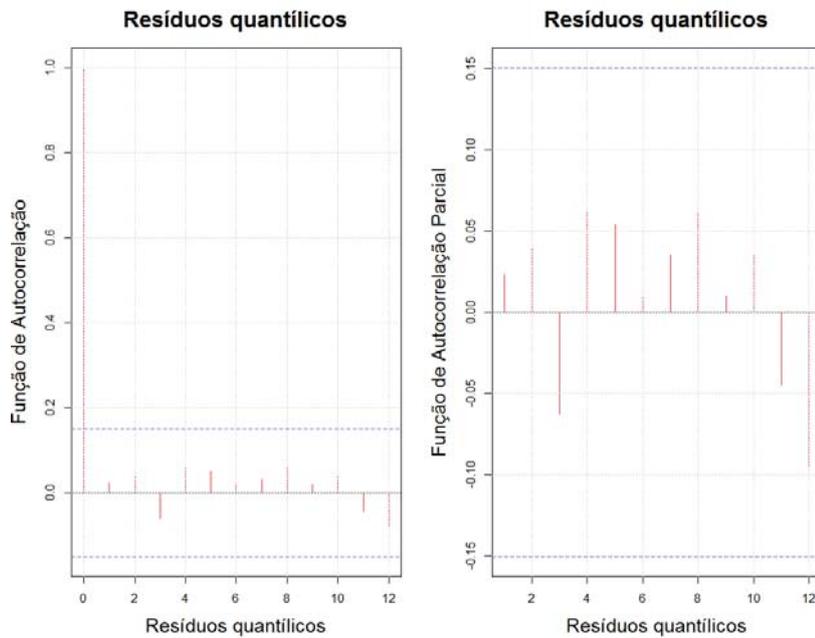
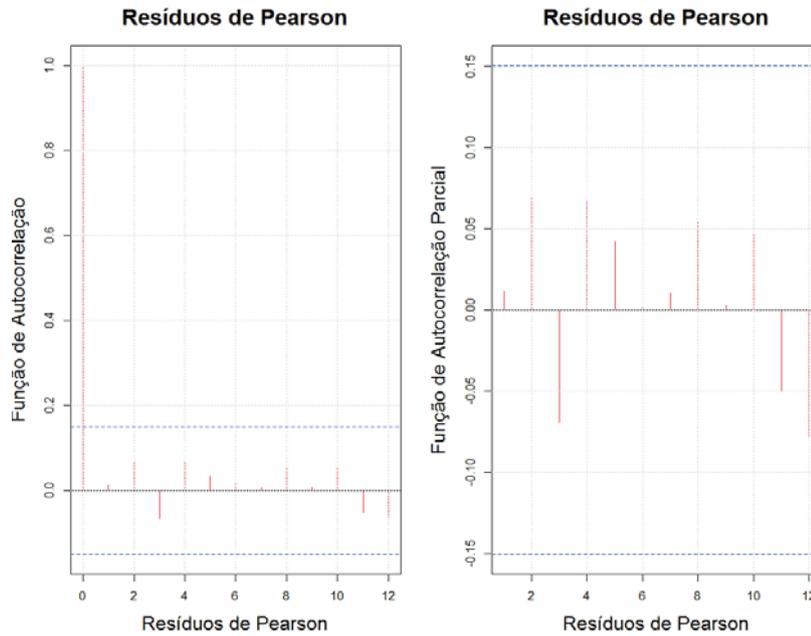
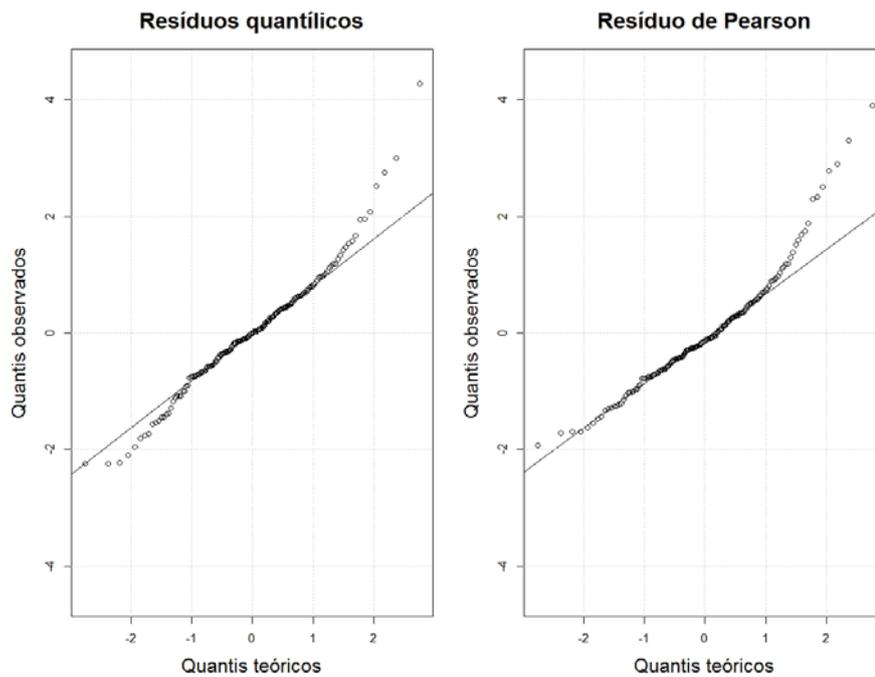


Figura 5.18: Casco - função de autocorrelação e função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.



Como pode ser visto nas Figuras 5.17 e 5.18, não há evidências de dependência linear nos resíduos quantílicos e de Pearson.

Figura 5.19: Casco - Q-Q plotes dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.



Como pode ser observado nos Q-Q plotes, os pontos se ajustam bem as retas apenas nas partes centrais dos gráficos. Há nos extremos alguns dados que não se ajustam às retas. Esses dados podem comprometer a normalidade das inovações padronizadas.

Para auxiliar na análise residual do modelo GAS (1,1), com distribuição log-normal, foram feitos os testes de Heterocedasticidade, Box-Ljung e Jarque-Bera nos resíduos quantílicos e de Pearson (Tabela 5.16).

Tabela 5.16: Cobertura Casco - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição log-normal.

Série/teste	estatísticas de teste e <i>p-valores</i>		
	Homocedasticidade (Qui-Quadrado)	Autocorrelação (Ljung-Box)	Normalidade (Jarque-Bera)
resíduo de Pearson	0,348 (0,999)	23,953 (0,774)	57,390 (0,000)
resíduo quantílico	0,363 (0,998)	23,593 (0,790)	35,615 (0,001)

Nos resíduos de Pearson e nos resíduos quantílicos, o teste de heterocedasticidade o *p-valor* não sugeriu a rejeição de H_0 . No teste Box-Ljung não há evidências para rejeitar a hipótese nula, isto é, os resíduos são descorrelatados. No teste de normalidade Jarque-Bera, rejeita-se a hipótese nula, a hipótese de normalidade dos resíduos de Pearson e dos resíduos quantílicos.

A não normalidade dos resíduos provavelmente ocorre devido as posições mal ajustadas nos extremos (ver Figura 5.19). Um modelo com variáveis *dummies* pode ser ajustado para acomodar essas posições mal ajustadas.

5.2.1.4

GAS (1,1) com distribuição log-normal - reserva IBNR e intervalos de confiança

Na Tabela 5.17, estão descritos os intervalos de confiança (Subseção 4.2.3) e as reservas IBNR por linhas e totais.

Tabela 5.17: Cobertura Casco - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição log-normal.

GAS (1,1) com distribuição log-normal			
Intervalo de Confiança de 95,0%			
Trimestre	2,5%	Reservas estimadas	97,5%
2 ^o trim.09	R\$ 43.298	R\$ 79.396	R\$ 134.712
3 ^o trim.09	R\$ 73.982	R\$ 118.028	R\$ 183.465
4 ^o trim.09	R\$ 144.346	R\$ 197.011	R\$ 271.675
1 ^o trim.10	R\$ 254.695	R\$ 334.223	R\$ 437.974
2 ^o trim.10	R\$ 417.836	R\$ 525.486	R\$ 676.075
3 ^o trim.10	R\$ 552.264	R\$ 698.263	R\$ 893.493
4 ^o trim.10	R\$ 740.427	R\$ 938.044	R\$ 1.205.470
1 ^o trim.11	R\$ 1.016.126	R\$ 1.271.350	R\$ 1.711.809
2 ^o trim.11	R\$ 1.223.759	R\$ 1.694.443	R\$ 2.178.414
3 ^o trim.11	R\$ 1.879.290	R\$ 2.443.478	R\$ 3.376.561
4 ^o trim.11	R\$ 2.047.181	R\$ 3.228.659	R\$ 4.245.666
1 ^o trim.12	R\$ 3.599.105	R\$ 4.705.416	R\$ 6.693.246
2 ^o trim.12	R\$ 3.789.619	R\$ 6.775.136	R\$ 8.774.454
3 ^o trim.12	R\$ 8.300.403	R\$ 11.183.181	R\$ 16.090.125
4 ^o trim.12	R\$ 8.014.019	R\$ 17.585.975	R\$ 23.500.875
1 ^o trim.13	R\$ 28.057.806	R\$ 44.838.971	R\$ 67.526.055
2 ^o trim.13	R\$ 45.124.937	R\$ 209.141.297	R\$ 288.096.062
Total	R\$ 245.179.141	R\$ 305.758.346	R\$ 413.807.187

Como pode ser visto na Tabela 5.17, o intervalo de confiança de 95,0% para as reservas IBNR totais é dado por:

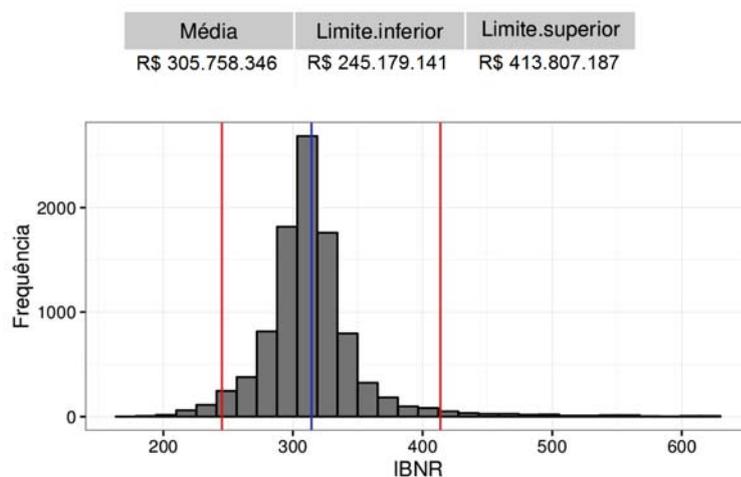
$$R\$245.179.141,00 \leq IBNR_{Total} \leq R\$413.807.187,00 \quad (5-4)$$

Ou seja, segundo o modelo GAS(1,1) com distribuição log-normal, as reservas IBNR totais têm 95,0% de chance de estarem entre R\$245.179.141,00 e R\$413.807.187,00. Os intervalos de confiança auxiliarão as seguradoras nas tomadas de decisões pautadas em maior segurança, visto que, podem a partir dos IC decidir manter uma reserva adicional.

O intervalo de confiança para as reservas IBNR totais, quando se utiliza o modelo GAS (1,1) com distribuição log-normal para estimá-las, tem amplitude maior, comparado ao intervalo de confiança para as reservas IBNR totais, quando se utiliza o modelo GAS (1,1) com distribuição gama, fato este que pode ser observado nas Tabelas 5.11 e 5.17.

A Figura 5.20 apresenta o intervalo de confiança para as reservas IBNR totais simuladas via modelo GAS (1,1) com distribuição log-normal.

Figura 5.20: Cobertura Casco - reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição log-normal (histograma).



No Anexo H, serão apresentados, para cada trimestre de origem, os gráficos contendo as 10.000 simulações que geraram os intervalos de confiança descritos na Tabela 5.17.

5.2.2

Previsão fora do triângulo (efeito linha ou reservas de prêmio) - Cobertura Casco

As Tabelas 5.18 e 5.19 apresentam os sinistros ocorridos e avisados (somatório por linha), os sinistros IBNR estimados e os sinistros “IBNR futuros” totais para quatro trimestres de origem, ou seja, os sinistros que ocorreram nos quatro trimestres seguintes aos apresentados nas Tabelas 5.1 e 5.2, que por algum motivo, não foram imediatamente informados, em sua totalidade, à seguradora.

A falta dos dados para este período inviabiliza os testes nas estimativas das reservas de prêmio. Contudo, podemos analisar a soma dos sinistros ocorridos e avisados por trimestres de ocorrência com as reservas IBNR estimadas pelos modelos propostos (Tabela 5.38). Essa soma ajudará a verificar se as reservas de prêmio são coerentes.

Tabela 5.18: Cobertura Casco - previsão fora da amostra (GAS (1,1) com distribuição gama).

GAS (1,1) com distribuição gama			
Trimestre de origem t	Sinistros ocorridos e avisados (A) Somatório por linha da Tabela 5.1	IBNR estimado por linha t (B) Tabela 5.38	Total por linha t (A+B)
1º trim. 09	R\$ 904,554,611	-	R\$ 904,554,611
2º trim. 09	R\$ 928,667,966	R\$ 71,751	R\$ 928,739,717
3º trim. 09	R\$ 954,791,394	R\$ 108,721	R\$ 954,900,115
4º trim. 09	R\$ 970,806,091	R\$ 183,876	R\$ 970,989,967
1º trim. 10	R\$ 1,012,629,209	R\$ 325,121	R\$ 1,012,954,330
2º trim. 10	R\$ 1,027,345,558	R\$ 509,150	R\$ 1,027,854,708
3º trim. 10	R\$ 1,052,709,678	R\$ 718,211	R\$ 1,053,427,889
4º trim. 10	R\$ 1,119,048,884	R\$ 990,344	R\$ 1,120,039,228
1º trim. 11	R\$ 1,149,377,474	R\$ 1,278,685	R\$ 1,150,656,159
2º trim. 11	R\$ 1,135,651,210	R\$ 1,751,508	R\$ 1,137,402,718
3º trim. 11	R\$ 1,183,489,986	R\$ 2,457,480	R\$ 1,185,947,467
4º trim. 11	R\$ 1,175,354,419	R\$ 3,361,105	R\$ 1,178,715,524
1º trim. 12	R\$ 1,148,467,539	R\$ 4,566,859	R\$ 1,153,034,399
2º trim. 12	R\$ 1,160,155,359	R\$ 7,031,399	R\$ 1,167,186,758
3º trim. 12	R\$ 1,179,710,743	R\$ 10,853,313	R\$ 1,190,564,056
4º trim. 12	R\$ 1,156,435,847	R\$ 17,882,216	R\$ 1,174,318,063
1º trim. 13	R\$ 1,120,912,599	R\$ 42,629,744	R\$ 1,163,542,344
2º trim. 13	R\$ 1,020,015,724	R\$ 210,843,198	R\$ 1,230,858,921
			Reservas de prêmio.
3º trim. 13	-	-	R\$ 1,106,099,644
4º trim. 13	-	-	R\$ 1,078,500,681
1º trim. 14	-	-	R\$ 1,095,991,016
2º trim. 14	-	-	R\$ 1,084,823,143

Tabela 5.19: Cobertura Casco - previsão fora da amostra (GAS (1,1) com distribuição log-normal).

GAS (1,1) com distribuição log-normal			
Trimestre de origem t	Sinistros ocorridos e avisados (A) Somatório por linha da Tabela 5.1	IBNR estimado por linha t (B) Tabela 5.38	Total por linha t (A+B)
1º trim. 09	R\$ 904,554,611	-	R\$ 904,554,611
2º trim. 09	R\$ 928,667,966	R\$ 75,220	R\$ 928,743,187
3º trim. 09	R\$ 954,791,394	R\$ 112,614	R\$ 954,904,008
4º trim. 09	R\$ 970,806,091	R\$ 188,516	R\$ 970,994,607
1º trim. 10	R\$ 1,012,629,209	R\$ 323,494	R\$ 1,012,952,703
2º trim. 10	R\$ 1,027,345,558	R\$ 509,387	R\$ 1,027,854,945
3º trim. 10	R\$ 1,052,709,678	R\$ 674,604	R\$ 1,053,384,281
4º trim. 10	R\$ 1,119,048,884	R\$ 922,604	R\$ 1,119,971,488
1º trim. 11	R\$ 1,149,377,474	R\$ 1,218,479	R\$ 1,150,595,953
2º trim. 11	R\$ 1,135,651,210	R\$ 1,686,129	R\$ 1,137,337,339
3º trim. 11	R\$ 1,183,489,986	R\$ 2,355,216	R\$ 1,185,845,202
4º trim. 11	R\$ 1,175,354,419	R\$ 3,251,133	R\$ 1,178,605,552
1º trim. 12	R\$ 1,148,467,539	R\$ 4,505,367	R\$ 1,152,972,906
2º trim. 12	R\$ 1,160,155,359	R\$ 6,910,524	R\$ 1,167,065,883
3º trim. 12	R\$ 1,179,710,743	R\$ 10,758,984	R\$ 1,190,469,727
4º trim. 12	R\$ 1,156,435,847	R\$ 17,996,991	R\$ 1,174,432,838
1º trim. 13	R\$ 1,120,912,599	R\$ 43,530,216	R\$ 1,164,442,815
2º trim. 13	R\$ 1,020,015,724	R\$ 217,127,496	R\$ 1,237,143,219
			Reservas de prêmio.
3º trim. 13	-	-	R\$ 1,128,984,269
4º trim. 13	-	-	R\$ 1,122,205,772
1º trim. 14	-	-	R\$ 1,126,714,846
2º trim. 14	-	-	R\$ 1,123,710,366

Como pode ser observado nas Tabelas 5.18 e 5.19, ao somar por linhas os sinistros ocorridos e avisados (A) e as reservas IBNR estimadas pelos modelos propostos (B), encontramos valores compatíveis com as estimativas das reservas de prêmio (partes cinzas das Tabelas 5.18 e 5.19).

5.2.3

Cobertura RCFV

5.2.3.1

Modelo GAS (1,1) com distribuição gama

Utilizando o triângulo de *run-off* com dados IBNR da cobertura RCFV (Tabela 5.2) e o modelo GAS(1,1) com distribuição gama, foram estimados os seguintes parâmetros variantes no tempo (Tabela 5.20):

Tabela 5.20: GAS (1,1) com distribuição gama - estimativas dos parâmetros variantes no tempo (cobertura RCFV).

Parâmetros variantes no tempo			
t	f_t	∇_t	S_t
1	2,715	-5,279	0,060
2	2,699	36,978	0,062
3	2,815	30,680	0,060
4	2,911	-1,293	0,059
5	2,911	14,874	0,061
6	2,962	-19,087	0,062
7	2,907	-5,712	0,067
8	2,892	34,309	0,070
9	3,018	-8,117	0,069
10	2,996	22,809	0,074
11	3,087	-6,540	0,075
12	3,070	-1,276	0,080
13	3,073	-16,990	0,087
14	3,006	-2,263	0,098
15	3,001	-4,355	0,110
16	2,983	0,899	0,128
17	2,994	-0,779	0,156
18	2,994	6,765	0,221

Como pode ser observado na Tabela 5.20, o *score* ∇_t apresenta variação significativa ao longo do tempo. O parâmetro em questão depende de dois outros parâmetros (α_t e β_i) e da própria variável $y_{t,i}$, como pode ser visto na Equação 4-40.

Na Tabela 5.7 estão descritos, para a série RCFV, os parâmetros denominados estáticos e os parâmetros que mudam de acordo com a coluna.

Tabela 5.21: Cobertura RCFV - estimativa dos parâmetros estáticos (vetor ψ). Modelo GAS (1,1) com distribuição gama.

ω	A	B							
-0,059 (0,375)	0,051 (0,001)	1,022 (0,051)							
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	
2,152 (0,021)	1,143 (0,020)	-0,114 (0,019)	-0,883 (0,017)	-1,263 (0,018)	-1,389 (0,019)	-1,488 (0,020)	-1,604 (0,020)	-1,906 (0,021)	
λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	
-2,137 (0,021)	-2,191 (0,020)	-2,358 (0,024)	-2,473 (0,025)	-2,582 (0,027)	-2,565 (0,030)	-3,006 (0,037)	-2,717 (0,047)	-3,761 (0,077)	

A estimativa do parâmetro estático B é superior a um (Tabela 5.21). Foi testado, tanto para a série RCFV quanto para a série Casco, a restrição para tal parâmetro (estar entre -1 e 1). Neste experimento, o tempo computacional quintuplicou e os resultados dos testes e as reservas IBNR não sofreram grandes mudanças, por esses motivos a opção de utilizar o modelo com a restrição foi preterida.

Como pode ser visto na Tabela 5.21, o parâmetro variante no tempo f_t depende positivamente do seu passado ($B=1,022$) e positivamente do vetor *score* ponderado ($A=0,051$).

Os resíduos quantílicos e de Pearson encontram-se descritos nas Tabelas 5.22 e 5.23.

Tabela 5.22: Cobertura RCFV - resíduos quantílicos, modelo GAS (1,1) com distribuição gama.

Trim. origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1º T 09	- 0,92	0,50	0,63	0,25	- 1,01	-0,57	0,10	-0,56	- 2,03	0,13	0,60	1,89	-0,84	-0,97	1,01	-1,03	0,82	0,59
2º T 09	- 0,32	0,96	0,05	0,44	1,82	1,12	1,64	-0,21	0,76	2,35	0,43	-0,72	0,77	-1,07	1,81	-0,42	0,20	
3º T 09	- 0,53	0,41	0,02	0,79	0,24	1,53	-1,18	-0,22	0,81	0,48	0,13	-1,34	1,59	1,48	1,24	2,04		
4º T 09	- 0,83	0,77	1,04	0,55	0,00	-0,77	0,11	2,02	1,02	-1,87	-1,24	1,98	1,30	0,68	- 4,11			
1º T 10	- 1,07	0,72	1,30	1,17	0,71	0,06	1,60	0,78	- 1,00	-0,90	0,91	1,20	-1,09	0,48				
2º T 10	- 0,77	0,91	- 0,62	-1,07	0,27	0,84	1,02	-1,31	- 1,44	0,80	0,37	-1,73	-1,01					
3º T 10	0,18	- 0,05	- 0,10	0,14	1,52	-0,31	-1,28	0,03	0,83	-0,38	-1,54	-0,55						
4º T 10	0,41	0,64	0,97	1,60	1,52	-0,94	-0,63	1,21	1,29	0,39	1,65							
1º T 11	- 0,33	- 0,43	- 0,07	0,24	1,41	-0,39	-0,66	-0,32	0,82	0,56								
2º T 11	0,49	0,14	0,93	0,57	0,40	2,55	0,82	-0,88	0,05									
3º T 11	0,23	- 0,59	- 0,54	-0,09	0,00	-0,86	-0,45	0,70										
4º T 11	- 0,07	- 0,07	0,35	-0,02	0,22	-0,81	-0,10											
1º T 12	- 0,01	- 0,35	- 0,32	-1,26	- 1,15	-0,64												
2º T 12	0,72	- 0,05	- 0,10	-1,03	- 0,17													
3º T 12	0,91	0,02	- 1,15	-0,79														
4º T 12	1,06	- 0,24	- 0,67															
1º T 13	0,73	- 0,93																
2º T 13	1,55																	

Tabela 5.23: Cobertura RCFV - resíduos de Pearson, modelo GAS (1,1) com distribuição gama.

Trim. origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1ª T 09	-0,93	0,43	0,57	0,16	-1,00	-0,62	0,02	-0,61	-1,75	0,04	0,53	2,09	-0,86	-0,97	1,00	-1,02	0,78	0,53
2ª T 09	-0,39	0,95	-0,03	0,36	2,00	1,13	1,78	-0,29	0,72	2,73	0,35	-0,76	0,72	-1,05	1,99	-0,49	0,12	
3ª T 09	-0,58	0,34	-0,06	0,75	0,16	1,63	-1,13	-0,29	0,78	0,41	0,05	-1,26	1,70	1,57	1,28	2,29		
4ª T 10	-0,85	0,73	1,03	0,49	-0,08	-0,80	0,04	2,25	1,01	-1,67	-1,19	2,20	1,34	0,63	-2,95			
1ª T 10	-1,05	0,67	1,34	1,19	-0,74	-0,02	1,71	0,74	-0,99	-0,91	0,88	1,22	-1,07	0,41				
2ª T 10	-0,80	0,89	-0,66	-1,05	-0,34	0,81	1,02	-1,25	-1,35	0,77	0,30	-1,57	-1,00					
3ª T 10	0,10	-0,12	-0,17	0,07	1,61	-0,38	-1,22	-0,05	0,80	-0,44	-1,42	-0,60						
4ª T 10	0,34	0,59	0,96	1,71	1,61	-0,94	-0,67	1,23	1,33	0,32	1,78							
1ª T 11	-0,39	-0,49	-0,14	0,17	-1,33	-0,45	-0,70	-0,38	0,78	0,51								
2ª T 11	0,43	0,06	0,91	0,51	0,33	2,95	0,79	-0,89	-0,03									
3ª T 11	0,16	-0,64	-0,58	-0,16	-0,07	-0,87	-0,50	0,66										
4ª T 11	-0,14	-0,14	0,28	-0,09	0,15	-0,83	-0,17											
1ª T 12	-0,08	-0,40	-0,38	-1,21	-1,12	-0,68												
2ª T 12	0,68	-0,12	-0,18	-1,02	-0,24													
3ª T 12	0,89	-0,06	-1,12	-0,81														
4ª T 12	1,06	-0,30	-0,71															
1ª T 13	0,69	-0,93																
2ª T 13	1,65																	

Menos do que 4,7% das posições apresentam valores absolutos superiores a 2,0, indicando um número desprezível de *outliers*.

As Figuras 5.21, 5.22, 5.23 e 5.24 apresentam os resíduos quantílicos e de Pearson empilhados, as funções de autocorrelação correspondentes e os Q-Q plots.

Figura 5.21: RCFV - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição gama.

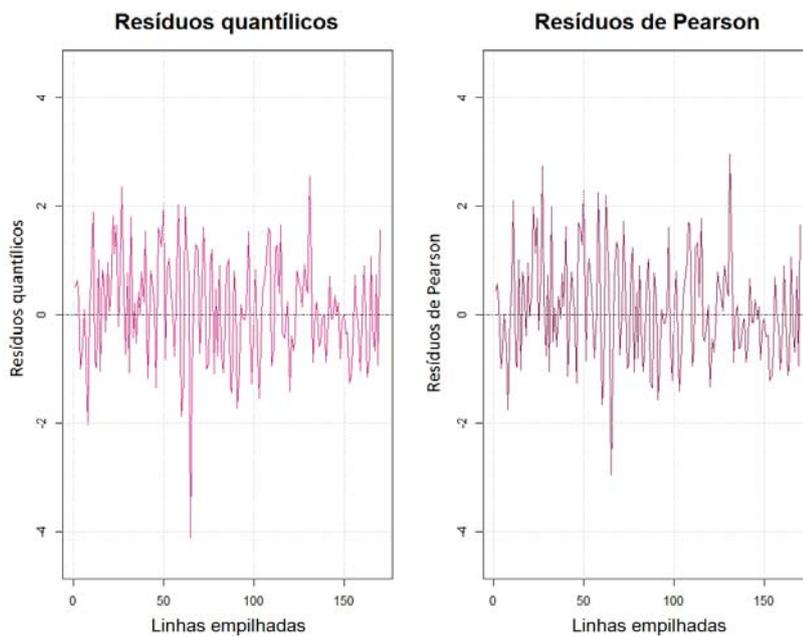


Figura 5.22: RCFV - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição gama.

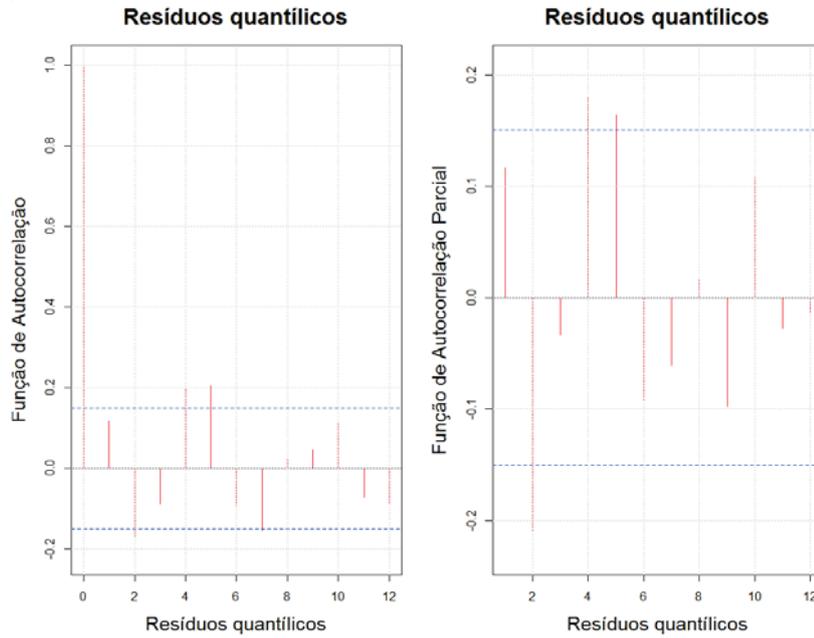
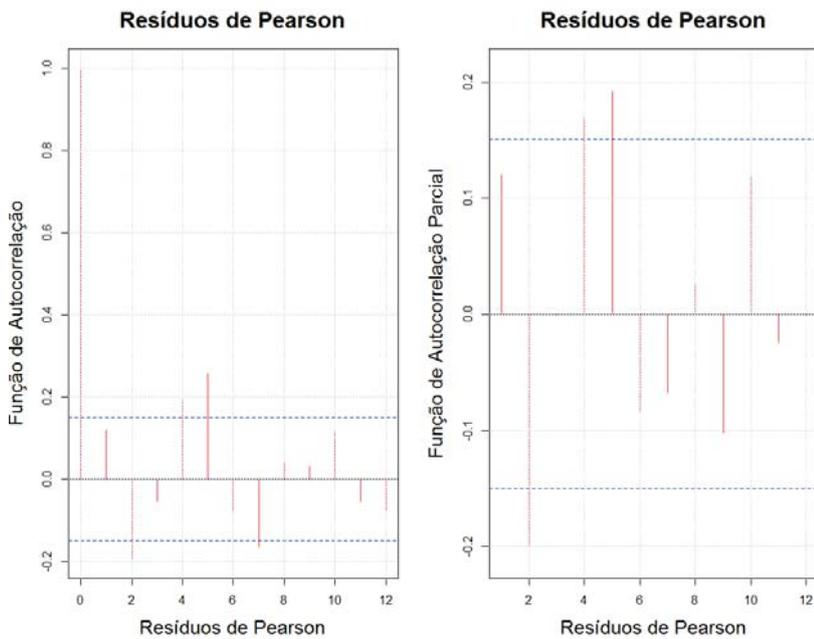
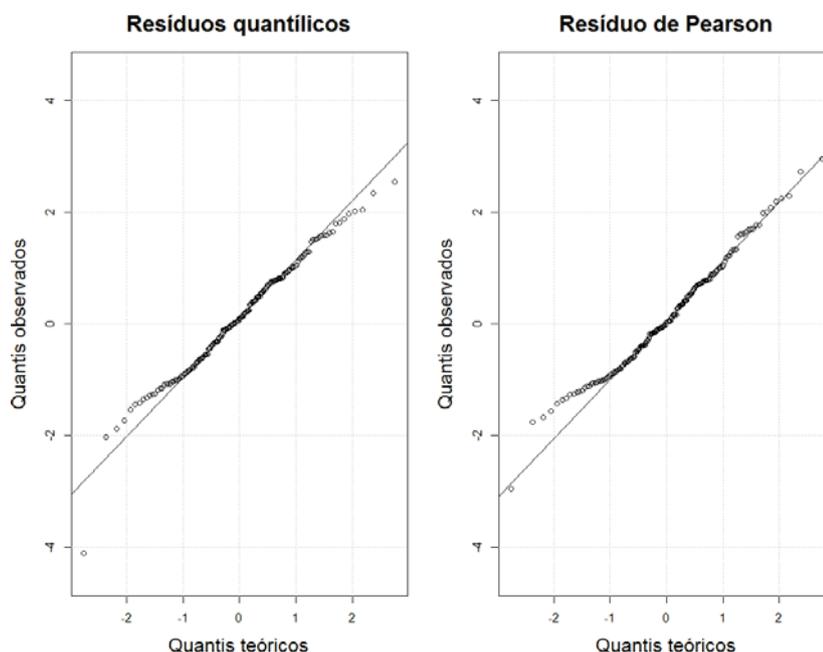


Figura 5.23: RCFV - função de autocorrelação e função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.



Como pode ser visto, os *lags* 2, 4 e 5 apresentaram uma correlação acima do limite considerável. Os testes apresentados na Tabela 5.24 auxiliam na análise.

Figura 5.24: RCFV - Q-Q plots dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição gama.



Como pode ser observado nos Q-Q plots, os pontos se ajustam bem às retas apenas nas partes centrais dos gráficos. Há nos extremos alguns dados que não se ajustam às retas.

Para auxiliar na análise residual do modelo GAS (1,1) com distribuição gama foram feitos os testes de Heterocedasticidade, Box-Ljung e Jarque-Bera nos resíduos quantílicos e de Pearson (Tabela 5.24).

Tabela 5.24: Cobertura RCFV - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição gama.

Série/teste	estatísticas de teste e <i>p</i> -valores		
	Homocedasticidade (Qui-Quadrado)	Autocorrelação (Ljung-Box)	Normalidade (Jarque-Bera)
resíduo de Pearson	0,529 (0,981)	2,479 (0,115)	2,126 (0,115)
resíduo quantílico	0,533 (0,981)	2,344 (0,125)	8,464 (0,015)

Nos resíduos de Pearson e nos resíduos quantílicos, o teste de heterocedasticidade não sugeriu a rejeição de H_0 . No teste Box-Ljung não há evidências para rejeitar a hipótese nula, isto é, os resíduos são decorrelatados.

No teste de normalidade Jarque-Bera, ao nível de significância de 1,0% aceita-se a hipótese nula, a hipótese de normalidade dos resíduos quantílicos. Ao nível de 5,0% de significância aceita-se a hipótese de normalidade dos resíduos de Pearson.

5.2.3.2

GAS (1,1) com distribuição gama - reserva IBNR e intervalo de confiança

Na Tabela 5.25, estão apresentados os intervalos de confiança para as reservas IBNR (cobertura RCFV), conforme procedimento descrito na Subseção 4.2.3.

Tabela 5.25: Cobertura RCFV - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição gama.

GAS (1,1) com distribuição gama			
Intervalo de Confiança de 95,0%			
Trimestre	2,50%	Reservas estimadas	97,5%
2 ^o trim.09	R\$ 205.500	R\$ 350.257	R\$ 544.140
3 ^o trim.09	R\$ 1.095.611	R\$ 1.514.947	R\$ 2.042.500
4 ^o trim.09	R\$ 1.955.817	R\$ 2.555.595	R\$ 3.231.294
1 ^o trim.10	R\$ 3.237.438	R\$ 3.992.836	R\$ 4.935.761
2 ^o trim.10	R\$ 4.676.257	R\$ 5.743.447	R\$ 7.175.444
3 ^o trim.10	R\$ 5.754.618	R\$ 7.026.649	R\$ 9.190.653
4 ^o trim.10	R\$ 7.321.498	R\$ 8.780.586	R\$ 11.677.278
1 ^o trim.11	R\$ 9.991.415	R\$ 12.383.909	R\$ 16.739.778
2 ^o trim.11	R\$ 11.579.552	R\$ 14.458.086	R\$ 20.772.501
3 ^o trim.11	R\$ 15.185.397	R\$ 19.502.851	R\$ 28.792.098
4 ^o trim.11	R\$ 18.158.905	R\$ 23.716.997	R\$ 37.620.272
1 ^o trim.12	R\$ 22.136.190	R\$ 29.305.481	R\$ 48.794.873
2 ^o trim.12	R\$ 24.746.726	R\$ 33.067.909	R\$ 57.892.762
3 ^o trim.12	R\$ 29.585.489	R\$ 39.478.463	R\$ 70.581.838
4 ^o trim.12	R\$ 35.687.492	R\$ 48.172.372	R\$ 90.268.272
1 ^o trim.13	R\$ 50.564.863	R\$ 68.996.916	R\$ 133.338.999
2 ^o trim.13	R\$ 100.859.046	R\$ 140.477.668	R\$ 291.968.536
Total	R\$ 346.753.303	R\$ 459.524.969	R\$ 830.502.455

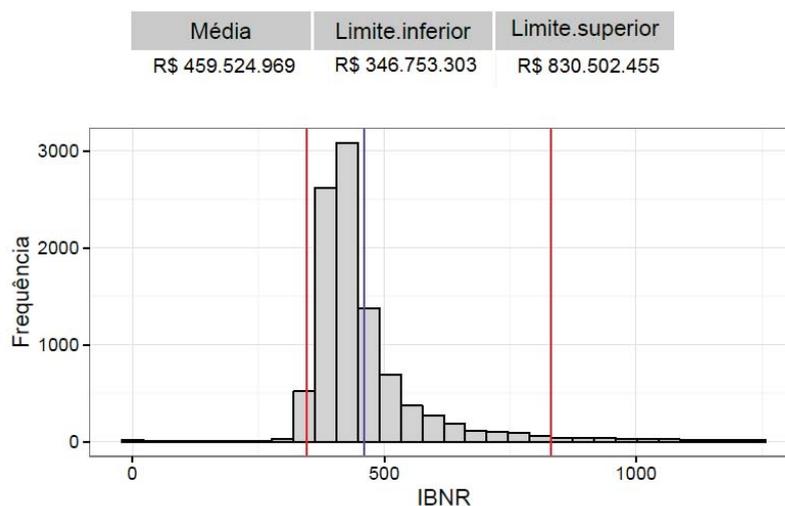
Como pode ser visto na Tabela 5.25, o intervalo de confiança de 95,0% para as reservas IBNR totais, no caso da série RCFV, é dado por:

$$R\$346.753.303,00 \leq IBNR_{Total} \leq R\$830.502.455,00 \quad (5-5)$$

Ou seja, de acordo com o modelo GAS(1,1) com distribuição gama, as reservas IBNR totais, para a cobertura RCFV, têm 95,0% de chance de estarem entre R\$346.753.303,00 e R\$830.502.455,00.

A Figura 5.25 apresenta o intervalo de confiança para as reservas IBNR totais simuladas via modelo GAS (1,1) com distribuição gama.

Figura 5.25: Cobertura RCFV - reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição gama (histograma).



No Anexo I, serão apresentados, para cada trimestre de origem, os gráficos contendo as 10.000 simulações que geraram os intervalos de confiança descritos na Tabela 5.25.

5.2.3.3 Cobertura RCFV - resultados do modelo GAS (1,1) com distribuição log-normal

Utilizando o triângulo de *run-off* com dados IBNR da cobertura RCFV e o modelo GAS(1,1) com distribuição log-normal (Subseção 4.3.2), foram estimados os seguintes parâmetros variantes no tempo (Tabela 5.26):

Tabela 5.26: GAS (1,1) com distribuição log-normal - estimativas dos parâmetros variantes no tempo (cobertura RCFV).

Parâmetros variantes no tempo			
t	f_t	∇_t	S_t
1	1,692	0,148	0,243
2	1,694	2,539	0,250
3	1,731	3,194	0,257
4	1,780	1,973	0,266
5	1,813	2,150	0,275
6	1,850	0,515	0,285
7	1,862	0,259	0,297
8	1,869	2,030	0,310
9	1,910	0,669	0,325
10	1,927	1,675	0,343
11	1,965	0,646	0,364
12	1,984	0,505	0,389
13	2,001	-0,352	0,420
14	1,998	-0,102	0,460
15	2,000	-0,239	0,515
16	1,999	-0,031	0,594
17	2,003	-0,070	0,728
18	2,006	0,312	1,029

Como pode ser observado na Tabela 5.26, ao longo das linhas do triângulo de *run-off*, o vetor *score* ∇_t apresenta variações mais significativas comparadas as variações dos parâmetros f_t e S_t .

Na Tabela 5.27 estão descritos, para o triângulo contendo os dados dos sinistros da cobertura RCFV, os parâmetros denominados estáticos, os que mudam de acordo com a coluna e o parâmetro σ da distribuição log-normal.

Tabela 5.27: Cobertura RCFV - estimativas dos parâmetros estáticos (vetor ψ) pelo modelo GAS (1,1) com distribuição log-normal.

ω	A	B	σ						
-0,031	0,059	1,018	0,057						
(0,128)	(0,019)	(0,075)	(1,114)						
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	
3,124	2,138	0,872	0,102	-0,289	-0,422	-0,520	-0,641	-0,956	
(0,310)	(0,317)	(0,311)	(0,312)	(0,318)	(0,313)	(0,312)	(0,313)	(0,315)	
λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	
-1,173	-1,224	-1,426	-1,517	-1,641	-1,703	-2,081	-1,716	-2,760	
(0,315)	(0,317)	(0,317)	(0,320)	(0,322)	(0,326)	(0,333)	(0,346)	(0,390)	

Como pode ser visto na Tabela 5.27, o parâmetro variante no tempo f_t depende positivamente do seu passado (B=1,018) e positivamente do vetor *score* ponderado (A=0,059). O erro padrão das estimativas ($EP(\lambda_i)$) cresce a medida que a coluna (i) avança. Esse fato ocorre, pois, quando i cresce (Tabela 5.2) temos menos observações para estimar cada parâmetro.

Os resíduos quantílicos e de Pearson encontram-se descritos nas Tabelas 5.28 e 5.29.

Tabela 5.28: Cobertura RCFV - resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.

Trimestre de origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1º trim. 09	-0,92	0,55	0,72	0,33	-1,03	-0,50	0,24	-0,46	-2,20	0,29	0,77	2,18	-0,76	-0,85	1,63	-0,86	0,86	0,63
2º trim. 09	-0,34	0,93	0,01	0,43	1,85	1,20	1,71	-0,17	0,92	2,38	0,50	-0,63	0,89	-1,09	2,31	-0,26	0,11	
3º trim. 09	-0,70	0,19	-0,18	0,60	0,10	1,37	-1,40	-0,32	0,77	0,38	0,02	-1,43	1,47	1,44	1,55	2,01		
4º trim. 09	-0,95	0,54	0,82	0,37	-0,11	-0,86	0,03	1,78	0,95	-2,05	-1,36	1,87	1,18	0,68	-4,44			
1º trim. 10	-1,12	0,57	1,13	1,03	-0,75	0,06	1,48	0,76	-0,95	-0,91	0,86	1,27	-1,07	0,57				
2º trim. 10	-0,78	0,73	-0,68	-1,14	-0,29	0,78	0,94	-1,30	-1,38	0,76	0,35	-1,62	-0,95					
3º trim. 10	0,22	-0,10	-0,12	0,13	1,44	-0,24	-1,27	0,11	0,93	-0,30	-1,54	-0,35						
4º trim. 10	0,44	0,56	0,91	1,48	1,45	-0,91	-0,59	1,22	1,35	0,45	1,60							
1º trim. 11	-0,35	-0,54	-0,16	0,13	-1,43	-0,38	-0,65	-0,29	0,80	0,51								
2º trim. 11	0,48	0,06	0,81	0,50	0,39	2,24	0,80	-0,78	0,17									
3º trim. 11	0,22	-0,63	-0,54	-0,12	0,00	-0,77	-0,39	0,66										
4º trim. 11	-0,02	-0,12	0,30	-0,02	0,24	-0,69	-0,03											
1º trim. 12	0,05	-0,35	-0,29	-1,19	-1,03	-0,51												
2º trim. 12	0,74	-0,06	-0,08	-0,98	-0,09													
3º trim. 12	0,92	0,01	-1,10	-0,73														
4º trim. 12	1,06	-0,22	-0,61															
1º trim. 13	0,77	-0,91																
2º trim. 13	1,48																	

Tabela 5.29: Cobertura RCFV - resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.

Trimestre de origem t	Trimestre de desenvolvimento i																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1º trim. 09	-0,91	0,45	0,64	0,22	-0,99	-0,56	0,12	-0,53	-1,75	0,18	0,70	2,63	-0,78	-0,85	1,80	-0,86	0,80	0,53
2º trim. 09	-0,43	0,88	-0,10	0,32	2,11	1,22	1,90	-0,28	0,86	2,96	0,39	-0,68	0,84	-1,03	2,84	-0,36	-0,01	
3º trim. 09	-0,73	0,07	-0,28	0,50	-0,02	1,44	-1,25	-0,41	0,70	0,27	-0,10	-1,27	1,57	1,53	1,68	2,35		
4º trim. 09	-0,93	0,43	0,75	0,26	-0,22	-0,86	-0,09	1,81	0,91	-1,67	-1,23	2,15	1,19	0,59	-2,74			
1º trim. 10	-1,06	0,47	1,13	1,00	-0,77	-0,06	1,58	0,68	-0,93	-0,90	0,80	1,31	-1,02	0,47				
2º trim. 10	-0,80	0,65	-0,72	-1,07	-0,39	0,71	0,89	-1,18	-1,24	0,68	0,24	-1,40	-0,93					
3º trim. 10	0,10	-0,21	-0,23	0,01	1,54	-0,34	-1,16	-0,01	0,88	-0,40	-1,35	-0,44						
4º trim. 10	0,32	0,46	0,86	1,58	1,55	-0,90	-0,65	1,24	1,41	0,34	1,75							
1º trim. 11	-0,43	-0,60	-0,26	0,01	-1,28	-0,47	-0,69	-0,39	0,72	0,41								
2º trim. 11	0,38	-0,06	0,74	0,39	0,28	2,73	0,72	-0,80	0,05									
3º trim. 11	0,10	-0,68	-0,60	-0,23	-0,11	-0,79	-0,47	0,57										
4º trim. 11	-0,13	-0,23	0,18	-0,14	0,12	-0,72	-0,14											
1º trim. 12	-0,06	-0,44	-0,38	-1,11	-0,99	-0,58												
2º trim. 12	0,66	-0,17	-0,19	-0,95	-0,20													
3º trim. 12	0,86	-0,11	-1,04	-0,76														
4º trim. 12	1,04	-0,32	-0,66															
1º trim. 13	0,69	-0,90																
2º trim. 13	1,58																	

Menos do que 4,7% das posições apresentam valores absolutos superiores a 2,0, sob normalidade desses resíduos. Isso indica a presença de poucos *outliers*. A correlação entre as duas séries de resíduos foi de 99,2%, indicando uma forte relação entre eles.

As Figuras 5.26, 5.27, 5.28 e 5.29 apresentam os resíduos quantílicos e de Pearson empilhados, as funções de autocorrelação correspondentes e os Q-Q plots.

Figura 5.26: RCFV - resíduos quantílicos e de Pearson, modelo GAS (1,1) com distribuição log-normal.

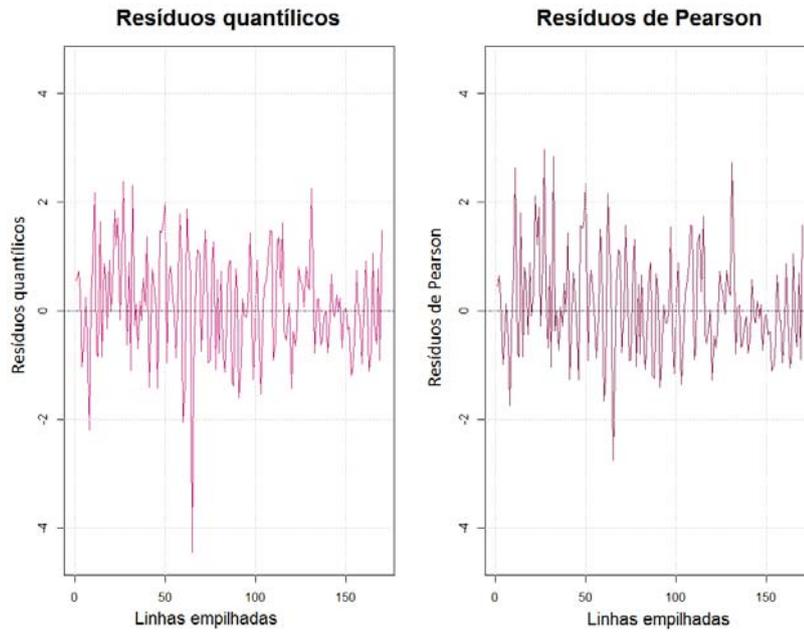


Figura 5.27: RCFV - função de autocorrelação e função de autocorrelação parcial dos resíduos quantílicos, modelo GAS (1,1) com distribuição log-normal.

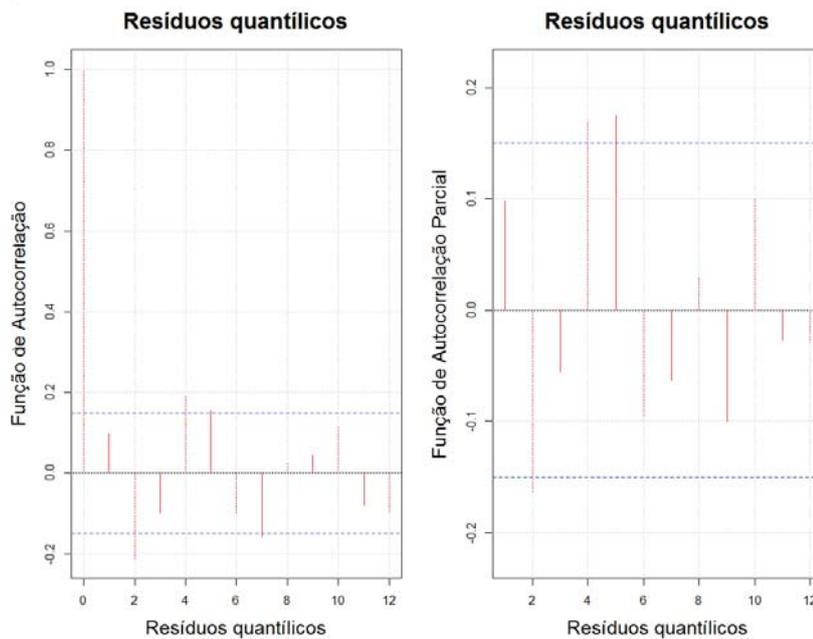
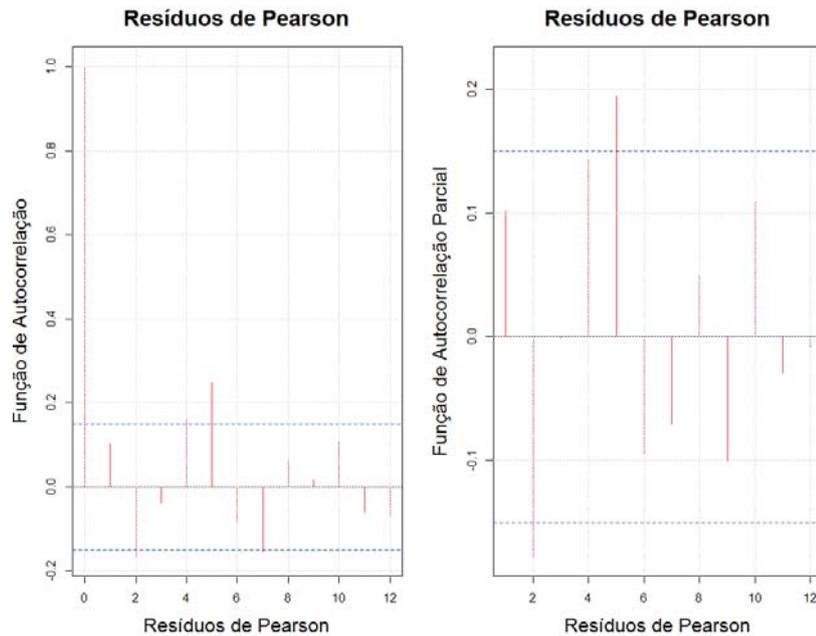
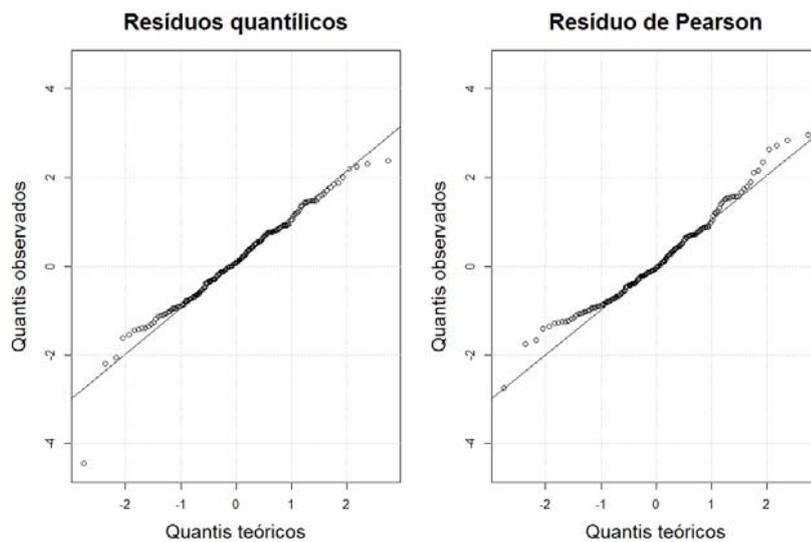


Figura 5.28: RCFV - função de autocorrelação função de autocorrelação parcial dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.



Como pode ser visto nos resíduos quantílicos (Figura 5.27), os *lags* 2 e 4 apresentaram uma correlação acima do limite considerável. Nos resíduos de Pearson, os *lags* 2, 4 e 5 apresentaram uma correlação acima do limite considerável. Os testes apresentados na Tabela 5.24 auxiliam na análise.

Figura 5.29: RCFV - Q-Q plotes dos resíduos quantílicos e dos resíduos de Pearson, modelo GAS (1,1) com distribuição log-normal.



Como pode ser observado nos Q-Q plotes, os pontos se ajustam bem às retas apenas nas partes centrais dos gráficos, tanto nos resíduos de Pearson quanto nos

resíduos quantílicos. Há nos extremos alguns dados que não se ajustam às retas.

Para auxiliar na análise residual do modelo GAS (1,1) com distribuição log-normal foram feitos os testes de Heterocedasticidade, Box-Ljung e Jarque-Bera nos resíduos quantílicos e de Pearson (Tabela 5.30).

Tabela 5.30: Cobertura RCFV - testes e diagnósticos aplicados aos resíduos, modelo GAS (1,1) com distribuição log-normal.

Série/teste	estatísticas de teste e <i>p</i> -valores		
	Homocedasticidade (Qui-Quadrado)	Autocorrelação (Ljung-Box)	Normalidade (Jarque-Bera)
Resíduo de Pearson	0,400 (0,999)	1.794 (0,1803)	6,914 (0,031)
Resíduo Quantílico	0,428 (0,998)	1.674 (0,1957)	22,265 (0,000)

Nos resíduos de Pearson e nos resíduos quantílicos, o *p*-valor não sugeriu a rejeição da hipótese de homocedasticidade. No teste Box-Ljung não há evidências para rejeitar a hipótese na qual os resíduos são descorrelatados.

No teste de normalidade Jarque-Bera, ao nível de significância de 1,0%, aceita-se a hipótese de normalidade dos resíduos de Pearson. No caso dos resíduos quantílicos rejeita-se a hipótese de normalidade.

5.2.3.4

GAS (1,1) com distribuição log-normal - reserva IBNR e intervalo de confiança

Na Tabela 5.31, estão apresentados os intervalos de confiança para as reservas IBNR (cobertura RCFV), conforme procedimento descrito na Subseção 4.2.3.

Tabela 5.31: Cobertura RCFV - reservas IBNR e intervalos de confiança, modelo GAS (1,1) com distribuição log-normal.

GAS (1,1) Distribuição log-normal			
Intervalo de Confiança de 95,0%			
Trimestre	2,50%	Reservas estimadas	97,5%
2 ^o trim.09	R\$ 198.988	R\$ 370.901	R\$ 625.639
3 ^o trim.09	R\$ 1.136.884	R\$ 1.634.701	R\$ 2.279.696
4 ^o trim.09	R\$ 2.021.659	R\$ 2.659.414	R\$ 3.407.929
1 ^o trim.10	R\$ 3.205.529	R\$ 3.900.955	R\$ 4.677.215
2 ^o trim.10	R\$ 4.678.603	R\$ 5.574.200	R\$ 6.575.588
3 ^o trim.10	R\$ 5.786.988	R\$ 6.696.323	R\$ 7.698.791
4 ^o trim.10	R\$ 7.319.356	R\$ 8.322.211	R\$ 9.447.254
1 ^o trim.11	R\$ 10.171.493	R\$ 11.948.140	R\$ 13.856.824
2 ^o trim.11	R\$ 12.065.869	R\$ 13.702.888	R\$ 15.423.241
3 ^o trim.11	R\$ 15.894.451	R\$ 18.491.877	R\$ 21.195.460
4 ^o trim.11	R\$ 19.511.047	R\$ 22.187.653	R\$ 25.151.382
1 ^o trim.12	R\$ 23.919.365	R\$ 27.165.397	R\$ 30.930.249
2 ^o trim.12	R\$ 26.668.527	R\$ 30.288.231	R\$ 35.022.294
3 ^o trim.12	R\$ 31.738.402	R\$ 35.946.225	R\$ 41.541.034
4 ^o trim.12	R\$ 38.268.519	R\$ 43.467.277	R\$ 50.590.208
1 ^o trim.13	R\$ 54.527.545	R\$ 62.146.972	R\$ 72.635.433
2 ^o trim.13	R\$ 109.477.281	R\$ 126.075.154	R\$ 149.692.988
Total	R\$ 373.127.103	R\$ 420.578.521	R\$ 482.704.363

Como pode ser visto na Tabela 5.31, o intervalo de confiança de 95,0% para as reservas IBNR totais, no caso da série RCFV, é dado por:

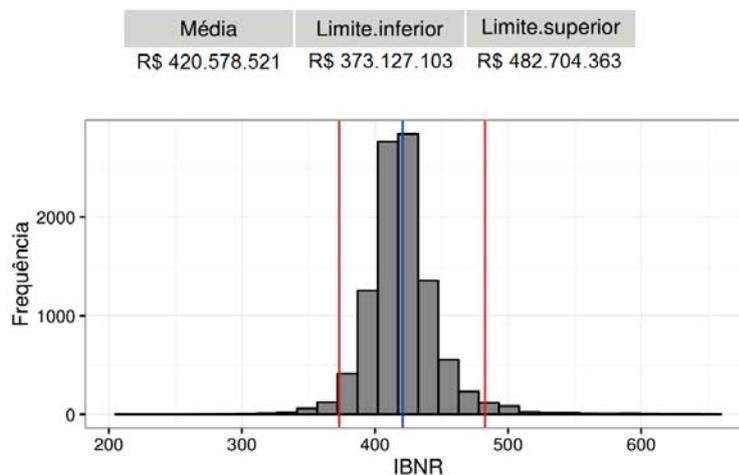
$$R\$373.127.103,00 \leq IBNR_{Total} \leq R\$482.704.363,00 \quad (5-6)$$

Ou seja, segundo o modelos GAS(1,1) com distribuição log-normal, reservas IBNR totais, para a cobertura RCFV, têm 95,0% de chance de estarem entre R\$373.127.103,00 e R\$482.704.363,00.

Como se pode observar nas Tabelas 5.25 e 5.31, a amplitude do intervalo de confiança para a reserva IBNR total (cobertura RCFV) estimada pelo modelo GAS(1,1) com distribuição log-normal é muito inferior à estimada pelo GAS(1,1) com distribuição gama.

A Figura 5.30 apresenta o intervalo de confiança para as reservas IBNR totais simuladas via modelo GAS (1,1) com distribuição log-norma.

Figura 5.30: Cobertura RCFV - Reservas IBNR totais e intervalo de confiança, modelo GAS (1,1) com distribuição log-normal (histograma).



No Anexo J, serão apresentados, para cada trimestre de origem, os gráficos contendo as 10.000 simulações que geraram os intervalos de confiança descritos na Tabela 5.31.

5.2.4

Previsão fora da amostra (efeito linha no triângulo de run-off)

As Tabelas 5.32 e 5.33 apresentam os sinistros ocorridos e avisados (somatório por linha), os sinistros IBNR estimados e os sinistros “IBNR futuros” totais para quatro trimestres de origem, ou seja, os sinistros que ocorreram nos quatro trimestres seguintes aos apresentados nas Tabelas 5.1 e 5.2, que por algum motivo, não foram imediatamente informados, em sua totalidade, à seguradora.

Tabela 5.32: Cobertura RCFV - previsão fora da amostra (GAS (1,1) com distribuição gama).

Gas (1,1) com distribuição gama			
Trimestre de origem t	Sinistros ocorridos e avisados (A) Somatório por linha da Tabela 5.2	IBNR estimado por linha t (B) Tabela 5.40	Total por linha t (A+B)
1º trim. 09	R\$ 198,420,821	-	R\$ 198,420,821
2º trim. 09	R\$ 225,801,869	R\$ 345,656	R\$ 226,147,525
3º trim. 09	R\$ 233,859,271	R\$ 1,490,939	R\$ 235,350,210
4º trim. 09	R\$ 254,511,338	R\$ 2,550,550	R\$ 257,061,888
1º trim. 10	R\$ 247,422,512	R\$ 3,964,959	R\$ 251,387,472
2º trim. 10	R\$ 258,113,287	R\$ 5,633,772	R\$ 263,747,059
3º trim. 10	R\$ 266,099,002	R\$ 6,875,604	R\$ 272,974,606
4º trim. 10	R\$ 289,147,404	R\$ 8,478,097	R\$ 297,625,500
1º trim. 11	R\$ 266,728,622	R\$ 11,902,216	R\$ 278,630,838
2º trim. 11	R\$ 308,613,548	R\$ 14,003,903	R\$ 322,617,451
3º trim. 11	R\$ 300,337,499	R\$ 18,602,775	R\$ 318,940,274
4º trim. 11	R\$ 289,799,199	R\$ 22,629,835	R\$ 312,429,034
1º trim. 12	R\$ 278,186,737	R\$ 27,564,644	R\$ 305,751,381
2º trim. 12	R\$ 290,991,029	R\$ 30,814,284	R\$ 321,805,313
3º trim. 12	R\$ 289,829,587	R\$ 36,345,571	R\$ 326,175,158
4º trim. 12	R\$ 282,646,319	R\$ 43,856,015	R\$ 326,502,334
1º trim. 13	R\$ 248,057,378	R\$ 62,193,791	R\$ 310,251,169
2º trim. 13	R\$ 235,076,407	R\$ 124,800,509	R\$ 359,876,916
			Reserva de prêmio.
3º trim. 13	-	-	R\$ 321,972,740
4º trim. 13	-	-	R\$ 324,519,612
1º trim. 14	-	-	R\$ 327,142,791
2º trim. 14	-	-	R\$ 329,845,032

Tabela 5.33: Cobertura RCFV - previsão fora da amostra (GAS (1,1) com distribuição log-normal).

Gas (1,1) com distribuição log-normal			
Trimestre de origem t	Sinistros ocorridos e avisados (A) Somatório por linha da Tabela 5.2	IBNR estimado por linha t (B) Tabela 5.40	Total por linha t (A+B)
1º trim. 09	R\$ 198,420,821	-	R\$ 198,420,821
2º trim. 09	R\$ 225,801,869	R\$ 356,866	R\$ 226,158,735
3º trim. 09	R\$ 233,859,271	R\$ 1,598,846	R\$ 235,458,117
4º trim. 09	R\$ 254,511,338	R\$ 2,646,674	R\$ 257,158,012
1º trim. 10	R\$ 247,422,512	R\$ 3,881,508	R\$ 251,304,020
2º trim. 10	R\$ 258,113,287	R\$ 5,522,077	R\$ 263,635,364
3º trim. 10	R\$ 266,099,002	R\$ 6,669,807	R\$ 272,768,809
4º trim. 10	R\$ 289,147,404	R\$ 8,211,417	R\$ 297,358,820
1º trim. 11	R\$ 266,728,622	R\$ 11,817,936	R\$ 278,546,558
2º trim. 11	R\$ 308,613,548	R\$ 13,730,001	R\$ 322,343,548
3º trim. 11	R\$ 300,337,499	R\$ 18,362,483	R\$ 318,699,982
4º trim. 11	R\$ 289,799,199	R\$ 22,206,567	R\$ 312,005,766
1º trim. 12	R\$ 278,186,737	R\$ 27,046,668	R\$ 305,233,405
2º trim. 12	R\$ 290,991,029	R\$ 30,110,603	R\$ 321,101,632
3º trim. 12	R\$ 289,829,587	R\$ 35,561,205	R\$ 325,390,792
4º trim. 12	R\$ 282,646,319	R\$ 42,977,246	R\$ 325,623,565
1º trim. 13	R\$ 248,057,378	R\$ 61,408,590	R\$ 309,465,968
2º trim. 13	R\$ 235,076,407	R\$ 124,715,203	R\$ 359,791,610
			Reserva de prêmio.
3º trim. 13	-	-	R\$ 309,129,457
4º trim. 13	-	-	R\$ 311,035,878
1º trim. 14	-	-	R\$ 312,988,964
2º trim. 14	-	-	R\$ 314,990,083

Como pode ser observado nas Tabelas 5.32 e 5.33, ao somar por linhas os sinistros ocorridos e avisados (A) e as reservas IBNR estimadas pelos modelos propostos (B) encontramos valores compatíveis com as estimativas das reservas de prêmio (partes cinzas das Tabelas).

5.3

Medidas de aderências - coberturas Casco e RCFV

Nesta seção, serão apresentadas as medidas de aderência do método *Chain Ladder* e dos quatro modelos aplicados nesta tese (Atherino univariado e bivariado, GAS (1,1) com distribuição gama e log-normal).

Cabe ressaltar que, para o cálculo do MAPE, do EQM e do Pseudo- R^2 (Subseção 3.4), não se levou em consideração as observações da primeira linha do triângulo, referente ao período difuso do filtro de Kalman e também as observações da primeira coluna, em que o método *Chain Ladder* não se aplica, por construção.

Originalmente a base de dados era constituída por triângulos de *run-off* com 19 linhas e 19 colunas. Para auxiliar a decisão sobre qual modelo é mais adequado, realizou-se uma validação fora da amostra. Considerou-se para a modelagem triângulos com 18 linhas e 18 colunas, retirando a diagonal principal (17 observações) para posteriormente calcular as medidas de aderências. Cabe ressaltar que no caso

dos modelos Atherino univariado e bivariado, os valores utilizados nos teste foram os filtrados e no caso dos modelos GAS, os resultados das Equações 4-32 e 4-42.

Tabela 5.34: Cobertura Casco - medidas de aderência (dentro da amostra).

Medidas de aderência	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
MAPE (%)	31,14	23,09	23,47	23,25	22,35
EQM (+10 ¹³)	38,04	7,95	9,52	6,41	6,74
Pseudo- R^2 (%)	94,73	97,29	96,77	97,79	97,84

Observa-se na Tabela 5.34 que as medidas MAPE e Pseudo- R^2 apontam para uma melhor aderência do modelo GAS (1,1) com distribuição log-normal.

Na Tabela 5.35, estão descritas as medidas de aderência calculadas para os dados fora da amostra, cobertura Casco.

Tabela 5.35: Cobertura Casco - medidas de aderência (fora da amostra).

Medidas de aderência	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
MAPE (%)	35,03	34,75	33,88	32,80	33,67
EQM (+10 ¹³)	20,49	5,30	7,51	5,05	7,03
Pseudo- R^2 (%)	99,44	99,61	99,68	99,61	99,60

O teste fora da amostra, no caso da cobertura Casco, aponta melhor aderência para o modelo GAS (1,1) com distribuição gama, pois apresentou menor MAPE e menor EQM.

As medidas de aderência para os dados RCFV “*in the sample*” serão descritos Tabela 5.36.

Tabela 5.36: Cobertura RCFV - medidas de aderência (dentro da amostra).

Medidas de aderência	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
MAPE (%)	28,33	20,02	19,30	19,85	19,77
EQM (+10 ¹³)	41,54	5,23	8,09	8,27	7,93
Pseudo- R^2 (%)	96,18	98,56	98,26	97,60	97,81

Duas, das três medidas de aderências descritas na Tabela 5.36 apontam para uma melhor aderência do modelo Atherino univariado.

Na Tabela 5.37, estão descritas as medidas de aderência calculadas para os dados fora da amostra.

Tabela 5.37: Cobertura RCFV - medidas de aderência (fora da amostra).

Medidas de aderência	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
MAPE (%)	31,29	25,30	24,86	27,10	27,41
EQM (+10 ¹³)	15,40	1,37	2,26	2,24	2,40
Pseudo- R^2 (%)	98,63	99,34	99,33	99,35	99,33

Como pode ser visto na Tabela 5.37, cada medida de aderência aponta para um modelo, logo, o teste fora da amostra foi inconclusivo.

5.4

Reservas IBNR estimadas - coberturas Casco e RCFV

Um ponto importante a ser destacado é o aprimoramento da mensuração das reservas IBNR, bem como a diminuição dos coeficientes de variação das estimativas. É relevante por conta das estimativas eficientes que se aproximam do real valor o qual a seguradora deverá pagar pelo sinistro ocorrido e não avisado, tal como provém do objetivo primário das seguradoras garantir um melhor uso do capital. Por conseguinte, garantir uma melhor rentabilidade sobre o seu Patrimônio Líquido de forma que a solvência não seja comprometida.

A Tabela 5.38 apresenta as reservas IBNR estimadas pelo método *Chain Ladder* e pelos modelos: Atherino univariado e bivariado, GAS (1,1) com distribuição gama e com distribuição log-normal.

Tabela 5.38: Cobertura Casco - reservas IBNR estimadas.

Trim.	Reservas IBNR estimadas				
	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
2 ^o trim.09	R\$ 79.571	R\$ 87.697	R\$ 89.740	R\$ 71.751	R\$ 75.220
3 ^o trim.09	R\$ 110.766	R\$ 122.192	R\$ 127.553	R\$ 108.721	R\$ 112.614
4 ^o trim.09	R\$ 202.764	R\$ 211.778	R\$ 217.965	R\$ 183.876	R\$ 188.516
1 ^o trim.10	R\$ 340.595	R\$ 333.818	R\$ 340.248	R\$ 325.121	R\$ 323.494
2 ^o trim.10	R\$ 570.213	R\$ 557.269	R\$ 561.911	R\$ 509.150	R\$ 509.387
3 ^o trim.10	R\$ 794.820	R\$ 734.821	R\$ 751.964	R\$ 718.211	R\$ 674.604
4 ^o trim.10	R\$ 1.140.532	R\$ 987.650	R\$ 1.028.946	R\$ 990.344	R\$ 922.604
1 ^o trim.11	R\$ 1.490.262	R\$ 1.261.573	R\$ 1.320.905	R\$ 1.278.685	R\$ 1.218.479
2 ^o trim.11	R\$ 2.092.927	R\$ 1.828.497	R\$ 1.926.569	R\$ 1.751.508	R\$ 1.686.129
3 ^o trim.11	R\$ 2.929.810	R\$ 2.491.632	R\$ 2.615.079	R\$ 2.457.480	R\$ 2.355.216
4 ^o trim.11	R\$ 3.851.017	R\$ 3.343.851	R\$ 3.457.641	R\$ 3.361.105	R\$ 3.251.133
1 ^o trim.12	R\$ 5.260.001	R\$ 4.760.721	R\$ 4.843.548	R\$ 4.566.859	R\$ 4.505.367
2 ^o trim.12	R\$ 7.983.283	R\$ 7.219.269	R\$ 7.323.018	R\$ 7.031.399	R\$ 6.910.524
3 ^o trim.12	R\$ 12.341.738	R\$ 11.231.433	R\$ 11.354.550	R\$ 10.853.313	R\$ 10.758.984
4 ^o trim.12	R\$ 19.892.074	R\$ 18.682.098	R\$ 18.951.131	R\$ 17.882.216	R\$ 17.996.991
1 ^o trim.13	R\$ 46.563.124	R\$ 45.109.827	R\$ 45.877.155	R\$ 42.629.744	R\$ 43.530.216
2 ^o trim.13	R\$ 248.936.595	R\$ 222.719.082	R\$ 230.239.959	R\$ 210.843.198	R\$ 217.127.496
Total	R\$ 354.580.092	R\$ 321.683.208	R\$ 331.027.882	R\$ 305.562.682	R\$ 312.146.973

Tabela 5.39: Cobertura Casco - coeficientes de variação estimados.

Trim.	Coeficientes de Variação				
	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
2 ^o trim.09	36,0%	43,3%	43,2%	29,3%	28,6%
3 ^o trim.09	27,3%	33,4%	33,1%	21,7%	22,1%
4 ^o trim.09	21,0%	24,1%	23,7%	19,0%	18,0%
1 ^o trim.10	20,7%	19,6%	19,0%	16,1%	15,4%
2 ^o trim.10	17,4%	17,6%	16,5%	15,9%	14,7%
3 ^o trim.10	26,1%	15,7%	14,0%	14,9%	13,2%
4 ^o trim.10	22,9%	14,5%	12,0%	14,2%	12,4%
1 ^o trim.11	20,0%	13,6%	9,9%	13,4%	11,7%
2 ^o trim.11	17,4%	13,9%	10,6%	13,7%	11,9%
3 ^o trim.11	17,4%	13,5%	11,8%	13,2%	11,6%
4 ^o trim.11	14,2%	13,4%	8,2%	12,8%	11,4%
1 ^o trim.12	11,4%	13,8%	9,6%	12,8%	11,9%
2 ^o trim.12	13,8%	14,6%	10,0%	12,6%	12,8%
3 ^o trim.12	11,4%	15,3%	10,0%	12,7%	12,9%
4 ^o trim.12	188,1%	16,4%	12,8%	12,7%	13,9%
1 ^o trim.13	133,0%	20,4%	16,3%	12,5%	17,6%
2 ^o trim.13	32,9%	26,2%	24,5%	12,9%	22,8%

Observa-se que as reservas IBNR totais, estimadas pelo modelo GAS (1,1) com distribuição log-normal e pelo modelo GAS (1,1) com distribuição gama, são menores que a reserva total estimada pelos outros três modelos.

Como visto na Tabela 5.39, os coeficientes de variação estimados, utilizando modelo GAS (1,1) com distribuição log-normal, são menores em 35,3% dos períodos de desenvolvimento.

A Tabela 5.40 apresenta as reservas IBNR estimadas pelo método: *Chain Ladder* e pelos modelos: Atherino univariado e bivariado, GAS (1,1) com distribuição gama e com distribuição log-normal para a cobertura RCFV.

Tabela 5.40: Cobertura RCFV - reservas IBNR estimados.

Trim.	Reservas IBNR estimadas				
	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
2 ^o trim.09	R\$ 455.112	R\$ 441.236	R\$ 443.171	R\$ 345.656	R\$ 356.866
3 ^o trim.09	R\$ 1.700.425	R\$ 1.653.500	R\$ 1.634.690	R\$ 1.490.939	R\$ 1.598.846
4 ^o trim.09	R\$ 2.825.468	R\$ 2.567.117	R\$ 2.541.211	R\$ 2.550.550	R\$ 2.646.674
1 ^o trim.10	R\$ 4.221.393	R\$ 3.747.349	R\$ 3.752.360	R\$ 3.964.959	R\$ 3.881.508
2 ^o trim.10	R\$ 5.905.591	R\$ 5.147.239	R\$ 5.127.717	R\$ 5.633.772	R\$ 5.522.077
3 ^o trim.10	R\$ 7.834.896	R\$ 6.870.035	R\$ 6.862.344	R\$ 6.875.604	R\$ 6.669.807
4 ^o trim.10	R\$ 10.674.355	R\$ 8.900.303	R\$ 8.667.537	R\$ 8.478.097	R\$ 8.211.417
1 ^o trim.11	R\$ 12.169.916	R\$ 11.370.296	R\$ 10.727.636	R\$ 11.902.216	R\$ 11.817.936
2 ^o trim.11	R\$ 16.996.994	R\$ 13.928.740	R\$ 13.098.889	R\$ 14.003.903	R\$ 13.730.001
3 ^o trim.11	R\$ 20.107.990	R\$ 16.900.112	R\$ 16.158.163	R\$ 18.602.775	R\$ 18.362.483
4 ^o trim.11	R\$ 24.147.477	R\$ 20.605.042	R\$ 19.889.041	R\$ 22.629.835	R\$ 22.206.567
1 ^o trim.12	R\$ 28.410.014	R\$ 24.546.765	R\$ 24.031.915	R\$ 27.564.644	R\$ 27.046.668
2 ^o trim.12	R\$ 35.984.032	R\$ 29.209.946	R\$ 29.005.835	R\$ 30.814.284	R\$ 30.110.603
3 ^o trim.12	R\$ 43.137.541	R\$ 34.829.964	R\$ 34.738.337	R\$ 36.345.571	R\$ 35.561.205
4 ^o trim.12	R\$ 52.972.839	R\$ 43.086.664	R\$ 43.176.837	R\$ 43.856.015	R\$ 42.977.246
1 ^o trim.13	R\$ 69.284.893	R\$ 61.205.441	R\$ 61.435.169	R\$ 62.193.791	R\$ 61.408.590
2 ^o trim.13	R\$ 177.751.781	R\$ 122.505.698	R\$ 126.640.991	R\$ 124.800.509	R\$ 124.715.203
Total	R\$ 514.580.717	R\$ 407.515.446	R\$ 407.931.842	R\$ 422.053.121	R\$ 416.823.697

Como pode ser observado na Tabela 5.40, as reservas totais estimadas pelos modelos Atherino univariado e bivariado são as menores.

Tabela 5.41: Cobertura RCFV - coeficientes de variação estimados.

Trim.	Coeficientes de Variação				
	Chain Ladder	Atherino univariado	Atherino bivariado	GAS (1,1) gama	GAS (1,1) log-normal
2 ^o trim.09	77,2%	34,7%	36,1%	25,8%	24,1%
3 ^o trim.09	26,0%	22,8%	23,4%	19,1%	19,0%
4 ^o trim.09	22,0%	18,0%	18,2%	14,8%	15,1%
1 ^o trim.10	22,0%	15,3%	15,2%	13,2%	12,9%
2 ^o trim.10	17,8%	13,7%	13,4%	11,5%	11,6%
3 ^o trim.10	15,0%	12,6%	12,2%	10,9%	10,3%
4 ^o trim.10	14,2%	11,8%	11,2%	10,4%	9,6%
1 ^o trim.11	12,8%	11,2%	10,6%	9,2%	9,1%
2 ^o trim.11	11,8%	10,8%	10,1%	8,8%	8,7%
3 ^o trim.11	10,1%	10,6%	9,8%	8,1%	8,2%
4 ^o trim.11	9,3%	10,5%	9,7%	7,9%	8,2%
1 ^o trim.12	8,9%	10,5%	9,5%	7,7%	8,0%
2 ^o trim.12	8,6%	10,5%	9,4%	7,8%	7,7%
3 ^o trim.12	7,9%	10,7%	9,4%	7,8%	7,6%
4 ^o trim.12	30,2%	11,2%	9,7%	7,7%	7,6%
1 ^o trim.13	33,2%	12,6%	11,2%	7,5%	9,0%
2 ^o trim.13	21,5%	16,6%	15,4%	7,5%	13,1%

Os coeficientes de variação estimados, quando se utiliza o modelo GAS(1,1) com distribuição log-normal para prever as reservas IBNR, são inferiores aos demais em 58,8% dos trimestres.

5.5

Teste Diebold e Mariano dentro da amostra - coberturas Casco e RCFV

Nesta tese, foram aplicados um método (*Chain Ladder*) e quatro modelos (Atherino univariado e bivariado, GAS (1,1) com distribuições gama e log-normal) para a estimação das reservas IBNR, utilizando os dados da cobertura Casco e da cobertura RCFV (conforme apresentado nas Tabelas 5.1 e 5.1). Para auxiliar na tomada de decisão a respeito do modelo mais eficiente, o teste Diebold e Mariano foi empregado. O teste em questão tem a finalidade de comparar a precisão das previsões obtidas através de dois modelos distintos (Anexo C). A função perda utilizada foi o MAPE.

A hipótese nula, $H_0 : E[d_{ij}] = 0$, é a de que o modelo i e o modelo j geram previsões de mesma precisão e a hipótese alternativa, $H_1 : E[d_{ij}] < 0$, é a hipótese que o modelo i é mais preciso que o modelo j .

A Tabela 5.42 apresenta os p -valores dos testes aplicados à cobertura Casco. O primeiro modelo citado na linha é considerado o modelos i e o segundo o modelo j .

Tabela 5.42: Cobertura Casco - teste Diebold e Mariano (p -valores).

Modelo (i)	Modelo (j)	p -valores
GAS (1,1) log-normal	<i>Chain Ladder</i>	(0,005)
GAS (1,1) log-normal	Atherino univariado	(0,263)
GAS (1,1) log-normal	Atherino bivariado	(0,234)
GAS (1,1) log-normal	GAS (1,1) gama	(0,056)
GAS (1,1) gama	<i>Chain Ladder</i>	(0,017)
GAS (1,1) gama	Atherino univariado	(0,556)
GAS (1,1) gama	Atherino bivariado	(0,516)
Atherino bivariado	<i>Chain Ladder</i>	(0,017)
Atherino bivariado	Atherino univariado	(0,937)
Atherino univariado	<i>Chain Ladder</i>	(0,014)

Ao nível de significância de 5,0%, todos os modelos são mais precisos, comparados ao método *Chain Ladder*. Nas demais comparações aceita-se a hipótese de que os modelos geram previsões de mesma precisão. Provavelmente porque as funções perdas (MAPE) dos quatro modelos - Atherino univariado, Atherino bivariado, GAS (1,1) log-normal e GAS (1,1) gama - são muito próximas (vide Tabela 5.34).

A Tabela 5.43 descreve os p -valores dos testes aplicados á cobertura RCFV.

Tabela 5.43: Cobertura RCFV - teste Diebold e Mariano (p -valores).

Modelo (i)	Modelo (j)	p -valores
GAS (1,1) log-normal	<i>Chain Ladder</i>	(0,017)
GAS (1,1) log-normal	Atherino univariado	(0,389)
GAS (1,1) log-normal	Atherino bivariado	(0,462)
GAS (1,1) log-normal	GAS (1,1) gama	(0,197)
GAS (1,1) gama	<i>Chain Ladder</i>	(0,024)
GAS (1,1) gama	Atherino univariado	(0,443)
GAS (1,1) gama	Atherino bivariado	(0,514)
Atherino bivariado	<i>Chain Ladder</i>	(0,022)
Atherino bivariado	Atherino univariado	(0,031)
Atherino univariado	<i>Chain Ladder</i>	(0,031)

Os testes aplicados à cobertura RCFV mostram, ao nível de significância de 5,0%, que todos os modelos são mais precisos, comparados ao método *Chain Ladder* e o modelo Atherino bivariado gera previsões melhores, comparado ao Atherino univariado.

Ressalta-se que as estatísticas MAPE (Tabelas 5.36 e 5.34) são muito próximas.

6 Conclusão

Esta tese apresenta duas contribuições para a modelagem estatística do triângulo de *run-off* com o objetivo de estimar as reservas IBNR. A primeira delas é o desenvolvimento de um modelo multivariado em espaço de estado, via formulação SUTSE para estimação conjunta de reservas IBNR associadas a dois ou mais segmentos de uma seguradora. O modelo proposto é capaz de prever as reservas IBNR levando em consideração, quando existente, as relações entre os pagamentos dos sinistros IBNR de dois segmentos distintos, porém, de alguma forma, relacionados.

A outra contribuição diz respeito à formulação de um arcabouço geral para estimação de reservas IBNR a partir de triângulos de *run-off* via modelos GAS. Adicionalmente, foi desenvolvido e aplicado um método para o cálculo dos intervalos de confiança dessas reservas e do efeito linha (previsões fora do triângulo de *run-off*).

Em uma perspectiva mais aplicada, nossa contribuição consiste em apresentar modelos com fundamentação estatística e mais flexíveis (GAS) para prever as reservas IBNR, otimizando o retorno sobre o patrimônio líquido (*Return on Equity* - ROE), sem afetar a solvência da seguradora.

Nas Tabelas 6.1 e 6.2, estão apresentados os critérios que foram utilizados para escolher o melhor modelo para prever reservas IBNR (coberturas Casco e RCFV). As células marcadas com X nas tabelas explicitam que, naquele quesito, o modelo em questão apresentou o melhor resultado comparado aos demais modelos. Ressalta-se que o modelo GAS(1,1), quer seja com distribuição gama, quer seja com distribuição log-normal, apresentou menor dificuldade na programação (dificuldade na elaboração do código no programa R-Project) comparado aos modelos Atherino univariado e bivariado (modelo SUTSE de modo concomitante com o modelo Atherino).

Tabela 6.1: Cobertura Casco - avaliação das métricas aplicadas.

	Métricas/Modelos				
	Chain ladder	Atherino univariado	Atherino bivariado	GAS(1,1) Gama	GAS(1,1) Log-normal
Med. de aderência - dentro da amostra Melhor avaliação: (maior nº de critérios - Tab. 5.34)					X
Med. de aderência - fora da amostra Melhor avaliação: (maior nº de critérios - Tab.5.35)				X	
Tamanho da reserva Melhor avaliação: (Menor reserva - Tab. 5.38)				X	
Coefficiente de Variação Melhor avaliação: (Menores CV e menos variabilidade ao longo dos 18 trim. - Tab. 5.39)					X
Intervalo de confiança Melhor avaliação: (Menor amplitude - Tab. 5.11 e 5.17)	-	-	-	X	
Teste Diebold e Mariano Melhor avaliação: (Precisão - Tab.5.42)		X	X	X	X

Nota:“-” - Não calculado.

Para o triângulo de *run-off* com dados de sinistros da cobertura Casco (Tabela 5.1), como pode ser visto na Tabela 6.1, quatro quesitos apontam para o modelo GAS (1,1) com distribuição gama como sendo o mais aconselhável. São eles: medidas de aderência fora da amostra, tamanho das reservas IBNR estimadas, intervalo de confiança e teste DM. De fato, para esta cobertura em questão, os modelos GAS(1,1), quer seja com distribuição gama, quer seja com distribuição log-normal, apresentaram resultados melhores do que o método *Chain Ladder*, o modelo Atherino e até mesmo melhores na comparação com o modelo Atherino bivariado, também proposto e desenvolvido nesta tese.

Baseado em nossos resultados, pode-se inferir que os modelos Atherino univariado e bivariado, são sem dúvidas, mais aconselháveis para a previsão das reservas IBNR, comparados ao método *Chain Lader*. Contudo, quando incluímos os resultados dos modelos GAS(1,1) na comparação, observamos que a dificuldade na programação para estimar as reservas, via modelo Atherino univariado ou modelo Atherino bivariado, o qual leva em consideração as relações internas e externas dos dois triângulos de *run-off*, não trazem maiores ganhos em termos de precisão.

Tabela 6.2: Cobertura RCFV - avaliação das métricas aplicadas.

	Métricas/Modelos				
	Chain ladder	Atherino univariado	Atherino bivariado	GAS(1,1) Gama	GAS(1,1) Log-normal
Med. de aderência - dentro da amostra Melhor avaliação: (maior nº de critérios - Tab. 5.36)		X			
Med. de aderência - fora da amostra Melhor avaliação: (maior nº de critérios - Tab.5.37)	*	*	*	*	*
Tamanho da reserva Melhor avaliação: (Menor reserva - Tab. 5.40)		X			
Coefficiente de Variação Melhor avaliação: (Menores CV e menos variabilidade ao longo dos 18 trim. - Tab. 5.41)					X
Intervalo de confiança Melhor avaliação: (Menor amplitude - Tab. 5.25 e 5.31)	-	-	-		X
Teste Diebold e Mariano Melhor avaliação: (Precisão - Tab.5.43)		X	X	X	X

Nota: "*" - Não conclusivo e "-" - Não calculado.

Para o triângulo de *run-off* com dados de sinistros da cobertura RCFV, Tabela 5.2, a disputa está entre o modelo Atherino univariado e o modelo GAS(1,1) com distribuição log-normal (Tabela 6.2). As medidas de aderência dentro da amostra e o tamanho das reservas estimadas apontam para o modelo Atherino univariado, ao passo que os coeficientes de variação e o intervalo de confiança indicam o modelo GAS (1,1) com distribuição log-normal.

O resultado encontrado nesta tese corrobora aqueles obtidos por Koopman et al. (2015). Segundo os autores, as previsões obtidas através dos modelos GAS e dos modelos em espaço de estado têm precisão semelhante. Na conclusão do artigo em questão, os autores salientaram que, para as séries utilizadas, os resultados dos modelos GAS são comparáveis aos resultados dos modelos dirigidos por parâmetro, contudo, a complexidade computacional é extremamente menor nos modelos GAS.

Desta pesquisa podem-se originar três possíveis extensões: (I) Uma possível extensão é utilizar os dois modelos propostos nesta tese (Atherino multivariado e GAS) para modelar o volume de sinistros IBNR (Martin (2015), por exemplo, utiliza um modelo log-linear com efeito linha e efeito coluna para calcular o volume de sinistros IBNR). No caso dos modelos GAS, distribuições adequadas aos dados de contagem em situações que a aproximação pela distribuição normal não seja adequada (binomial negativa e Poisson) por exemplo, podem ser utilizadas. (II) Outra possível extensão deste estudo é o cálculo do efeitos linha (previsões fora do triângulo de *run-off*, vide Subseção 4.2.4), tanto para o modelo Atherino univariado quanto para o modelo Atherino bivariado. (III) A última extensão proposta nesta tese refere-se à elaboração e à aplicação de um modelo GAS bivariado (utilizando

as distribuições gama ou a log normal bivariadas, por exemplo). Neste caso ao invés de um parâmetro variante no tempo f_t , teremos um vetor de parâmetros. Nesta possível extensão a dependência entre as coberturas poderão ser captadas.

A

Abordagem univariada para caso multivariado

Lima (2010) observa que a principal motivação que sugere o uso do tratamento univariado para séries multivariadas, é a eficiência computacional, pois é bem mais simples trabalhar com valores escalares do que com matrizes.

[...] O exemplo mais ilustrativo deste ganho fica aparente considerando um processo de otimização com inicialização difusa, pois neste caso, é necessário “zerar” duas matrizes para poder estabilizar o processo de otimização. Quando adota-se o tratamento univariado do modelo multivariado uma dessas matrizes torna-se um escalar, sendo bem mais fácil obter um zero nesta condição do que fazendo uma matriz de zeros (LIMA, 2010, P. 36).

A abordagem será utilizada na programação, porém, a forma exposta neste trabalho será a forma espaço de estado apresentada nas Equações 3-41 e 3-45.

Assumindo primeiramente que a matriz H_t é diagonal, têm-se as observações e as inovações, sendo:

$$x_t = \begin{pmatrix} x_{t,1} \\ \vdots \\ x_{t,p_t} \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{t,1} \\ \vdots \\ \varepsilon_{t,p_t} \end{pmatrix} \quad (\text{A-1})$$

Com as matrizes do sistema:

$$Z_t = \begin{pmatrix} Z_{t,1} \\ \vdots \\ Z_{t,p_t} \end{pmatrix}, \quad H_t = \begin{pmatrix} \sigma_{t,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{t,p_t}^2 \end{pmatrix} \quad (\text{A-2})$$

Sendo $x_{t,i}$, $\varepsilon_{t,i}$ e $\sigma_{t,i}^2$ escalares e $Z_{t,i}$ vetores ($1 \times m$), para $i = 1, \dots, p_t$.

De acordo com Lima (2010), o propósito desta abordagem é reconhecer o processo $x_t = (x_{t,1}, \dots, x_{t,p_t})$ descrito no tempo por:

$$\begin{pmatrix} x_{1,1} \\ x_{1,2} \\ x_{1,3} \\ \vdots \\ x_{1,p_1} \end{pmatrix}, \begin{pmatrix} x_{2,1} \\ x_{2,2} \\ x_{2,3} \\ \vdots \\ x_{2,p_2} \end{pmatrix}, \begin{pmatrix} x_{3,1} \\ x_{3,2} \\ x_{3,3} \\ \vdots \\ x_{3,p_3} \end{pmatrix}, \dots, \begin{pmatrix} x_{T,1} \\ x_{T,2} \\ x_{T,3} \\ \vdots \\ x_{T,p_t} \end{pmatrix} \quad (\text{A-3})$$

Para a formação do vetor x_t , as coordenadas $x_{t,i}$, $i = 1, \dots, p_t$ chegam uma de cada vez, ou seja, primeiramente as observações no tempo t de todas as p_t séries, após, todas as observações em $t+1$ de todas as p_t séries, e assim sucessivamente, até o tempo $t = T$. Sabendo que a última observação no tempo t da série i é p_t . Logo, tem-se uma série univariada como descrito em A-4.

$$x_{1,1}, x_{1,2}, \dots, x_{1,p_1}, x_{2,1}, x_{2,2}, \dots, x_{2,p_2}, x_{3,1}, x_{3,2}, \dots, x_{3,p_3}, \dots, x_{n,1}, x_{n,2}, \dots, x_{T,p_T} \quad (\text{A-4})$$

Segundo Lima (2010), o propósito é achar um modelo univariado equivalente para A-3 e assim as recursões de Kalman tornam-se aplicáveis. Uma vez que A-3 é reescrita como A-4 a dinâmica do modelo multivariado deve ser estritamente respeitada, de forma que A-3 e A-4 sejam probabilisticamente equivalentes.

A equação das observações na forma univariada é dada por:

$$x_{t,i} = Z_{t,i}\alpha_{t,i} + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \sim WN(0, \sigma_{t,i}^2) \quad (\text{A-5})$$

Sendo $t = 1, \dots, T$, $i = 1, \dots, p_t$ e $\alpha_{t,i} = \alpha_t$.

O vetor de estado é dado pelas Equações A-6 e A-7:

$$\alpha_{t,i+1} = \alpha_{t,i} \quad i = 1, \dots, p_{t-1} \quad (\text{A-6})$$

$$\alpha_{t+1,1} = T_t\alpha_{t,p_t} + R_t\eta_t \quad i = 1, \dots, T \quad (\text{A-7})$$

Com o vetor de estado $\alpha_{1,1} = \alpha_1 \sim N(a, P)$ e $\eta_{t,p_t} \sim WN(0, Q_t)$.

A.1

Matriz H_t não diagonal - Modelo Atherino bivariado

Quando H_t é não diagonal, a representação univariada do modelo multivariado não leva a um modelo equivalente, pois as correlações entre as séries contidas na equação das observações são perdidas. Neste caso, segundo Koopman e Durbin (1998), pode-se introduzir o vetor ε_t no vetor de estado. Sendo assim:

$$\bar{\alpha}_t = \begin{pmatrix} \alpha_t \\ \varepsilon_t \end{pmatrix} \quad \bar{Z}_t = \begin{pmatrix} Z_t & I_m \end{pmatrix} \quad (\text{A-8})$$

e para a equação de estado define-se:

$$\bar{\eta}_t = \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \quad \bar{T}_t = \begin{pmatrix} T_t & 0 \\ 0 & 0 \end{pmatrix} \quad (\text{A-9})$$

$$\bar{R}_t = \begin{pmatrix} R_t & 0 \\ 0 & I_m \end{pmatrix} \quad \bar{Q}_t = \begin{pmatrix} Q_t & 0 \\ 0 & H_t \end{pmatrix} \quad (\text{A-10})$$

levando a:

$$x_t = \bar{Z}_t \bar{\alpha}_t \quad \bar{\alpha}_{t+1} = \bar{T}_t \bar{\alpha}_t + \bar{R}_t \bar{\eta}_t \quad \bar{\eta}_t \sim N(0, \bar{Q}_t) \quad (\text{A-11})$$

Para $t = 1, \dots, T$. I_m é a matriz identidade e m é a dimensão da matriz H_t . Segundo Koopman e Durbin (1998), procede-se com a mesma estratégia para o caso onde H_t é diagonal, tratando cada elemento do vetor de observação individualmente.

A.2

Filtragem e suavização - MEE multivariado na forma univariada

Assim como definido em Koopman e Durbin (2000):

$$\begin{aligned} a_{t,1} &= E(\alpha_{t,1} | X_{t-1}) \\ a_{t,i} &= E(\alpha_{t,i} | X_{t-1}, x_{t,1}, \dots, x_{t,i-1}) \\ P_{t,1} &= \text{Var}(\alpha_{t,1} | X_{t-1}) \\ P_{t,i} &= \text{Var}(\alpha_{t,i} | X_{t-1}, x_{t,1}, \dots, x_{t,i-1}) \end{aligned} \quad (\text{A-12})$$

Para $i = 2, \dots, p_t$. As esperanças e variâncias do vetor de estado explicitadas no conjunto de Equações A-12 mostram que no tempo t o vetor de estado i é condicionado a todo o passado das outras séries e ao passado da própria série i .

As equações de filtragem na qual H_t é diagonal são:

$$a_{t,i+1} = a_{t,i} + K_{t,i} v_{t,i} \quad (\text{A-13})$$

$$P_{t,i+1} = P_{t,i} - K_{t,i} F_{t,i}^{-1} K_{t,i}' \quad (\text{A-14})$$

No qual:

$$v_{t,i} = x_{t,i} - Z_{t,i} a_{t,i}, \quad F_{t,i} = Z_{t,i} P_{t,i} Z_{t,i}' + \sigma_{t,i}^2 \text{ e } K_{t,i} = P_{t,i} Z_{t,i}' F_{t,i}^{-1} \quad (\text{A-15})$$

Para $i = 1, \dots, p_t$ e $t = 1, \dots, n$. Sabe-se que $v_{t,i}$ e $F_{t,i}$ são escalares e $K_{t,i}$ é um vetor coluna. De acordo com Koopman e Durbin (2000), a transição do tempo t para o tempo $t + 1$ é dada pelas Equações A-16 e A-17:

$$a_{t+1,1} = T_t a_{t,p_t+1} \quad (\text{A-16})$$

$$P_{t+1,1} = T_t P_{t,p_t+1} T_t' + R_t Q_t R_t' \quad (\text{A-17})$$

As equações de suavização utilizadas nesta tese foram deduzidas em Koopman e Durbin (2000).

A abordagem da suavização univariada evita duas multiplicações de matrizes e a implementação se torna mais simples, segundo Koopman e Durbin (2000).

A.3

Missing values - MEE multivariado na forma univariada

Supondo que $I_{t,i} = (t_{1,i}, t_{2,i}, \dots, t_{T,i})$ corresponda ao conjunto de índices em que há informações faltantes. Tem-se, no filtro de Kalman:

$$a_{t|t,i} = E[\alpha_{t,i}|x_{t,i}] = E[\alpha_{t,i}|x_{t-1,i}] = a_{t|t-1,i} \quad (\text{A-18})$$

$$P_{t|t,i} = Var[\alpha_{t,i}|x_{t,i}] = Var[\alpha_{t,i}|x_{t-1,i}] = P_{t|t-1,i} \quad (\text{A-19})$$

Observa-se que o problema está no cômputo, assim como na abordagem tradicional multivariada e univariada, da $E[\cdot|X_{t,i}]$ e da $V[\cdot|X_{t,i}]$ quando $X_{t,i}$ não existir.

Para o instante de tempo em que houver observações faltantes, as equações de suavização utilizadas nesta tese também estão detalhadas em Koopman e Durbin (2000).

A.4

Inicialização difusa exata - MEE multivariado na forma univariada

Sendo d o instante de estabilidade do processo composto pelas equações de suavização, até d (t,i) são descritos da seguinte forma:

$$(t, i) = (t^*, i^*), (t^*, i^* - 1), \dots, (t^*, 1), (t^* - 1, p_{t^*-1}), \dots, (1, 1) \quad (\text{A-20})$$

Assim, como na abordagem univariada, o suavizador de Kalman inicial exato consiste em uma solução baseada na escrita das equações de suavização como funções

suaves de k , aproximadas por fórmulas de Maclaurin e fazendo $k \rightarrow +\infty$. As equações utilizadas foram:

$$\begin{aligned}
 P_t &= P_{*,t,i} + kP_{\infty,t,i} \\
 r_{t,i} &= r_{t,i}^{(0)} + k^{-1}r_{t,i}^{(1)} + O(k^{-2}) \\
 N_{t,i} &= N_{t,i}^{(0)} + k^{-1}N_{t,i}^{(1)} + k^{-2}N_{t,i}^{(2)} + O(k^{-3}) \\
 F_{t,i} &= F_{*,t,i} + kF_{\infty,t,i} \\
 F_{t,i}^{-1} &= k^{-1}F_{\infty,t,i}^{-1} - k^{-2}F_{*,t,i}F_{\infty,t,i}^{-2} + O(k^{-3}) \\
 F_{t,i}^{-1}(F_{*,t,i} + kF_{\infty,t,i}) &= 1
 \end{aligned} \tag{A-21}$$

Sendo $P_{t,i}^\dagger = (P_{*,t,i}, P_{\infty,t,i})$ e $L_{t,i} = I - K_{t,i}F_{t,i}^{-1}Z_{t,i}$.

Como pode ser observado o suavizador inicial exato pode ser reescrito na forma matricial. Sendo:

$$r_{t,i}^\dagger = \begin{pmatrix} r_{t,i}^{(0)} \\ r_{t,i}^{(1)} \end{pmatrix} \tag{A-22}$$

$$N_{t,i}^\dagger = \begin{pmatrix} N_{t,i}^{(0)} & N_{t,i}^{(1)} \\ N_{t,i}^{(1)} & N_{t,i}^{(2)} \end{pmatrix} \tag{A-23}$$

$$L_{t,i-1}^\dagger = \begin{pmatrix} L_{\infty,t,i} & L_{0,t,i} \\ 0 & L_{\infty,t,i} \end{pmatrix} \tag{A-24}$$

No qual:

$$\begin{aligned}
 L_{\infty,t,i} &= I - K_{\infty,t,i}Z_{t,i}F_{\infty,t,i}^{-1} \\
 L_{0,t,i} &= (K_{\infty,t,i}F_{*,t,i}F_{\infty,t,i}^{-1} - K_{*,t,i})Z_{t,i}F_{\infty,t,i}^{-1}
 \end{aligned}$$

Como descrito em Koopman e Durbin (2000) e Durbin e Koopman (2001), usando as equações de suavização e fazendo $F_{\infty,t,i} > 0$, tem-se:

$$r_{t,i-1}^\dagger = \begin{pmatrix} 0 \\ Z'_{t,i}F_{t,i}^{(1)}v_{t,i}^{(0)} \end{pmatrix} + L'_{t,i}r_{t,i}^\dagger \tag{A-25}$$

$$N_{t,i-1}^\dagger = \begin{pmatrix} 0 & Z'_{t,i}F_{\infty,t,i}^{-1}Z_{t,i} \\ Z'_{t,i}F_{\infty,t,i}^{-1}Z_{t,i} & Z'_{t,i}F_{\infty,t,i}^{-2}Z_{t,i}F_{*,t,i} \end{pmatrix} + L'_{t,i}N_{t,i}^\dagger L_{t,i}^\dagger \tag{A-26}$$

Tem-se ainda:

$$r_{t-1, P_t}^\dagger = \begin{pmatrix} T_{t-1} & 0 \\ 0 & T_{t-1} \end{pmatrix}' r_{t,0}^\dagger \quad (\text{A-27})$$

$$N_{t-1, P_t}^\dagger = \begin{pmatrix} T_{t-1} & 0 \\ 0 & T_{t-1} \end{pmatrix}' N_{t,0}^\dagger \begin{pmatrix} T_{t-1} & 0 \\ 0 & T_{t-1} \end{pmatrix} \quad (\text{A-28})$$

Para $t = t^*, \dots, 1$. Têm-se as equações suavizadas difusas dadas por:

$$\hat{\alpha}_t = a_{t,1} + P_{t,1}^\dagger r_{t,0}^\dagger \quad (\text{A-29})$$

$$V_t = P_{*,t,1} - P_{t,1}^\dagger N_{t,0}^\dagger P_{t,1}' \quad (\text{A-30})$$

Para mais detalhes sobre inicialização difusa exata do suavizador e do filtro de Kalman pesquisar no artigo: *Fast Filtering and smoothing for multivariate state space models*, publicado por Koopman e Durbin em 2000 e no livro dos mesmos autores *Time Series Analysis by Space State Methods*, publicado no ano de 2001.

A.5

Verossimilhança exata - MEE multivariado na forma univariada

Para o caso multivariado na forma univariada, segundo Koopman e Durbin (2000), os hiperparâmetros serão estimados utilizando a decomposição da verossimilhança pelo erro de predição, como pode ser visto nas equações abaixo:

$$\log(L(\psi)) = -\frac{Tm}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^{p_t} \left(\log F_{t,i} + v_{t,i}^2 F_{t,i}^{-1} \right) \quad (\text{A-31})$$

No qual, $v_{t,i}$ e $F_{t,i}^{-1}$ foram definidos na Seção A.2, p_t é o número de series temporais empilhadas, m número de variáveis não observadas e T representa o número de observações de cada série.

A verossimilhança pelo erro de predição com inicialização difusa é dada por:

$$\begin{aligned} \log(L(\psi)) &= \text{constante} - \frac{1}{2} \sum_{t=1}^{t^*} \sum_{i=1}^{i^*} \log F_{\infty,t,i} \\ &\quad - \frac{1}{2} \sum_{t=t^*}^T \sum_{i=i^*+1}^{p_t} \left(\log F_{t,i} + v_{t,i}^2 F_{t,i}^{-1} \right) \end{aligned}$$

No qual:

$$F_{\infty,t,i} = Z_{t,i} P_{\infty,t,i} Z'_{t,i} \quad (\text{A-32})$$

$$K_{\infty,t,i} = P_{\infty,t,i} Z'_{t,i} \quad (\text{A-33})$$

As variáveis $F_{\infty,t,i}$, $K_{\infty,t,i}$ e $P_{\infty,t,i}$ da filtragem difusa não dependem das matrizes H_t , R_t e Q_t . Portanto, os cálculos de $F_{\infty,t,i}$, $K_{\infty,t,i}$ e $P_{\infty,t,i}$ não se repetem cada vez que é necessária uma nova avaliação de probabilidade para um novo vetor de parâmetros ψ . Isso leva a esforços computacionais consideravelmente menores.

Visto que o algoritmo EM sempre vai na direção do máximo da função, como em Atherino et al. (2010), utilizou-se o mesmo para gerar as componentes do vetor e inicializar a otimização.

A.6 Algoritmo EM

A utilização do algoritmo EM garante que na maximização a função de verossimilhança cresça sempre a cada iteração, diminuindo assim o tempo computacional para se atingir o máximo da função, por este motivo utilizamos o EM para inicializar a estimação.

Segundo McLachlan, Krishnan e Ng (2004), o EM é um algoritmo iterativo, no qual, em cada iteração há duas etapas: a *Expectation*, (E-passo) e a *Maximization*, (M-passo). O nome do algoritmo EM foi elaborado por Dempster et al. (1977), que sintetizaram formulações desse algoritmo em muitos casos e apresentaram uma formulação geral para este método.

Assim como na revisão do livro de Durbin e Koopman (2001), publicado em 2014, descreve-se os passos do algoritmo EM para um modelo univariado contendo apenas nível. Essa é a elaboração mais simples do algoritmo, contudo, se pode adaptar para representações com sazonalidade ou periodicidade e até mesmo modelos multivariados com todas estas componentes.

O logaritmo da função de verossimilhança pode ser escrita como:

$$\text{Log}p(\alpha, X|\psi) = -\frac{1}{2} \sum_{t=1}^n \text{Log}|H_t| + \text{Log}|Q_{t-1}| + \varepsilon'_t H_t^{-1} \varepsilon_t + \eta'_{t-1} Q_{t-1}^{-1} \eta_{t-1} \quad (\text{A-34})$$

Sabendo que $\hat{\varepsilon}_t = E[x_t|\alpha_t]$ e $\hat{\eta}_t = E[\alpha_t|\alpha_{t-1}]$. A partir deste ponto é necessário avaliar a função de verossimilhança em relação à densidade $p(\alpha, x|\psi)$.

$$\begin{aligned} E[\text{Log}p(\alpha, X|\psi)] &= -\frac{1}{2} \sum_{t=1}^n \text{Log}|H_t| \\ &+ \text{Log}|Q_{t-1}| + \det[(\hat{\varepsilon}_t \hat{\varepsilon}'_t + \text{var}(\varepsilon_t|X)) H_t^{-1}] \\ &+ \det[(\hat{\eta}_{t-1} \hat{\eta}'_{t-1} + \text{var}(\eta_{t-1}|X)) Q_{t-1}^{-1}] \end{aligned} \quad (\text{A-35})$$

O próximo passo é maximizar a Função A-35, aplicando a derivada.

$$\begin{aligned} \frac{\partial}{\partial \psi} \left[E[\text{Log}p(\alpha, X | \psi)] \right] &= -\frac{1}{2} \frac{\partial}{\partial \psi} \sum_{t=1}^n \text{Log}|H_t| \\ &+ \text{Log}|Q_{t-1}| + \det[(\hat{\varepsilon}_t \hat{\varepsilon}_t' + \text{var}(\varepsilon_t | X)) H_t^{-1}] \\ &+ \det[(\hat{\eta}_{t-1} \hat{\eta}_{t-1}' + \text{var}(\eta_{t-1} | X)) Q_{t-1}^{-1}] \end{aligned}$$

Os valores para as próximas iterações são:

$$\begin{aligned} H_t^{\text{new}} &= n^{-1} [(\varepsilon_t^{\text{old}} \varepsilon_t^{\text{old}'} + \text{var}(\varepsilon_t^{\text{old}} | X))] \\ Q_t^{\text{new}} &= n^{-1} [(\eta_{t-1}^{\text{old}} \eta_{t-1}^{\text{old}'} + \text{var}(\eta_{t-1}^{\text{old}} | X))] \end{aligned} \quad (\text{A-36})$$

A.7

MEE - diagnósticos e medidas de aderência

Um passo crucial, após aplicar qualquer modelo estatístico a uma base de dados, é verificar se o modelo aplicado é uma aproximação razoável para o verdadeiro mecanismo probabilístico gerador dos dados. Segundo Pizzinga (2004), a essa tarefa dá-se o nome de diagnósticos, os quais se caracterizam basicamente pela análise gráfica e/ou inferencial de resíduos, pela aplicação de testes estatísticos e pelo poder preditivo do modelo ajustado, isto é, se as expressões matemáticas estimadas conseguirem “reproduzir” relativamente bem os dados.

A.8

Análise dos resíduos

Quanto aos diagnósticos, no contexto de espaço de estado, podem ser feitos sobre as inovações padronizadas, $v_t^{(P)}$, as quais são calculadas, utilizando a Equação A-37 ou sobre os resíduos auxiliares que são calculadas usando a Equação A-39.

Inovações padronizadas

$$v_t^{(P)} = F_t^{-\frac{1}{2}} v_t \quad (\text{A-37})$$

No qual $F_t^{-1} = F_t^{-\frac{1}{2}} F_t^{-\frac{1}{2}}$.

Resíduos auxiliares

De acordo com a Subseção 4.8 do livro de Durbin e Koopman (2001), para as posições que forem *missing*, tem-se:

$$\hat{\varepsilon}_{t|n} \sim N(0, H_t)$$

Para as posições em que os dados não são ausentes, $\mathfrak{I} \equiv \{t: \text{é não ausente}\}$, tem-se:

$$\begin{aligned}\hat{\varepsilon}_{t|n} &= E[\varepsilon_t|X] = E[\varepsilon_t|X_{t-1}, v_t, \dots, v_n] \\ &= \sum_{j=t}^n E(\varepsilon_t, v'_j) F_j^{-1} v_j\end{aligned}\tag{A-38}$$

Tem-se que: $E[\varepsilon_t|Y_{t-1}] = 0$, $E(\varepsilon_t, v'_j) = E(\varepsilon_t, u'_j)Z'_j + E[\varepsilon_j \varepsilon'_j]$ com $E[\varepsilon_t u'_j] = 0$ caso $t = 1, \dots, n$ e $j = 1, \dots, n$. Ou seja:

$$E(\varepsilon_t, v'_j) = \begin{cases} H_t, & j = t; \\ E(\varepsilon_t, u'_j)Z'_j, & j = t + 1, \dots, n. \end{cases}\tag{A-39}$$

B

Teste de homogeneidade da variância

De acordo com Durbin e Koopman (2001), a seguinte estatística de teste pode ser utilizada para testar homogeneidade da variância de uma série de resíduos estacionários.

$$H(h) = \frac{\sum_{t=n-h+1}^n e_t^2}{\sum_{t=1}^h e_t^2} \quad (\text{B-1})$$

No qual $e_t = x_t - E[x_t|X_{t-1}]$.

O teste busca selecionar dois segmentos de um vetor de inovações padronizadas - no caso dos modelos Atherino univariado e bivariado - de um vetor com resíduos quantílicos e de Pearson - no caso dos modelos GAS - e verificar se ambos possuem a mesma variância. Como se observa na Equação B-1, temos a soma de h variáveis normais padrão quadráticas no numerador e no denominador. Com isso, fica claro que se trata da divisão de duas variáveis com distribuições qui-quadrado com mesmo grau de liberdade, h . Como consequência, tem-se uma distribuição $F(h, h)$.

Por definição, o teste tem como hipótese nula que a série é homocedástica, isto é, possui variância constante, e sob esta hipótese $H(h) \sim F(h, h)$.

Para executar o teste divide-se a série de inovações padronizadas em três segmentos: início, meio e fim. São testados os segmentos de inovações do início contra os demais. Após aplicar tal procedimento, pode-se decidir se há rejeição de H_0 , ou seja, se há evidências da série ser heterocedástica. Maiores detalhes sobre o teste de homogeneidade da variância, ver Koenker (1981) ou Durbin e Koopman (2001).

C

Teste Diebold e Mariano

No teste de Diebold e Mariano, a hipótese nula afirma não existir diferença entre duas previsões em comparação. A abordagem deste teste, de acordo com Diebold e Mariano (1995), é diretamente focada nos desempenhos das previsões. O teste pode ser usado para uma ampla classe de medidas de previsão. Diebold (2012) salienta que o teste baseia-se em suposições feitas diretamente no erro de previsão denotado como perda associada. Ressalta-se que a função de perda não precisa ser de segundo grau nem simétrica e os erros de previsão podem ser não Gaussianos e correlacionados.

Segundo Diebold e Mariano (1995), considera-se duas previsões para a série temporal $\{Y_t\}_{t=1}^T$, sendo elas $\{\hat{Y}_{it}\}_{t=1}^T$ e $\{\hat{Y}_{jt}\}_{t=1}^T$, com erros associados $\{e_{it}\}_{t=1}^T$ e $\{e_{jt}\}_{t=1}^T$. Assim, a perda - digamos i - será uma função arbitrária da realização e previsão $g(Y_t, \hat{Y}_{it})$, o que pode ser chamado de $g(e_{it})$. Nesta tese a função perda utilizada foi o MAPE (Equação 3-52).

O teste DM assume as seguintes premissas:

$$\text{Teste DM} \begin{cases} E[d_{ij}] = \mu, & \forall t \text{ e } i \neq j. \\ \text{cov}(d_{ij,(t)}, d_{ij,(t+\tau)}) = \gamma_\tau \\ 0 \leq \text{var}[d_{ij}] = \sigma_{d_{ij}}^2 < \infty, \end{cases} \quad (\text{C-1})$$

Ou seja, a diferença entre dois erros de previsão é um processo estacionário de 2ª ordem.

Sendo d_{ij} a diferença entre duas previsões no mesmo período de tempo.

$$d_{ij} = g(e_{it}) - g(e_{jt}) \quad t = 1, \dots, T \quad (\text{C-2})$$

A hipótese nula de igualdade de desempenho de previsão esperada é dada por:

$$\begin{aligned} E[g(e_{it})] &= E[g(e_{jt})] \\ E[g(e_{it}) - g(e_{jt})] &= 0 \\ E[d_{ij}] &= 0 \quad \implies \mu = 0 \quad (\text{C-1}) \end{aligned} \quad (\text{C-3})$$

Logo, em H_0 aceita-se a hipótese de que a previsão dos modelos testados é igual. H_1 pode assumir 3 alternativas a serem escolhidas:

$$H_1 = \begin{cases} E[d_{ij}] \neq 0, & \text{a acurácia dos modelos é diferente ou} \\ E[d_{ij}] < 0, & \text{o modelo i é mais preciso que o modelo j ou} \\ E[d_{ij}] > 0, & \text{o modelo j é mais preciso que o modelo i.} \end{cases} \quad (\text{C-4})$$

Tem-se que:

$$DN_{ij} = \frac{\bar{d}_{ij}}{\hat{\sigma}_{\bar{d}_{ij}}} \longrightarrow N(0, 1) \quad (\text{C-5})$$

No qual \bar{d}_{ij} é a média amostral dada pela Equação C-6:

$$\bar{d}_{ij} = T^{-1} \sum_{t=1}^T d_{ij}(t) \quad (\text{C-6})$$

E $\sigma_{\bar{d}_{ij}}$, encontrado na Equação C-5 é a estimativa consistente do desvio padrão de \bar{d}_{ij} . Segundo Wink Junior (2011), a variância de \bar{d}_{ij} pode ser escrita da seguinte forma:

$$\sigma_{\bar{d}_{ij}}^2 = T^{-1} [\gamma_0 + 2 \sum_{k=1}^{T-1} \hat{\gamma}_k] \quad (\text{C-7})$$

No qual, $\hat{\gamma}_k$ é a k-ésima auto-covariância de d_{ij} . A distribuição empírica do teste do teste DM é normal.

Logo, a hipótese nula será rejeitada ao nível de 5% de significância se o *p-valor* do teste for menor que 5%.

Vale ressaltar que nesta tese o horizonte de previsão é igual a 1, ou seja, $h = 1$, que caracteriza a previsão *in-sample*.

D

Resultado do modelo Atherino univariado

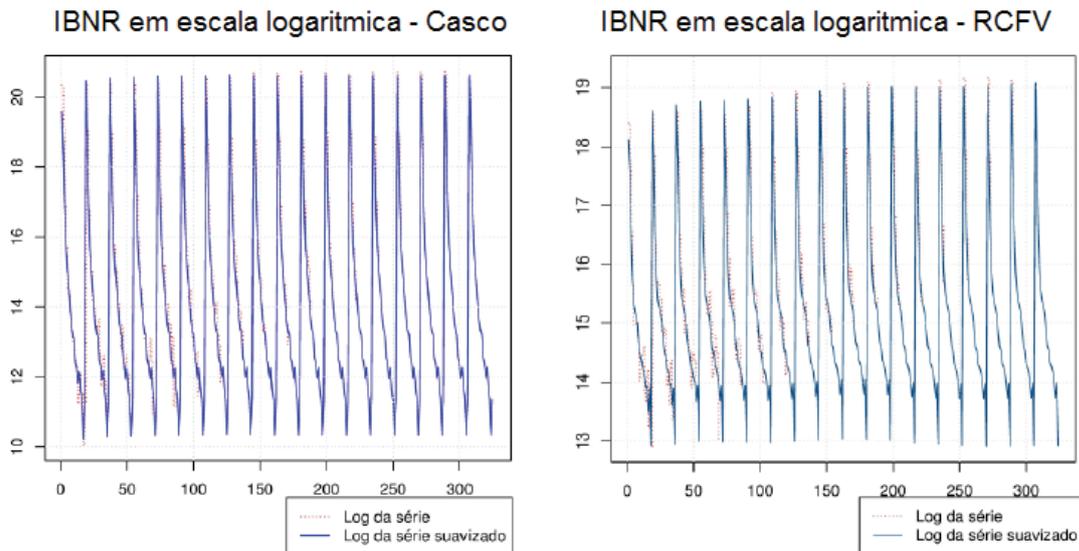
Após a estimação via algoritmo EM e máxima verossimilhança, utilizando a inicialização difusa exata, obtêm-se as seguintes estimativas para as variâncias dos choques aleatórios (Equações 3-1, 3-2 e 3-3) associadas aos modelos estruturais ajustados as duas séries dos triângulos de *run-off* (Tabelas 5.1 e 5.2).

Tabela D.1: Parâmetros estimados e relações sinal/ruído para as séries Casco e RCFV - modelo Atherino univariado.

Parâmetros estimados	Casco	RCFV
σ_ε^2	$8,52 \times 10^{-02}$	$5,51 \times 10^{-02}$
σ_η^2	$1,12 \times 10^{-04}$	$1,84 \times 10^{-04}$
σ_ω^2	$8,06 \times 10^{-05}$	$8,16 \times 10^{-04}$
$q_\eta = \sigma_\eta^2 / \sigma_\varepsilon^2$	$1,31 \times 10^{-03}$	$3,35 \times 10^{-03}$
$q_\omega = \sigma_\omega^2 / \sigma_\varepsilon^2$	$9,47 \times 10^{-04}$	$1,48 \times 10^{-02}$

As estimações dos parâmetros, tanto para a série Casco quanto para a série RCFV, revelaram indícios de nível e periodicidade determinísticos.

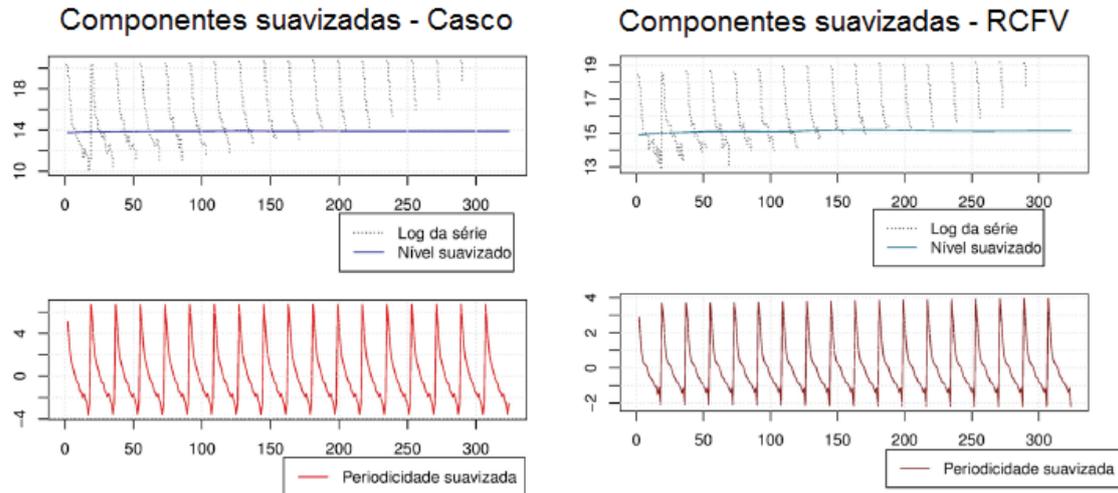
Figura D.1: O logaritmo da série e a série suavizada - coberturas Casco e RCFV.



Como pode ser observado, as séries obtidas através da abordagem de suavização aderem-se de forma adequada às séries originais observadas.

Na Figura D.2, apresentam-se as componentes suavizadas, μ_t e γ_t , provenientes da Equação 3-1 para ambas as séries.

Figura D.2: O logaritmo da série IBNR, o nível e a periodicidade suavizadas - coberturas Casco e RCFV.



Observa-se que no caso da cobertura Casco, a componente nível estimada, μ_t , é praticamente constante, em torno do valor 13 e a componente periodicidade estimada, γ_t , varia de -4 a 5. Para o logaritmo da série RCFV, a componente nível estimada é também constante e um pouco mais elevada, em torno do valor 14 enquanto a componente periódica estimada varia em torno de -2 a 4.

Como pode ser observado na Figura D.2, a amplitude do fator periódico da série Casco é relativamente maior, comparado ao fator periódico da série RCFV. Esse resultado já era esperado, uma vez que, observando o Gráfico 5.1, evidencia-se que a diferença entre o valor pago no primeiro trimestre e no último trimestre de desenvolvimento (com dados disponíveis), em termos monetários, é muito maior na cobertura Casco, comparada à diferença entre os valores na cobertura RCFV (Figura 5.2).

Para a análise residual do modelo Atherino univariado foram calculadas duas séries de resíduos, as inovações padronizadas (Equação A-37) e os resíduos suavizados ou auxiliares (Equação A-39). Esses encontram-se descritos nas Figuras D.2 e D.3.

Tabela D.2: Cobertura Casco - inovações padronizadas e resíduos auxiliares (Atherino univariado).

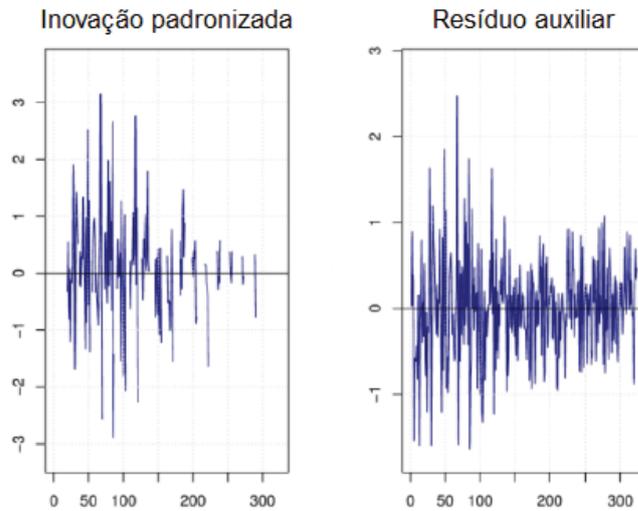
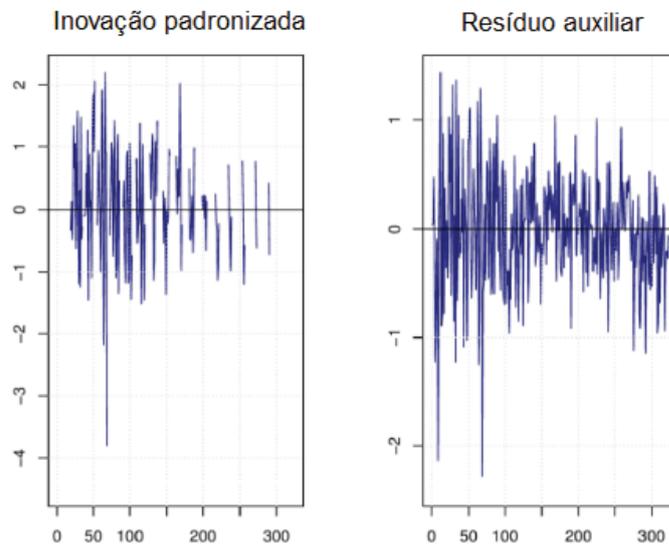


Tabela D.3: Cobertura RCFV - inovações padronizadas e resíduos auxiliares (Atherino univariado).



As Figuras D.3, D.4 e D.5 apresentam as funções de autocorrelação, autocorrelação parcial e os Q-Q plotes dos resíduos das duas séries (Casco e RCFV) que foram modeladas através do modelo Atherino univariado.

Figura D.3: Funções de autocorrelação das inovações padronizadas e das inovações padronizadas ao quadrado - Casco.

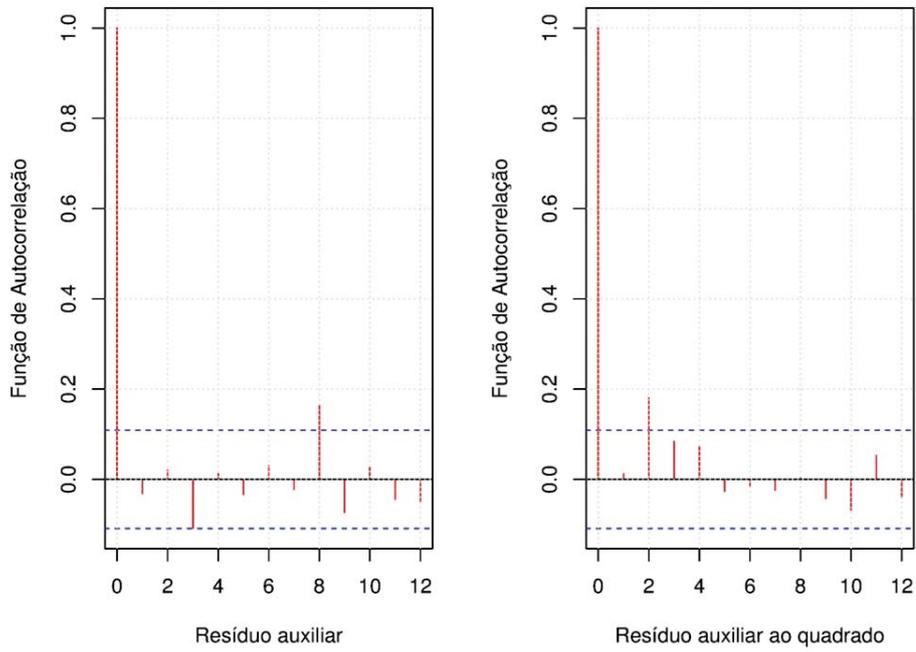


Figura D.4: Funções de autocorrelação das inovações padronizadas e das inovações padronizadas ao quadrado - RCFV.

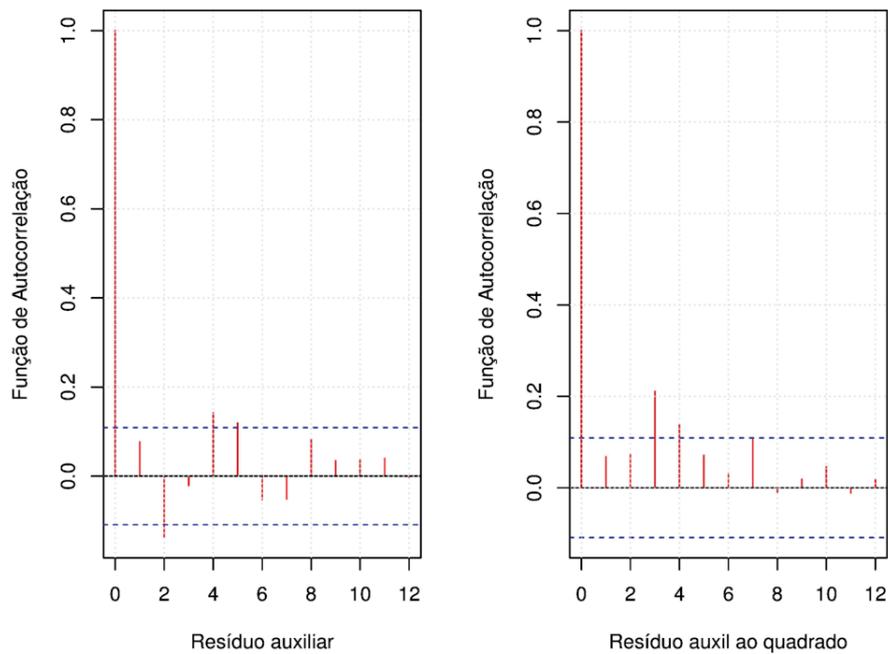
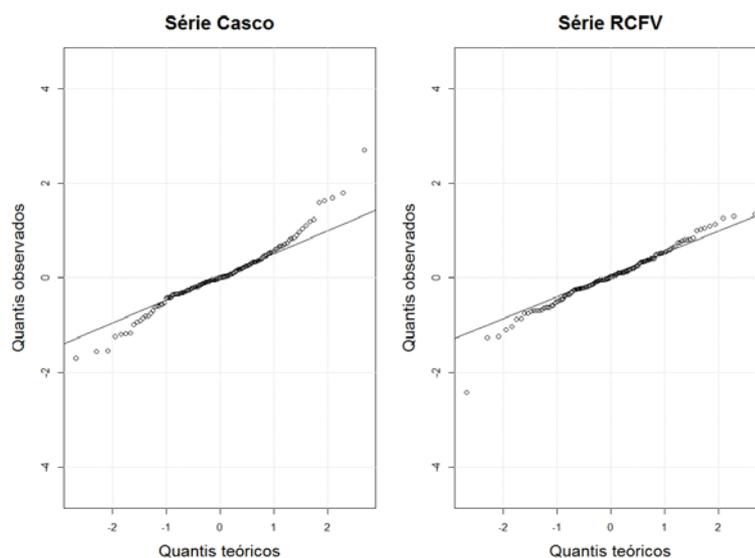


Figura D.5: Coberturas Casco e RCFV - Q-Q plots das inovações padronizadas.



Em vista do padrão da FAC, observa-se que não há evidências fortes de dependência linear nos resíduos. A FAC dos resíduos ao quadrado indica que existe, potencialmente, alguma estrutura de dependência não linear que não foi capturada pelo modelo proposto, mas sendo desprezível. Essa dependência não será tratada nesta tese.

Como pode ser observado nos Q-Q plots, os pontos se ajustam bem as retas apenas nas partes centrais dos gráficos, tanto nos resíduos de Pearson quanto nos resíduos quantílicos. Há nos extremos alguns dados que não se ajustam às retas.

Em complemento, foram efetuados o teste de auto-correlação, o teste de normalidade e o teste de homogeneidade da variância, que serão apresentados nas próximas tabelas. Maiores informações sobre o teste de homogeneidade encontra-se no Anexo B.

Tabela D.4: Testes e diagnósticos aplicados à série dos resíduos (Modelo de Atherino univariado) - cobertura Casco.

Modelo de Atherino univariado - Casco	
Testes e diagnósticos	<i>p-valor</i>
Homocedasticidade (Qui-Quadrado)	0,998
Autocorrelação (Box-Ljung)	0,455
Normalidade (Jarque-Bera)	0,001

Tabela D.5: Testes e diagnósticos aplicados à série dos resíduos (Modelo de Atherino univariado) - cobertura RCFV.

Modelo de Atherino univariado - RCFV	
Testes e diagnósticos	<i>p-valor</i>
Homocedasticidade (Qui-Quadrado)	0,999
Autocorrelação (Box-Ljung)	0,601
Normalidade (Jarque-Bera)	0,003

Os diagnósticos do modelo Atherino univariado, para a série Casco, são apresentados na Tabela D.4. No teste de heterocedasticidade, obteve-se um *p-valor* igual a 0,998, não sugerindo a rejeição de H_0 , ou seja, não há evidências de heterocedasticidade. No teste Box-Ljung, não há evidências para rejeitar a hipótese nula, isto é, a série é descorrelatada. No teste de normalidade Jarque-Bera, rejeita-se a hipótese nula, a hipótese de normalidade dos resíduos.

Os diagnósticos aplicados às inovações do método Atherino univariado para a série RCFV (Tabela D.5) apresentaram os seguintes resultados: no teste de heterocedasticidade, obteve-se um *p-valor* próximo a 0,999, não sugerindo evidências de heterocedasticidade. No teste de autocorrelação Box-Ljung, não há evidências para rejeitar a hipótese de ausência de correlação residual. Já no teste de normalidade de Jarque-bera, rejeita-se a hipótese de normalidade para os resíduos auxiliares.

As medidas de aderência, as reservas IBNR e os coeficientes de variação provenientes da aplicação do modelo Atherino univariado foram apresentados na última seção do Capítulo 5, juntamente com os resultados dos modelos propostos nesta tese.

E Modelo GAS simplificado - cobertura Casco

Utilizando o triângulo de *run-off* com dados IBNR da cobertura Casco (Figura 5.1) e o modelo simplificado com distribuição gama ($A = 0$, $B = 0$, $\alpha = e^w$ e $f_t = \omega$, ou seja f_t constante), foram estimados os seguintes parâmetros estáticos (Tabela E.1):

Tabela E.1: Cobertura Casco - parâmetros estimados (modelo simples com distribuição gama).

$f_t = \omega$								
2,515								
(0,100)								
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
4,268	2,621	0,699	-0,561	-1,191	-1,634	-2,207	-2,702	-2,951
(0,120)	(0,122)	(0,1233)	(0,124)	(0,125)	(0,128)	(0,129)	(0,131)	(0,134)
λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
-3,108	-3,803	-3,869	-4,159	-4,083	-4,637	-4,965	-6,096	-5,071
(0,137)	(0,141)	(0,146)	(0,153)	(0,161)	(0,173)	(0,191)	(0,224)	(0,299)

Na Tabela E.2 estão descritas as medidas de aderência calculadas para os dados dentro e fora da amostra.

Tabela E.2: Cobertura Casco - medidas de aderência com distribuição gama.

	Modelo simplificado	
	Dentro da amostra	Fora da amostra
MAPE (%)	24,49	34,96
EQM (+10 ¹³)	11,10	15,14
Pseudo R2 (%)	97,96	99,37

A Tabela E.3 apresenta as reservas IBNR e os coeficientes de variação estimados (modelo simples com distribuição gama).

Tabela E.3: Cobertura Casco - reservas IBNR e coeficiente de variação (Modelo simples com distribuição gama).

Trimestre	Modelo simplificado	
	Reservas IBNR	CV (%)
2	R\$ 77.616,2	29,7%
3	R\$ 105.443,7	24,0%
4	R\$ 191.747,4	18,8%
5	R\$ 311.502,9	16,9%
6	R\$ 519.831,0	16,1%
7	R\$ 712.868,6	14,9%
8	R\$ 971.043,2	14,3%
9	R\$ 1.246.863,9	14,7%
10	R\$ 1.799.076,0	14,5%
11	R\$ 2.445.267,8	13,3%
12	R\$ 3.274.007,0	13,2%
13	R\$ 4.634.262,0	14,0%
14	R\$ 7.046.846,0	13,9%
15	R\$ 10.804.024,0	13,8%
16	R\$ 17.856.732,0	15,0%
17	R\$ 42.724.852,0	13,8%
18	R\$ 212.695.252,0	13,8%
Total	R\$ 307.417.235,4	

O resultado do modelo GAS(1,1) com distribuição gama aponta que não se deve desconsiderar o fator temporal ao selecionar o modelo a ser aplicado aos dados.

F Modelo GAS simplificado - cobertura RCFV

Utilizando o triângulo de *run-off* com dados IBNR da cobertura RCFV (Figura 5.2) e o modelo simplificado com distribuição gama ($A = 0$, $B = 0$, $\alpha = e^w$ e $f_t = w$, ou seja f_t constante), foram estimados os seguintes parâmetros estáticos (Tabela F.1):

Tabela F.1: Cobertura RCFV - parâmetros estimados (Modelo simples com distribuição gama).

$f_t = \omega$								
2,869								
(0,102)								
λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
2,249	1,237	-0,025	-0,797	-1,182	-1,311	-1,420	-1,541	-1,856
(0,116)	(0,117)	(0,118)	(0,119)	(0,120)	(0,121)	(0,123)	(0,124)	(0,126)
λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
-2,092	-2,168	-2,334	-2,471	-2,601	-2,584	-3,058	-2,769	-3,788
(0,129)	(0,132)	(0,136)	(0,140)	(0,147)	(0,157)	(0,170)	(0,195)	(0,257)

Na Tabela F.2 estão descritas as medidas de aderência calculadas para os dados dentro e fora da amostra.

Tabela F.2: Cobertura RCFV - medidas de aderência com distribuição gama.

	Modelo simplificado	
	Dentro da amostra	Fora da amostra
MAPE (%)	20,64	24,63
EQM*	13,16	4,94
Pseudo R2 (%)	96,39	99,42

Nota: "*" - Dentro da amostra ($+10^{12}$) e fora da amostra ($+10^{13}$).

A Tabela F.3 apresenta as reservas IBNR e os coeficientes de variação estimados (modelo simples com distribuição gama).

Tabela F.3: Cobertura RCFV - reservas IBNR e coeficiente de variação (Modelo simples com distribuição gama).

Trimestre	Modelo simplificado	
	Reservas IBNR	CV (%)
2	R\$ 399.161,7	24,0%
3	R\$ 1.505.010,7	18,4%
4	R\$ 2.332.783,0	15,0%
5	R\$ 3.662.325,0	13,3%
6	R\$ 4.970.206,0	12,2%
7	R\$ 6.459.213,0	11,3%
8	R\$ 8.166.622,0	10,3%
9	R\$ 10.182.000,0	9,9%
10	R\$ 12.357.403,0	9,5%
11	R\$ 15.112.539,0	9,0%
12	R\$ 18.888.333,0	8,8%
13	R\$ 23.147.763,0	8,7%
14	R\$ 27.899.740,0	8,5%
15	R\$ 33.304.366,0	8,4%
16	R\$ 41.246.899,0	8,3%
17	R\$ 58.443.529,0	8,3%
18	R\$ 119.187.079,0	8,1%
Total	R\$ 387.264.972,4	

A comparação com o resultado do modelo GAS(1,1) com distribuição gama aponta que não se deve desconsiderar o fator temporal ao selecionar o modelo a ser aplicado aos dados.

G Cobertura Casco - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição gama)

Neste anexo serão apresentados os 18 gráficos contendo as 10.000 simulações que geraram os intervalos de confiança para todos os trimestres descritos na Tabela 5.11.

Figura G.1: GAS (1,1) gama
- IC simulado para $t=2$
(2^otrim.09), Casco.

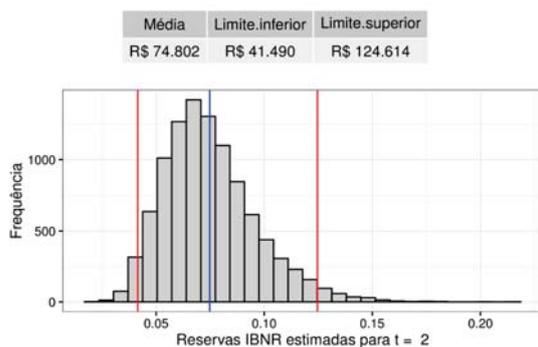


Figura G.2: GAS (1,1) gama
- IC simulado para $t=3$
(3^otrim.09), Casco.

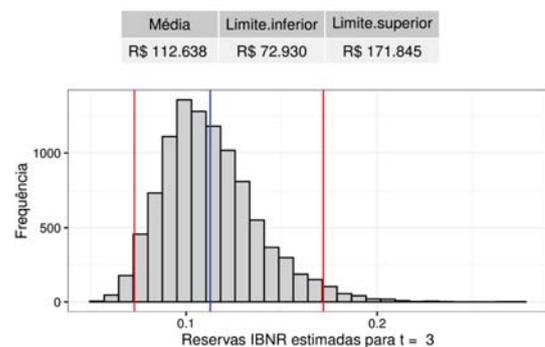


Figura G.3: GAS (1,1) gama
- IC simulado para $t=4$
(4^otrim.09), Casco.

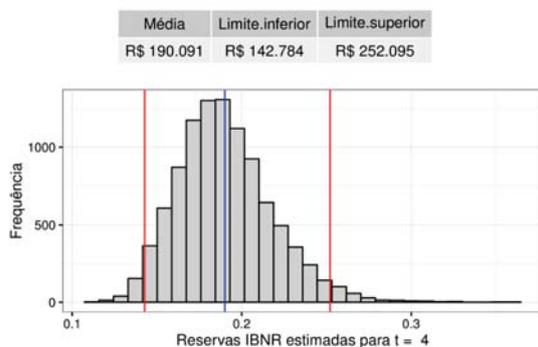


Figura G.4: GAS (1,1) gama
- IC simulado para $t=5$
(1^otrim.10), Casco.

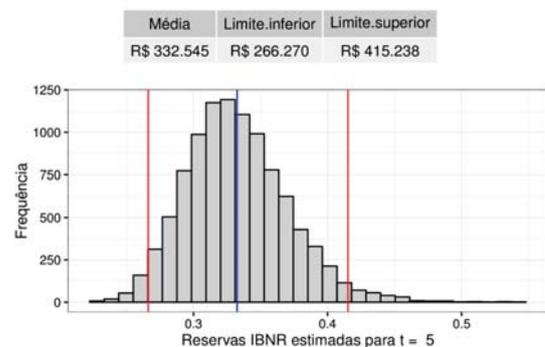


Figura G.5: GAS (1,1) gama - IC simulado para t=6 (2^otrim.10), Casco.

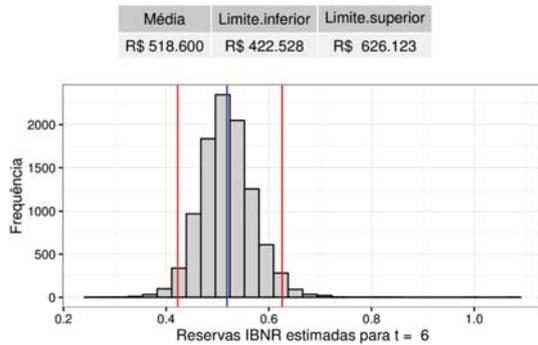


Figura G.6: GAS (1,1) gama - IC simulado para t=7 (3^otrim.10), Casco.

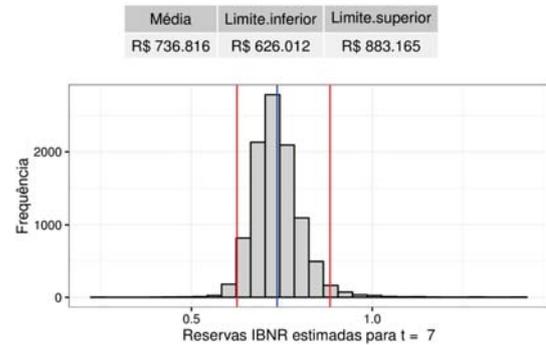


Figura G.7: GAS (1,1) gama - IC simulado para t=8 (4^otrim.10), Casco.

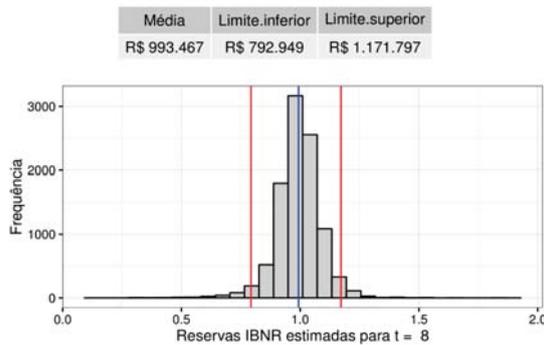


Figura G.8: GAS (1,1) gama - IC simulado para t=9 (1^otrim.11), Casco.

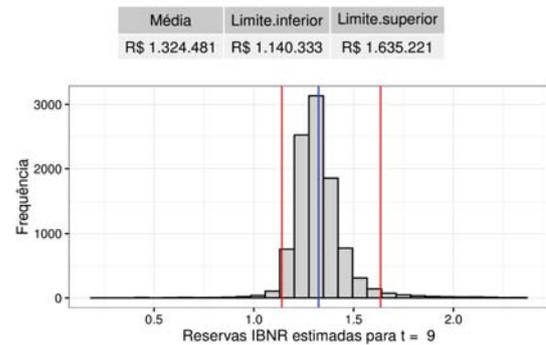


Figura G.9: GAS (1,1) gama - IC simulado para t=10 (2^otrim.11), Casco.

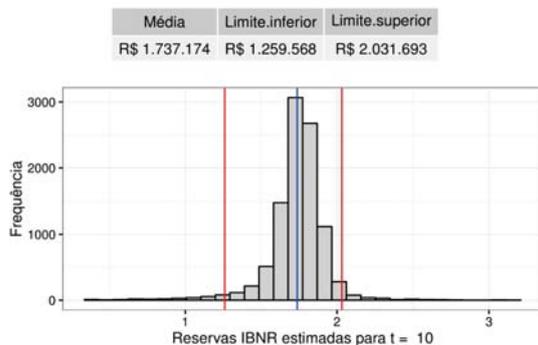


Figura G.10: GAS (1,1) gama - IC simulado para t=11 (3^otrim.11), Casco.

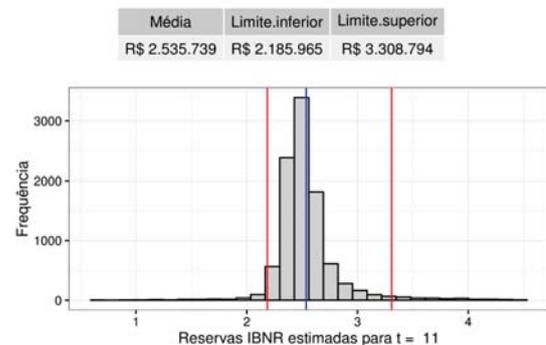


Figura G.11: GAS (1,1) gama
- IC simulado para t=12
(4^otrim.11), Casco.

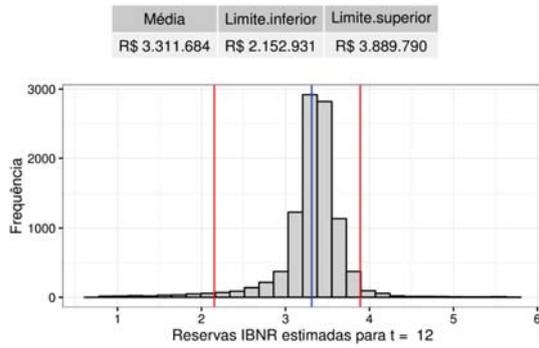


Figura G.12: GAS (1,1) gama
- IC simulado para t=13
(1^otrim.12), Casco.

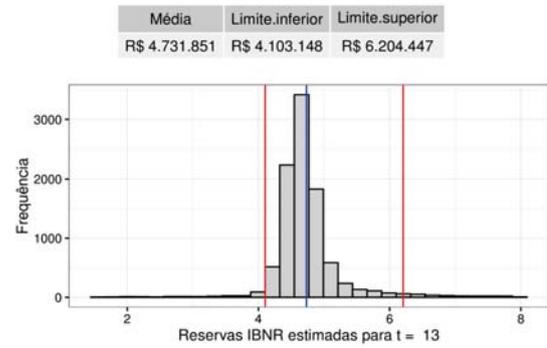


Figura G.13: GAS (1,1) gama
- IC simulado para t=14
(2^otrim.12), Casco.

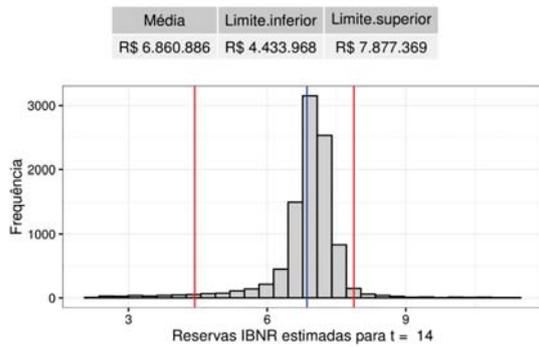


Figura G.14: GAS (1,1) gama
- IC simulado para t=15
(3^otrim.12), Casco.

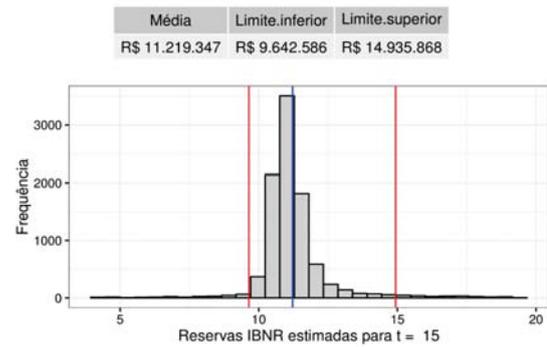


Figura G.15: GAS (1,1) gama
- IC simulado para t=16
(4^otrim.12), Casco.

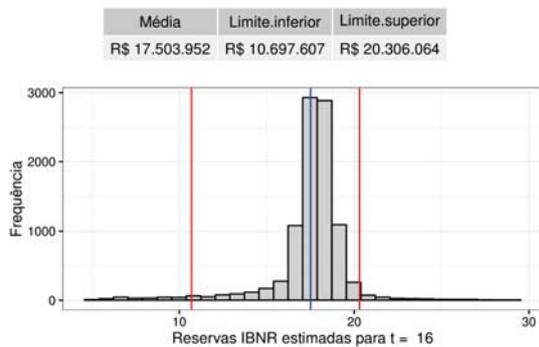


Figura G.16: GAS (1,1) gama
- IC simulado para t=17
(1^otrim.13), Casco.

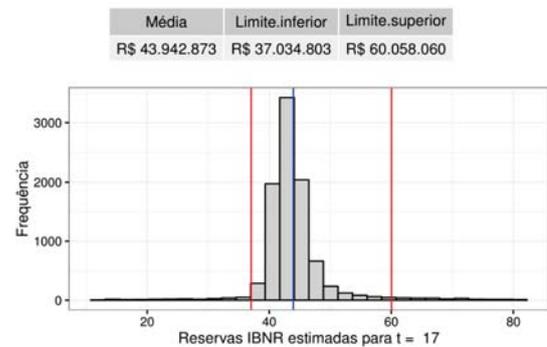
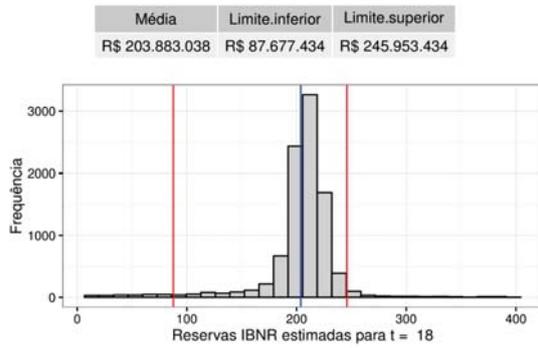


Figura G.17: GAS (1,1) gama
 - IC simulado para $t=18$
 (2^{o} trim.13), Casco.



As simulações quando $t = 2, 3, 4, 5, 6, 7, 9, 11, 13, 15$ e 17 são assimétricas para a direita, enquanto as simulações quando $t = 8, 10, 12, 14, 16$ e 18 são assimétricas para a esquerda.

H

Cobertura Casco - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição log-normal)

Figura H.1: GAS (1,1) Log-normal - IC simulado para t=2 (2^otrim.09), Casco.

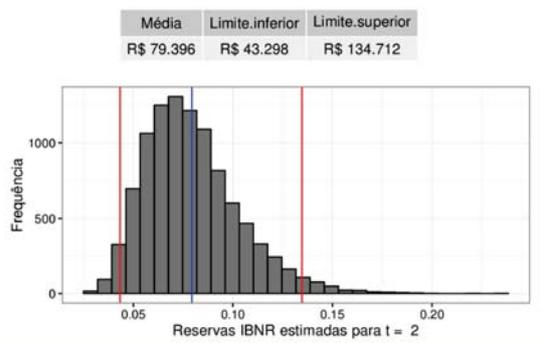


Figura H.2: GAS (1,1) Log-normal - IC simulado para t=3 (3^otrim.09), Casco.

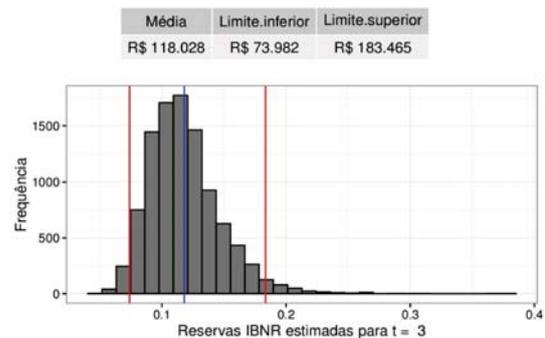


Figura H.3: GAS (1,1) Log-normal - IC simulado para t=4 (4^otrim.09), Casco.

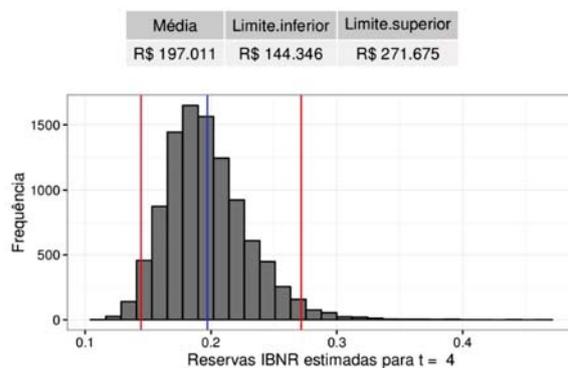


Figura H.4: GAS (1,1) Log-normal - IC simulado para t=5 (1^otrim.10), Casco.

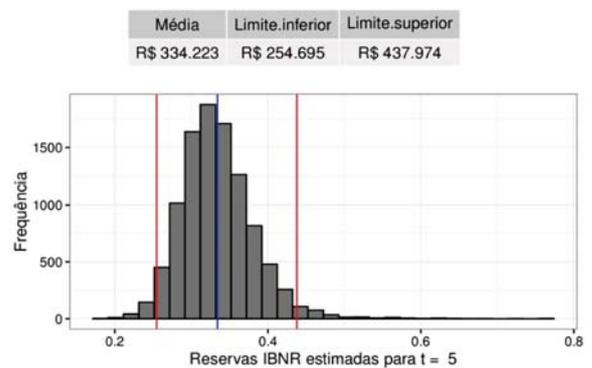


Figura H.5: GAS (1,1) Log-normal - IC simulado para t=6 (2^otrim.10), Casco.

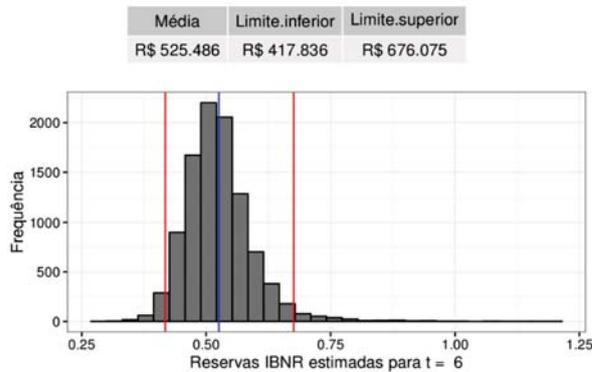


Figura H.6: GAS (1,1) Log-normal - IC simulado para t=7 (3^otrim.10), Casco.

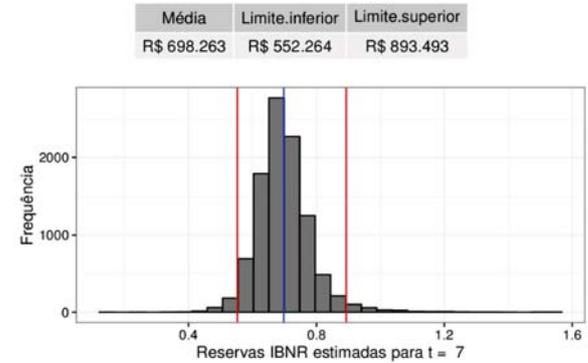


Figura H.7: GAS (1,1) Log-normal - IC simulado para t=8 (4^otrim.10), Casco.

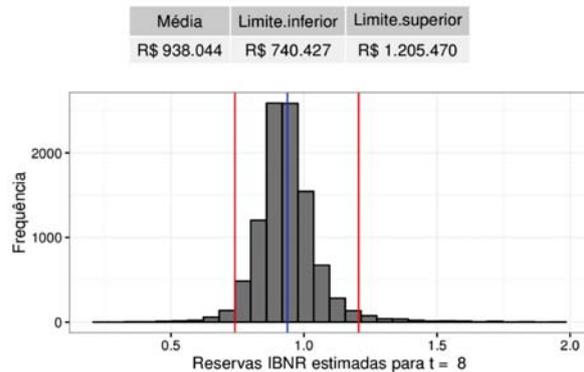


Figura H.8: GAS (1,1) Log-normal - IC simulado para t=9 (1^otrim.11), Casco.

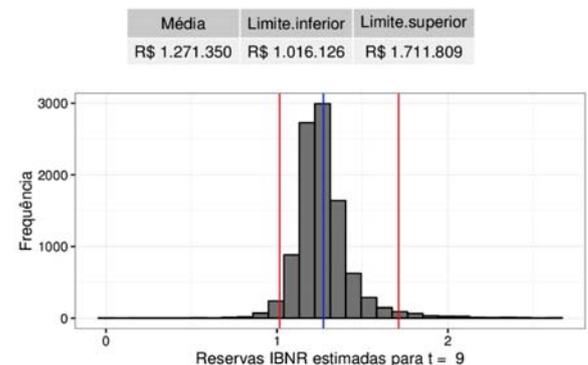


Figura H.9: GAS (1,1) Log-normal - IC simulado para t=10 (2^otrim.11), Casco.

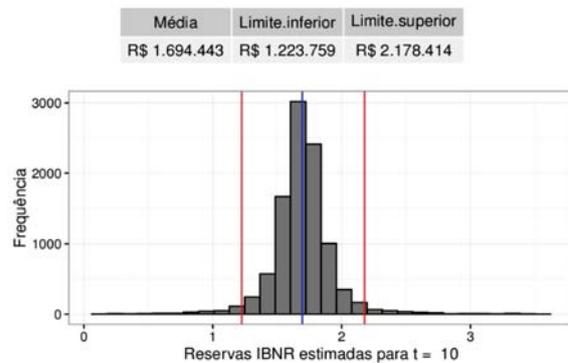


Figura H.10: GAS (1,1) Log-normal - IC simulado para t=11 (3^otrim.11), Casco.

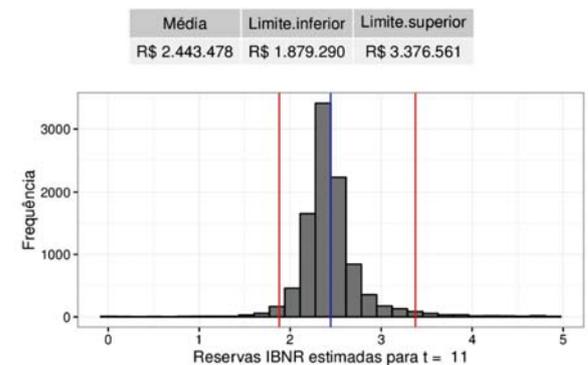


Figura H.11: GAS (1,1) Log-normal - IC simulado para $t=12$ (4^otrim.11), Casco.

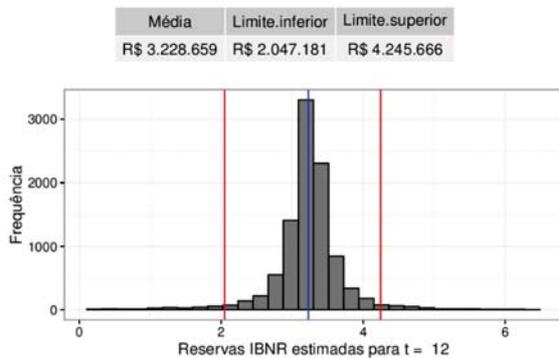


Figura H.12: GAS (1,1) Log-normal - IC simulado para $t=13$ (1^otrim.12), Casco.

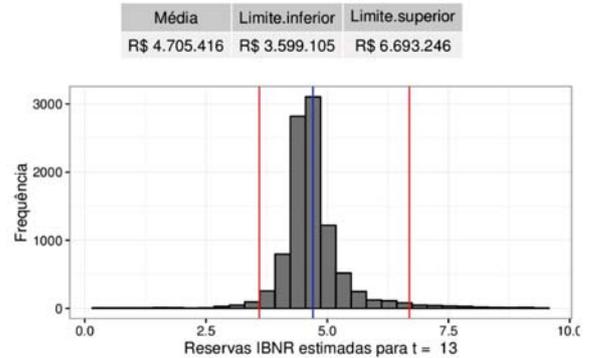


Figura H.13: GAS (1,1) Log-normal - IC simulado para $t=14$ (2^otrim.12), Casco.

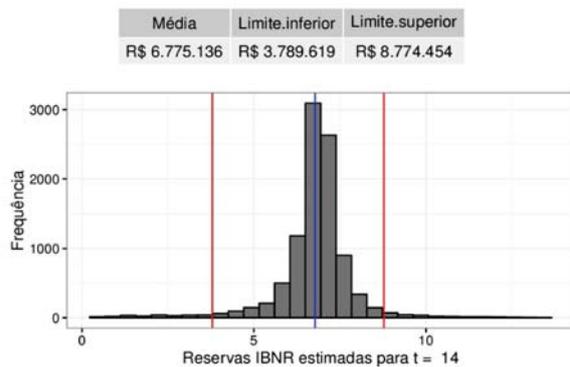


Figura H.14: GAS (1,1) Log-normal - IC simulado para $t=15$ (3^otrim.12), Casco.

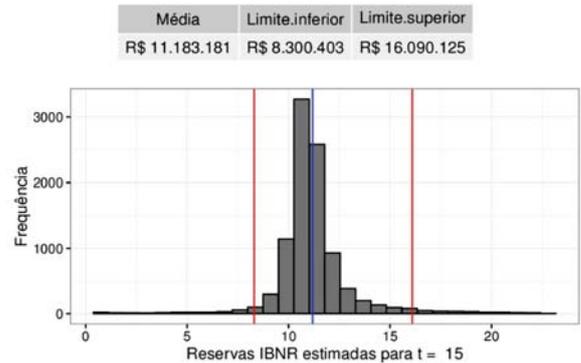


Figura H.15: GAS (1,1) Log-normal - IC simulado para $t=16$ (4^otrim.12), Casco.

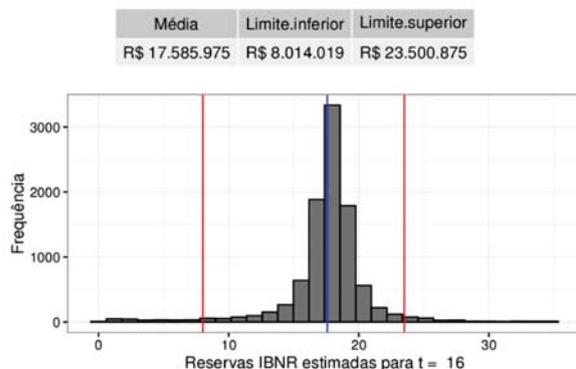


Figura H.16: GAS (1,1) Log-normal - IC simulado para $t=17$ (1^otrim.13), Casco.

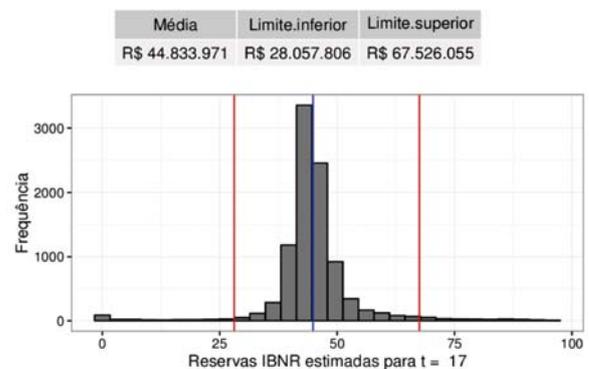
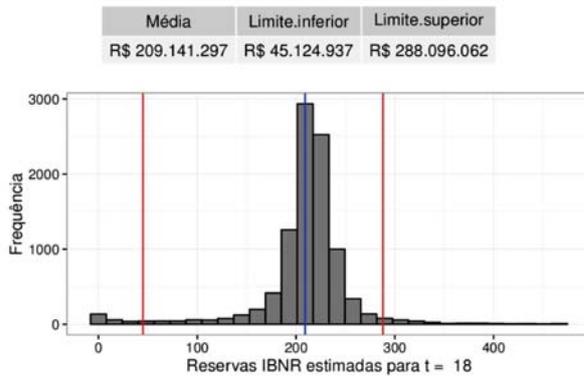


Figura H.17: GAS (1,1) Log-normal - IC simulado para $t=18$ (2^otrim.13), Casco.



Como pode ser observado nos gráficos acima, as simulações quando $t = 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 15$ e 17 são assimétricas para a direita, enquanto as simulações quando $t = 12, 14, 16$ e 18 são assimétricas para a esquerda. O resultado das simulações quando $t = 10$ apresenta um formato mais simétrico.

I

Cobertura RCFV - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição gama)

Figura I.1: GAS (1,1) gama
- IC simulado para t=2
(2^otrim.09), RCFV.

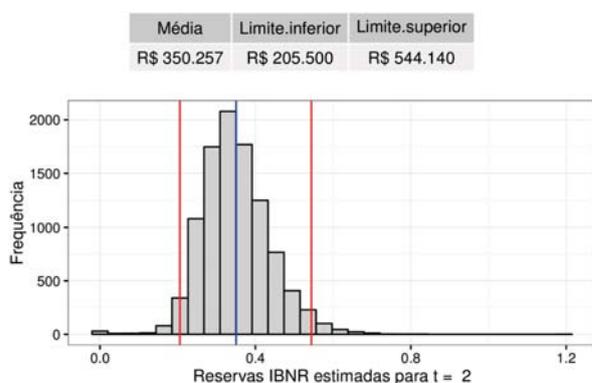


Figura I.2: GAS (1,1) gama
- IC simulado para t=3
(3^otrim.09), RCFV.

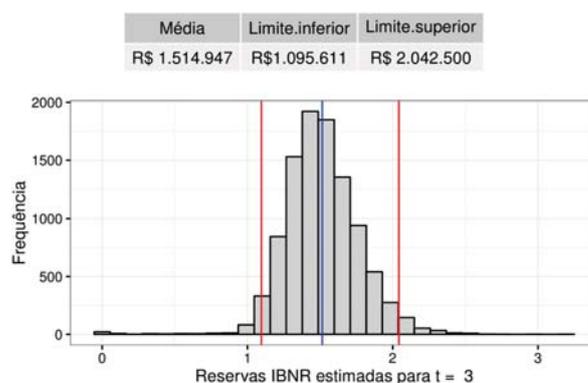


Figura I.3: GAS (1,1) gama
- IC simulado para t=4
(4^otrim.09), RCFV.

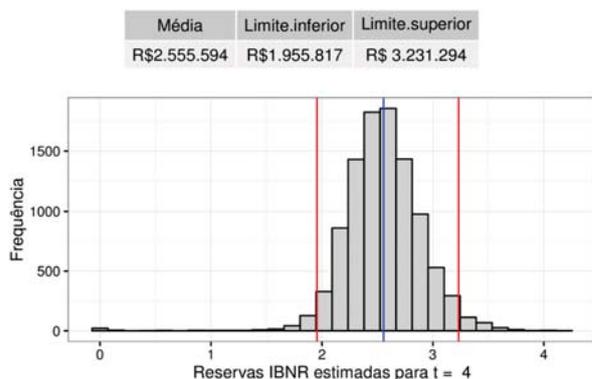


Figura I.4: GAS (1,1) gama
- IC simulado para t=5
(1^otrim.10), RCFV.

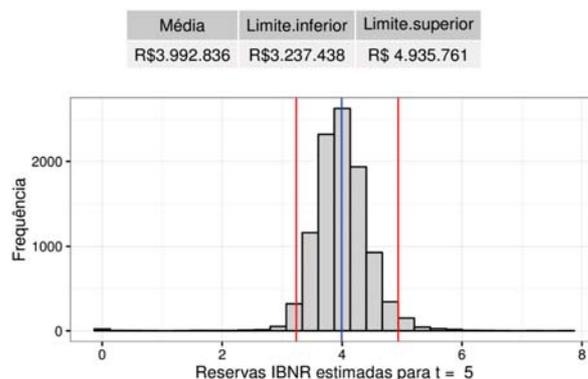


Figura I.5: GAS (1,1) gama - IC simulado para t=6 (2^otrim.10), RCFV.

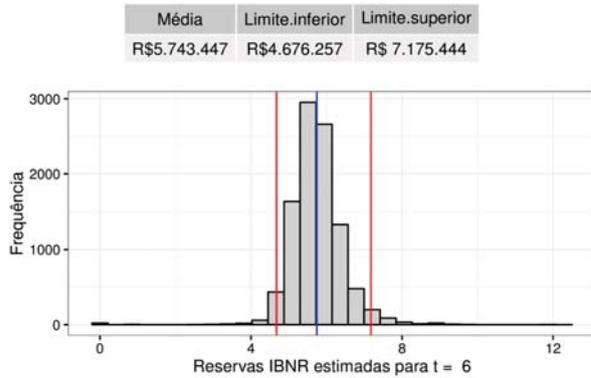


Figura I.6: GAS (1,1) gama - IC simulado para t=7 (3^otrim.10), RCFV.

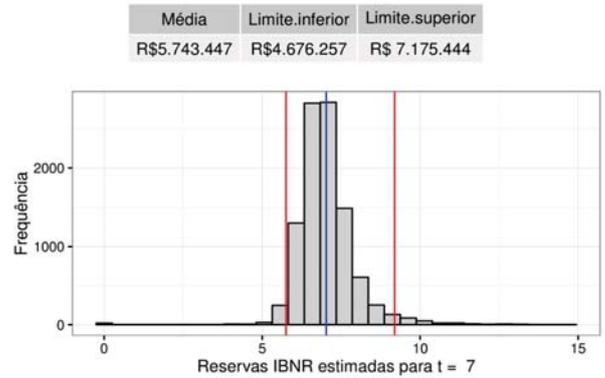


Figura I.7: GAS (1,1) gama - IC simulado para t=8 (4^otrim.10), RCFV.

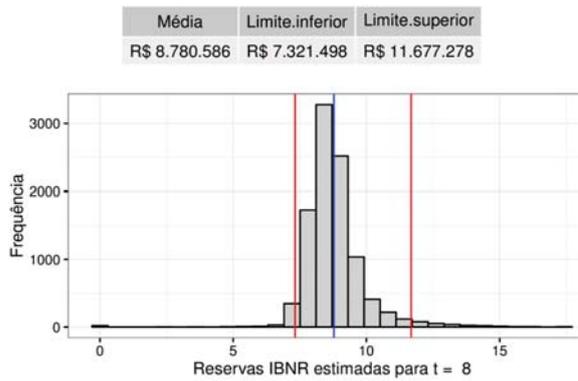


Figura I.8: GAS (1,1) gama - IC simulado para t=9 (1^otrim.11), RCFV.

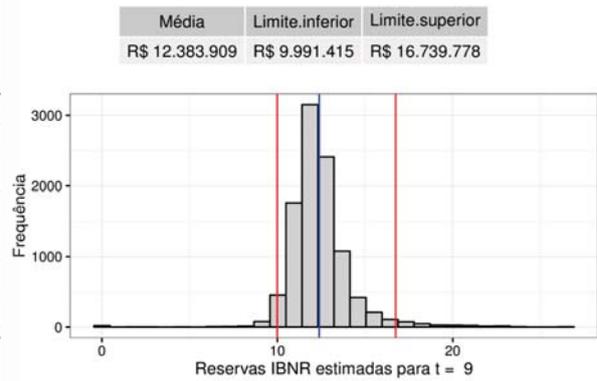


Figura I.9: GAS (1,1) gama - IC simulado para t=10 (2^otrim.11), RCFV.

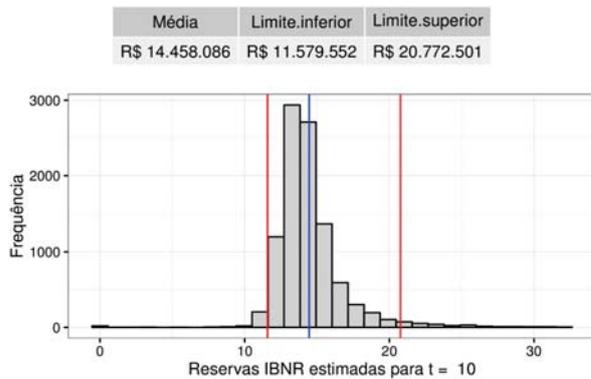


Figura I.10: GAS (1,1) gama - IC simulado para t=11 (3^otrim.11), RCFV.

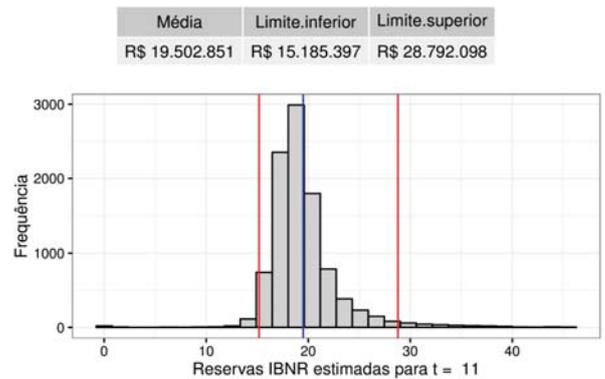


Figura I.11: GAS (1,1) gama
- IC simulado para t=12
(4^otrim.11), RCFV.

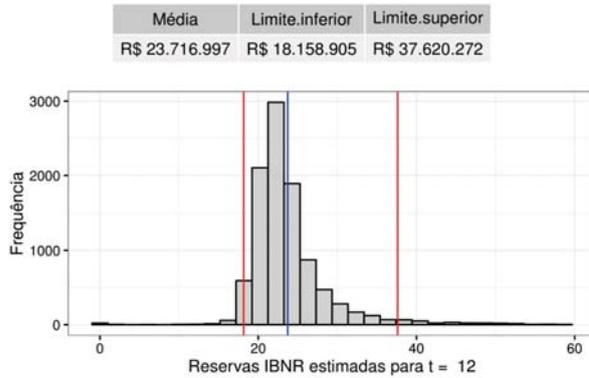


Figura I.12: GAS (1,1) gama
- IC simulado para t=13
(1^otrim.12), RCFV.

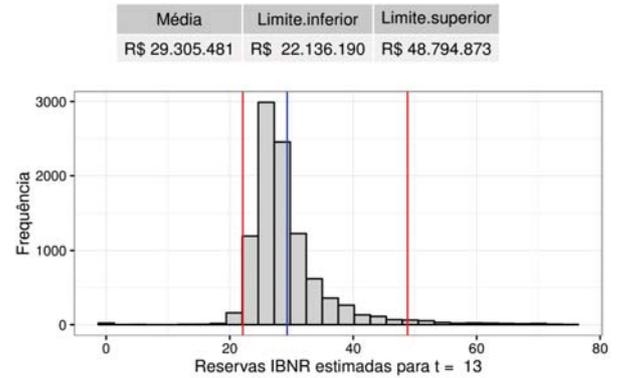


Figura I.13: GAS (1,1) gama
- IC simulado para t=14
(2^otrim.12), RCFV.

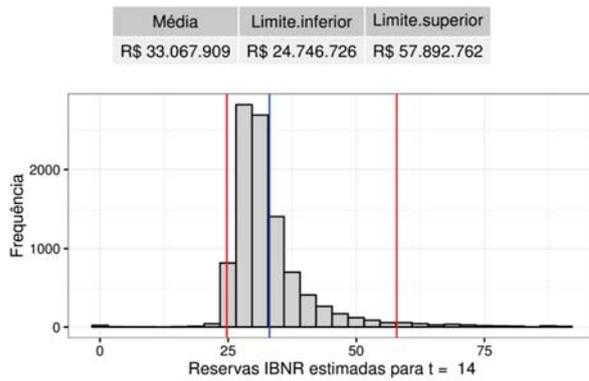


Figura I.14: GAS (1,1) gama
- IC simulado para t=15
(3^otrim.12), RCFV.

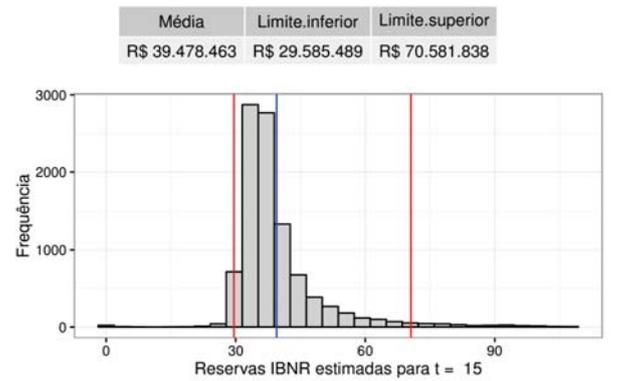


Figura I.15: GAS (1,1) gama
- IC simulado para t=16
(4^otrim.12), RCFV.

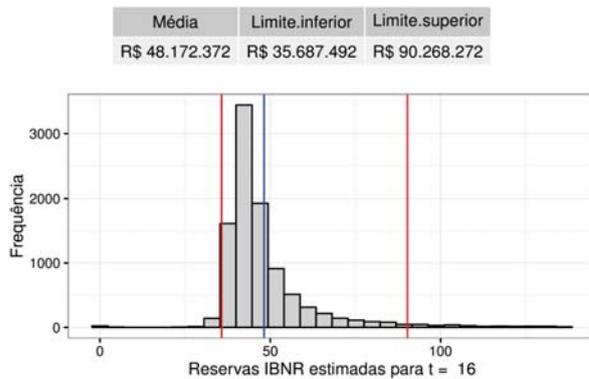


Figura I.16: GAS (1,1) gama
- IC simulado para t=17
(1^otrim.13), RCFV.

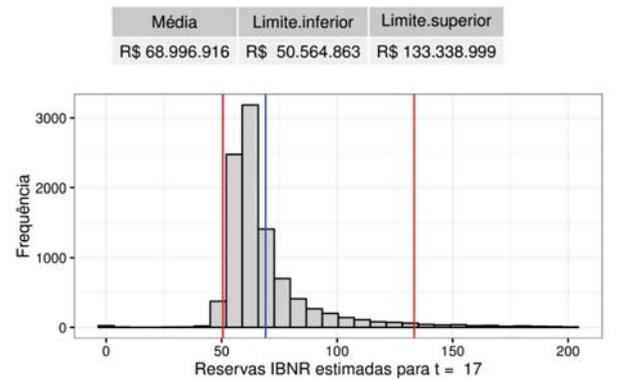
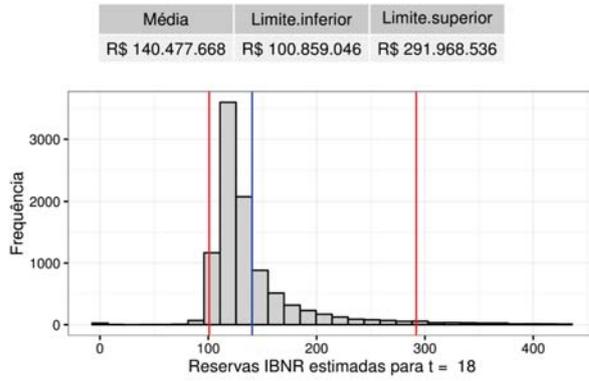


Figura I.17: GAS (1,1) gama
- IC simulado para $t=18$
(2^otrim.13), RCFV.



Para a cobertura RCFV, como pode ser observado nos gráficos acima, as simulações das reservas IBNR, para cada trimestre, apresentam simetria à direita.

J

Cobertura RCFV - gráficos de dispersão dos intervalos de confiança para o IBNR (modelo GAS (1,1) com distribuição log-normal)

Figura J.1: GAS (1,1) log-normal - IC simulado para t=2 (2^otrim.09), RCFV.

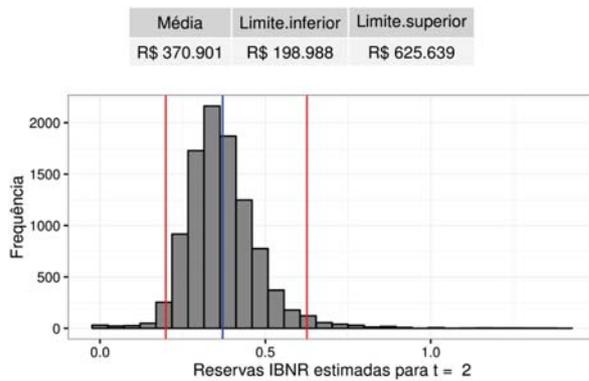


Figura J.2: GAS (1,1) log-normal - IC simulado para t=3 (3^otrim.09), RCFV.

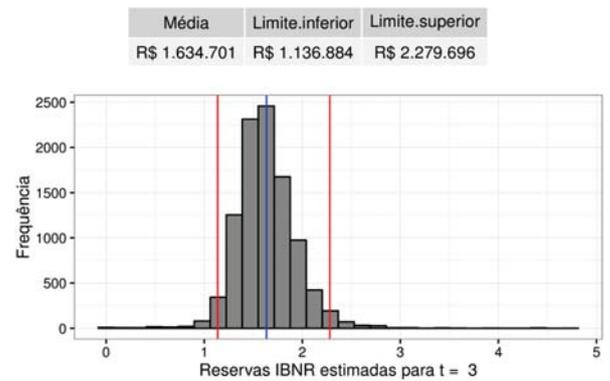


Figura J.3: GAS (1,1) log-normal - IC simulado para t=4 (4^otrim.09), RCFV.

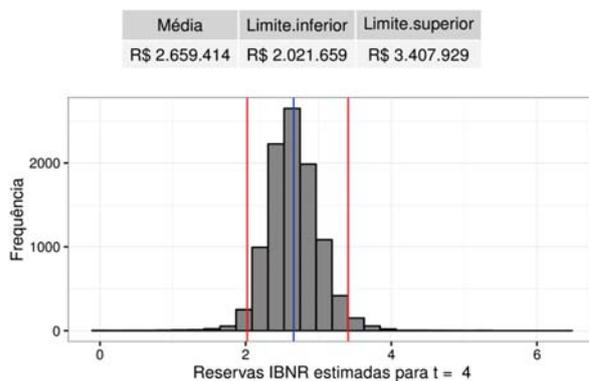


Figura J.4: GAS (1,1) log-normal - IC simulado para t=5 (1^otrim.10), RCFV.

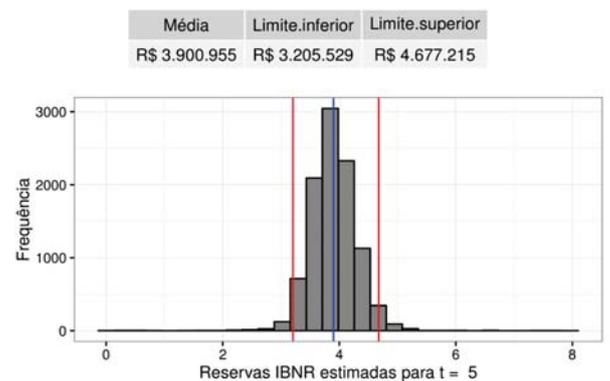


Figura J.5: GAS (1,1) log-normal - IC simulado para t=6 (2^otrim.10), RCFV.

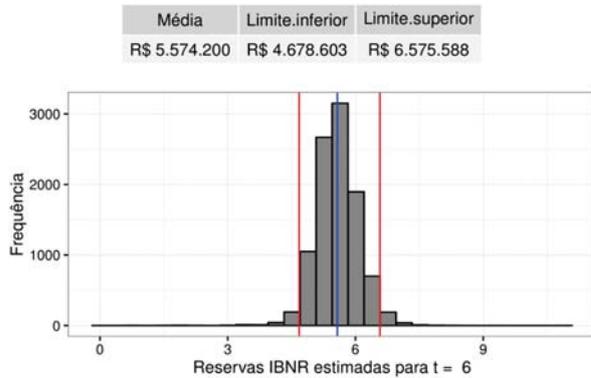


Figura J.6: GAS (1,1) log-normal - IC simulado para t=7 (3^otrim.10), RCFV.

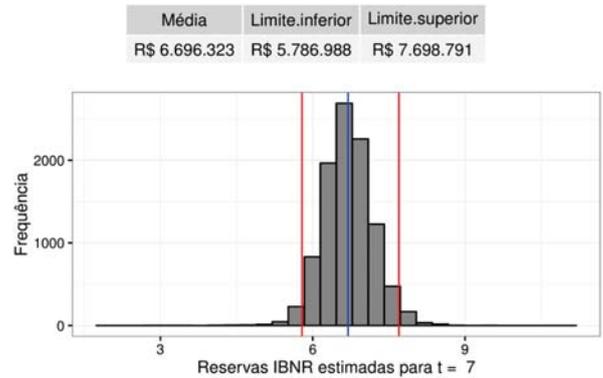


Figura J.7: GAS (1,1) log-normal - IC simulado para t=8 (4^otrim.10), RCFV.

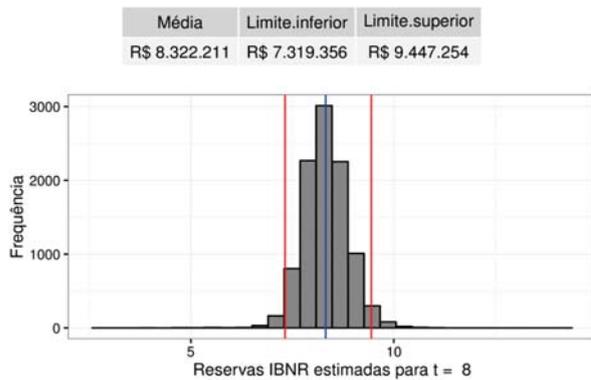


Figura J.8: GAS (1,1) log-normal - IC simulado para t=9 (1^otrim.11), RCFV.

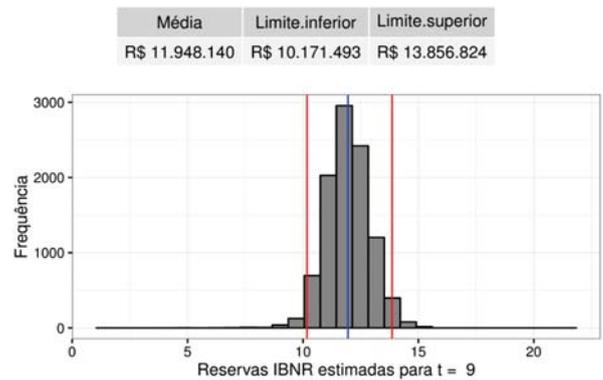


Figura J.9: GAS (1,1) log-normal - IC simulado para t=10 (2^otrim.11), RCFV.

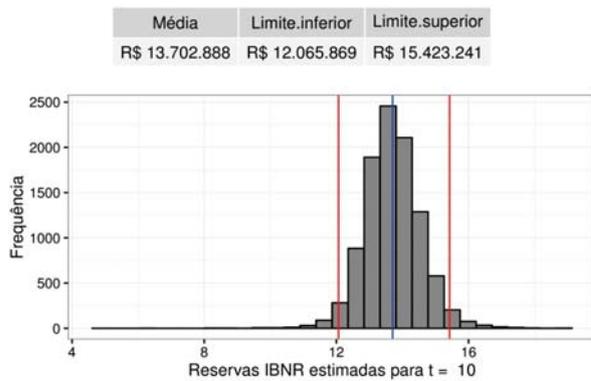


Figura J.10: GAS (1,1) log-normal - IC simulado para t=11 (3^otrim.11), RCFV.

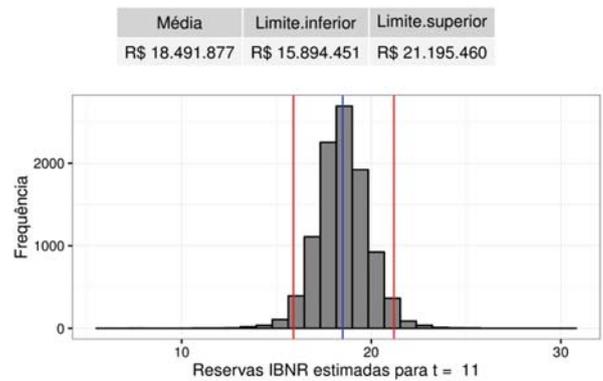


Figura J.11: GAS (1,1) log-normal - IC simulado para $t=12$ (4^otrim.11), RCFV.

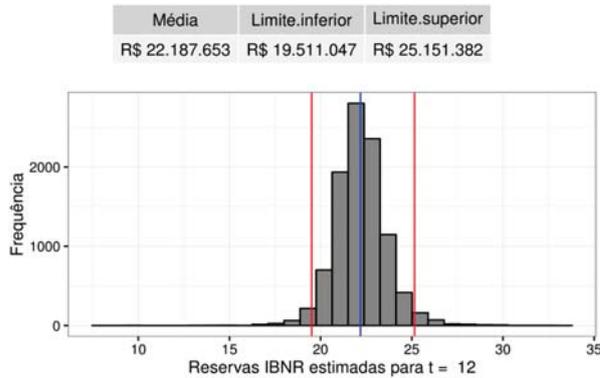


Figura J.12: GAS (1,1) log-normal - IC simulado para $t=13$ (1^otrim.12), RCFV.

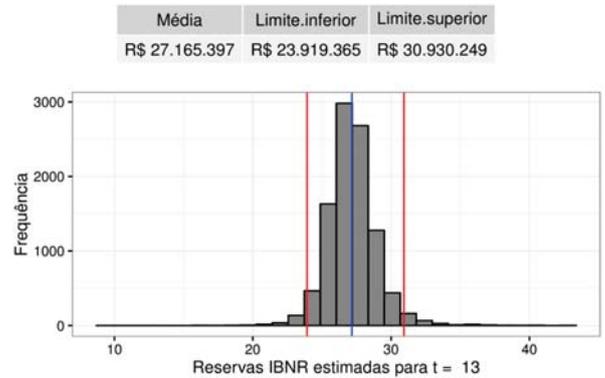


Figura J.13: GAS (1,1) log-normal - IC simulado para $t=14$ (2^otrim.12), RCFV.

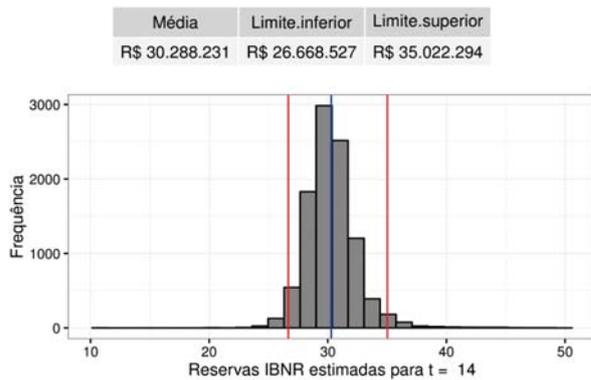


Figura J.14: GAS (1,1) log-normal - IC simulado para $t=15$ (3^otrim.12), RCFV.

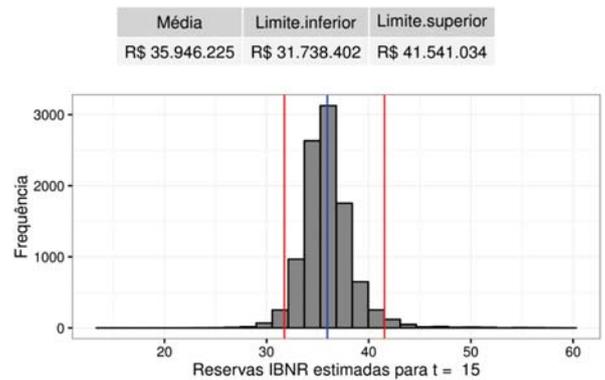


Figura J.15: GAS (1,1) log-normal - IC simulado para $t=16$ (4^otrim.12), RCFV.

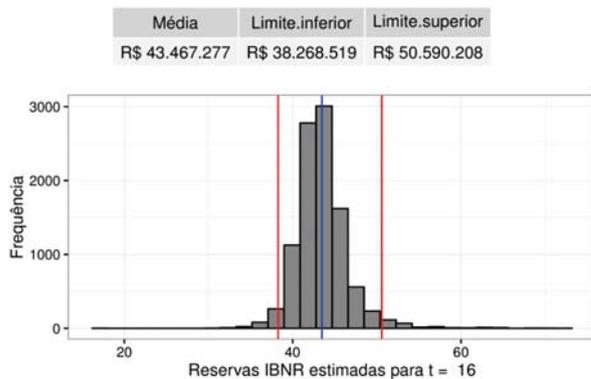


Figura J.16: GAS (1,1) log-normal - IC simulado para $t=17$ (1^otrim.13), RCFV.

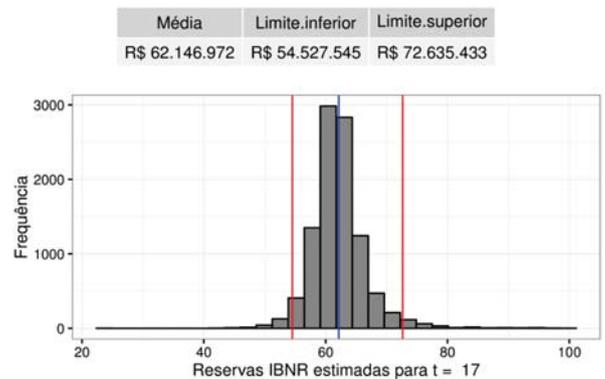
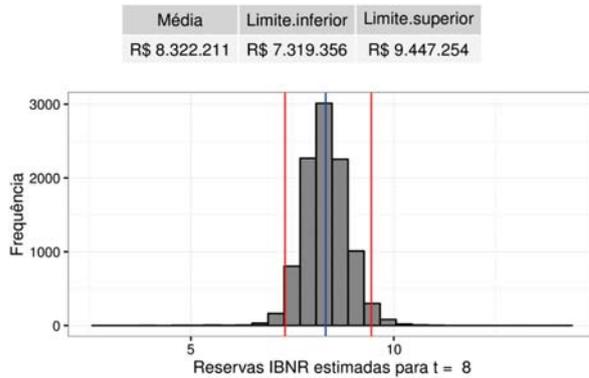


Figura J.17: GAS (1,1) log-normal - IC simulado para $t=18$ (2^otrim.13), RCFV.



Para a cobertura RCFV, como pode ser observado nos gráficos acima, as simulações dos sinistros IBNR obtidas por simulação, para cada trimestre, apresentam formatos distintos, as distribuições quando $t = 2, 3, 4, 13, 14, 16$ e 17 são assimétricas para a direita, enquanto as distribuições quando $t = 5, 6, 7, 9, 10, 11, 12, 15$ e 18 apresentam formato mais simétrico.